SOCIAL IMPACT

IMPROVING DEVELOPMENT EFFECTIVENESS

Millennium Challenge Corporation

Impact Evaluation of the MCA Jordan Compact Results of Component 1 – Impacts of Water Network and Wastewater Network Projects

July 2015

This publication was produced by Social Impact at the request of the US Millennium Challenge Corporation. It was prepared independently by Marc Jeuland, Jenny Orgill, Sabreen Alikhan, and Nathan Cutler.

Impact Evaluation of the MCA Jordan Compact Results of Component 1 – Water Network and Wastewater Network Projects

Prepared by Social Impact, Inc.
for the Millennium Challenge Corporation

July 15, 2015

Marc Jeuland[1], Jenny Orgill[1],
Sabreen Alikhan[2], Nathan Cutler[2]

[1] Duke University, Durham, NC USA

[2] Social Impact, Arlington, VA USA. www.socialimpact.com Please address communication to Sabreen Alikhan at salikhan@socialimpact.com

LIST OF ACRONYMS

CI	Confidence Interval
DA	Development Area
DiD	Difference-in-Differences
DHS	Demographic and Health Survey
DMA	District Metering Area
DoS	Jordanian Department of Statistics
ERR	Economic Rate of Return
FGD	Focus Group Discussion
GIS	Geographic Information System
GOJ	Government of Jordan
GPSM	Generalized Propensity Score Matching
ICC	Intra-Cluster Correlation
IE	Impact Evaluation
IRB	Institutional Review Board
ITT	Intention-to-Treat
IV	Instrumental Variable
JVA	Jordan Valley Authority
KAC	King Abdullah Canal
KII	Key Informant Interview
KTD	King Talal Dam
KTR	King Talal Reservoir
M&E	Monitoring and Evaluation
MCA	Millennium Challenge Account
MCA-J	Millennium Challenge Account – Jordan
MCC	Millennium Challenge Corporation
MoH	Ministry of Health
MWI	Ministry of Water and Irrigation
NAF	National Aid Fund
NRW	Non-Revenue Water
PMC	Program Management Consultant
PMU	Project Management Unit
PSM	Propensity Score Matching
PSU	Primary Sampling Unit
RCT	Randomized Control Trial
RD	Regression Discontinuity
SES	Socio-Economic Status
SI	Social Impact
SOW	Scope of Work
WAJ	Water Authority of Jordan
WASH	Water, Sanitation and Hygiene
WSH	Water Smart Homes Activity
WHO	World Health Organization
WTP	Willingness to pay

WWTP	Wastewater Treatment Plant

TABLE OF CONTENTS

1 INTRODUCTION

The Hashemite Kingdom of Jordan is one of the four driest countries in the world facing severe water scarcity with declining per capita water resources as a result of population growth and decreasing water availability (Hashemite Kingdom of Jordan 2008). The scarcity of water in Jordan has been called the single most important constraint inhibiting the country's future growth and poverty alleviation. Water constraints not only affect economic development, but also have consequences for food production, health, social and human development. According to the Ministry of Water and Irrigation (MWI) and the Food and Agriculture Organization, renewable water availability in Jordan declined from 3600m^3/capita-year in 1946 to 125m^3/capita-year in 2014, well below the international water poverty line of 500m^3/year (Hashemite Kingdom of Jordan 2009, FAO 2015). Moreover, water sector costs represent a major drain on the Government of Jordan (GOJ)'s fiscal resources, accounting for about 5% of the 2010 national budget and 17% of the 2010-2013 capital investment program (USAID 2011).

The challenges of water scarcity and its consequences for economic activity and poverty are amplified in Zarqa Governorate, a populous, dry, and mostly urban governorate west of the capital Amman. Nearly three in ten households in Zarqa consume less than the minimum amount of water considered essential for personal hygiene and food safety by the World Health Organization (WHO) (MCC 2009). This is mainly due to irregular water availability: many households receive piped water only for a limited number of hours, one or two days per week.

Against this backdrop, the MCC has been working with the GOJ to implement an investment Compact that is aimed at addressing water challenges in Zarqa. Social Impact (SI) has been contracted by the MCC to measure the impacts of the Compact activities on economic and social outcomes. The Impact Evaluation (IE) aims to establish a *causal* relationship between program interventions and observed changes in household availability and consumption of different sources of water, income, expenditure, and health indicators. The IE also aims to measure potential impacts on other sectors (agriculture, utility financial performance, and local enterprises) should these occur in parallel to, or instead of, the expected impacts on households. This IE is, to our knowledge, the first attempt to conduct a rigorous IE design of a large infrastructure project in Jordan. It provides a unique opportunity for the MCC, the GOJ, and the broader development community to understand the impact of a large water investment on income and poverty of urban households and others who are affected by it. The IE design as a whole attempts to measure a diverse set of impacts that are differentially related to the three Compact projects, in order for the MCC to better understand which specific component(s) led to the observed changes in outcomes. A comparison of these different impacts will further allow for conclusions about the relative cost-effectiveness of each intervention.

This report details the results of baseline data collection activities conducted in association with the first of three IE components. Component 1 of the IE specifically aims to measure the impacts of the Water Network Project (WNP) and Wastewater Network Project (WWNP) on households and enterprises in Zarqa. For this purpose, two data collection activities were undertaken. The first was a baseline household-level survey, used to gather information on household demographics; water sourcing, storage, and use behaviors; preferences and satisfaction with water supply and sewer service; water quality measured at the tap and in in-house storage containers (chlorine residual, salinity, turbidity, and *E. coli* or *thermo-tolerant coliform* counts); coping and health costs related to intermittent water supply and poor water quality; and expenditures, income, and other socio-economic characteristics. The second was a baseline enterprise survey, which focused on enterprise characteristics, production inputs and outputs, costs and revenues, and constraints with regards to use of water as an input to production. The baseline household and enterprise surveys were conducted in the spring and winter of 2014, respectively, in conjunction with the Jordanian Department of Statistics (DoS).

2 OVERVIEW OF THE COMPACT AND THE IMPACT EVALUATION DESIGN

2.1 Summary of Compact Activities

The Millennium Challenge Corporation (MCC)'s five-year, $275 million Compact aims to reduce poverty and increase income in Zarqa Governorate through improvements to the water network, the extension of wastewater collection and the expansion of wastewater treatment. The combined Compact projects should theoretically improve the efficiency of water delivery in Zarqa. The MCC Compact investment was motivated in large part by the **primary substitution effect**, through which increases in the use of recycled wastewater[3] in agriculture enables increases in conventional freshwater availability for higher-value municipal uses. The entity charged with implementing the Compact in Jordan is the Millennium Challenge Account – Jordan (MCA-J).

The MCC Jordan Compact includes three inter-linked projects in the water sector in Zarqa Governorate:

(i) The Water Network Project (WNP) consists of two activities, a) the rehabilitation and restructuring of water supply transmission and distribution infrastructure, and replacement of domestic water meters, with the aim of improving the overall water system efficiency, reducing water losses and facilitating the transition from periodic distribution under high pressure to more consistent, gravity-fed distribution; and b) the Water Smart Homes (WSH) activity, a household-level intervention aimed at improving in-house water storage and sanitation that consists of a general outreach campaign, as well as delivery of infrastructure subsidies and technical assistance to poor households.

(ii) The Wastewater Network Project (WWNP) encompasses the expansion, rehabilitation and reinforcement of the wastewater network in West and East Zarqa, as well as West Ruseifa, aimed at improving the overall wastewater system efficiency and expanding the capture of municipal wastewater for reuse in agriculture downstream, possibly making additional freshwater available to the population of Zarqa Governorate through future wastewater substitutions for conventional freshwater.

(iii) The As-Samra Expansion Project (AEP) is designed to raise the capacity of the existing treatment plant with the aim of providing proper handling of increased volumes and

[3] Throughout this report, we refer to **recycled wastewater** as **treated wastewater** (i.e. wastewater treated to discharge standards governing wastewater releases in a particular location, which itself need not be reused) that is reused for any purpose. **Blended water** is a water supply that combines both runoff (from precipitation) and discharges of treated wastewater.

loads of both oxygen-demanding material and suspended solids, allowing treatment of the additional wastewater volumes resulting from the WNP and WWNP investments.

2.2 Rationale for this Impact Evaluation

As with all Impact Evaluations (IEs) funded by MCC, the Jordan Water IE is designed to meet the dual goals of learning and accountability. The research questions, evaluation methodology, and outcomes of interest are selected to maximize the utility of evaluation findings. In addition to answering programmatic questions about the effectiveness of the intervention and how benefits accrue to population sub-groups (e.g., women), the evaluation seeks to inform future MCC programming, and to improve the effectiveness and efficiency of investment decisions. By documenting and substantiating lessons learned with rigorous research methodology, the evaluation will provide useful and actionable information to MCC and the MCA-J senior management, project managers, beneficiaries, implementing partners, evaluators, and other evaluation stakeholders, most notably the Government of Jordan (GOJ). Lastly, with MCC's emphasis on transparency, the findings and data will be shared with the broader donor and development community, supplementing the global knowledge pool and amplifying the utility of the Jordan Water IE.

The IE will also help MCC to recalculate the Economic Rate of Return (ERR) of the Compact investment in Jordan following the investments. The SI team has, in close collaboration with the MCC and MCA-J technical teams, reviewed assumptions behind the original ERR calculation and identified areas in which the IE will provide MCC with new inputs to update this calculation in the future. In the same vein, the IE design has been developed in a manner that allows for accurate determination of the most appropriate and necessary inputs to the ERR calculations in order to maximize the utility of the IE. It is important to note, however, that not all inputs to the final ERR are to be supplied by the IE as some of these indicators are not impact estimates.

This IE also has the potential to contribute in meaningful ways to the existing literature on economic development and poverty reduction. Given the scale and anticipated impact of the Jordan Water interventions, MCC and the broader donor community have much to learn about which intervention or combination of interventions can be most effective and efficient in increasing available income through reduced water expenditures. In particular, there remains relatively little rigorous evidence on the impact of urban infrastructure interventions on household level outcomes, and even less on the private (enterprise) sector.

This IE is, to our knowledge, the first attempt to conduct a rigorous counterfactual IE design of a large infrastructure project in Jordan and will provide a unique opportunity for the MCC, the GOJ, and the broader development community to understand the true impacts of a large urban water investment. Finally, this IE will provide a unique and new dataset that can be used by other researchers to look at questions related to the effect of improved water and wastewater supply and systems on a series of household level outcomes. Following end line data collection, the team will synthesize the data into a report that will also be submitted for publication in the form of one or more articles in a peer-reviewed journal. As with all evaluations conducted by MCC, anonymized data will be made available for public use. This transparency will further facilitate the MCC goal of promoting learning.

2.3 Logic of the Impact Evaluation

As emphasized in pre-project feasibility studies and economic analyses of the Compact investments, the economic case for the MCC investments rests on a complex and interrelated set of hypothesized changes that flow from projects to outputs, then to outcomes and finally overarching goals. The linkages between the various components and these intermediate and final changes, respectively, are depicted in the IE logic (Figure 2.1). Importantly, as described in greater detail in the Evaluation Design Report (EDR), this diagram does not directly follow the categorization of impacts promulgated in other prior descriptions of Compact impacts (e.g., in those accompanying the MCC's economic rate of return analysis), for the following main reasons:

1) The impacts included in those analyses were admittedly non-exhaustive, due to data limitations (for example, effects on enterprises and/or on property values were omitted from the analyses)

2) The purposes of the IE logic are a) to trace out the *hypothesized* relationships between projects and intermediate outputs (grouped by color), and final outcomes and goals, b) to illustrate the overlapping relationships between project activities and desired outcomes, and c) to draw attention to some of the key assumptions underlying the case for the investment (as shown in annotations to Figure 2.1).

The IE logic aims to identify the set of final outcomes (and to a lesser extent the intermediate outputs) we intend to measure and track through the evaluation. Importantly, the so-called **primary substitution effect** (the increased use of blended KTR water in irrigation in the place of freshwater) is not and cannot be measured or shown as a single outcome. Rather, the quantification of this possible benefit stems from analysis that integrates several outcomes and outputs – to be carried out at the conclusion of the IE using data we proposed to collect – that flow through the following connections: a) reduced physical losses (WNP) and b) increased wastewater capture (WNP and WWNP); which lead to c) increased wastewater use in agriculture and d) substitution of King Talal Reservoir (KTR) water for King Abdullah Canal (KAC) water in the Jordan Valley; which together e) change per-capita use of utility water and lead to f) end-user time savings; g) consumer cost savings; h) aesthetic and health benefits; and i) are capitalized in land values. Similarly, understanding the net value of the **secondary substitution effect**, or the increased use of network water in place of tanker and/or vended water, flows through a complex chain that includes (not in order of importance), a) improved water quality at the point of delivery and b) changes in per capita use of utility water (due to the factors listed above as well as these quality improvements) which are embedded in reduced purchase of c) tanker water and d) vended water; both of which should ultimately appear as consumer e) cost and f) time savings, but may also result in reduced sales and/or profits in the water tanker and vended water industries. In addition, the extent of these primary and secondary substitution effects will likely be mediated (positively or negatively) by changes in utility performance, itself a function of the delivery of improved services.

Thus, we emphasize that measurement of these effects, as well as several others mentioned in the previous analyses of MCC Compact's economic feasibility (see Table 2.1 for a complete list), does not stem from any of the three individual data collection components described in this report, but rather from analysis and integration of their specific results. These components are:

1. Component 1: Household and enterprise surveys conducted in both intervention and control areas of the Zarqa/Ruseifa conurbation (and similar control areas selected from the Amman Governorate);
2. Component 2: Detailed longitudinal water balance analysis for the system spanning the Zarqa and Amman water and wastewater networks, the King Talal Reservoir, and the complex irrigation network of the Jordan Valley; as well as farm surveys in the Jordan Valley to estimate the magnitude and economic impacts of changes in the availability and utilization of water from different sources;
3. Component 3: Detailed monitoring of District Metering Area (DMA) and utility-level data on water delivery and wastewater collection in Zarqa, as well as indicators of financial and technical performance of the WAJ-Zarqa.

The present report focuses only on the sampling procedure and baseline data collected as part of Component 1. For details on the other components, readers can refer to the EDR.

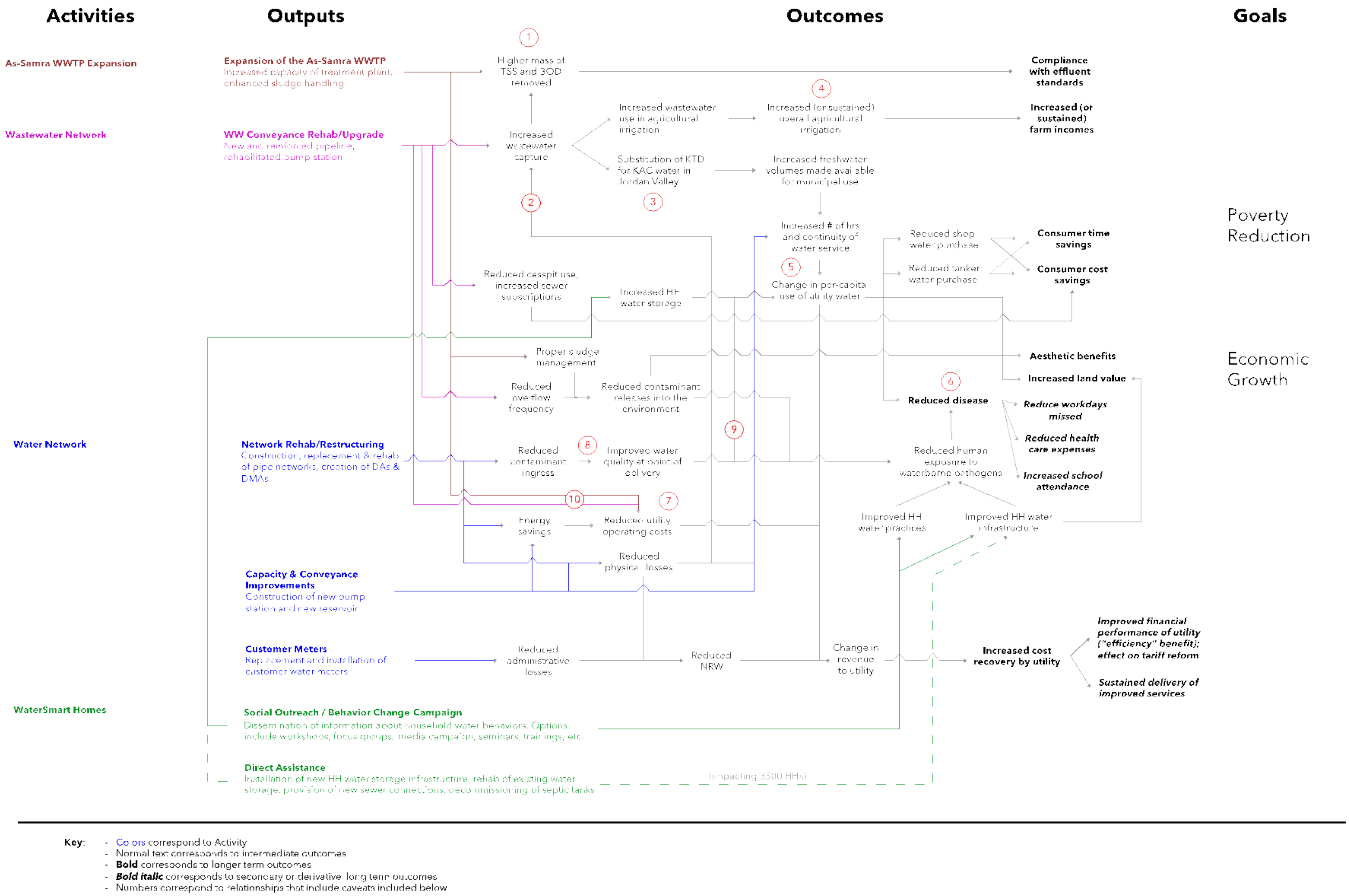

Figure 2.1. IE Logic Diagram.

Figure 2.2. Annotation List.

1. The As-Samra Facility expansion will enable removals of suspended matter and oxygen-demanding materials from increased volumes of wastewater that would not be treated in the absence of the expansion, as well as potentially facilitating the proper management of sewage sludge. In other words, it will not affect the volume of wastewater production from Zarqa, but it will ensure that increased effluent volumes will continue to meet internationally recognized wastewater treatment standards.
2. Wastewater volume increases will result from increased wastewater capture, a product of the wastewater network rehab/upgrade. It could also result from reductions in physical losses from the water network, assuming that the reduction of those losses lead in turn to increased municipal water usage.
3. Measuring the specific amount of replacement of freshwater by blended water (blended = treated effluent plus freshwater from the Zarqa watershed) in the Jordan Valley or elsewhere downstream of the As-Samra Plant will require careful construction of a water balance for the system.
4. Depending on the degree of substitution taking place, the amount of water used for irrigation downstream of the As-Samra facility may remain static or actually increase. Alternatively, freshwater allocation to farmers may decline at a rate higher than the increase in blended KTR water that is made available, in which case overall irrigation may actually decline, and the Compact benefit will be in slowing the decline of irrigation.
5. Changes in per-capita water use will be influenced by 1) increased # of hours and continuity of municipal water service, 2) improved water quality at the tap, if perceived by consumers, and 3) increased HH storage infrastructure resulting from WaterSmart Homes - though this will be only from a small number of homes. However, increases in usage could be modulated by increased metering, which will change household water use behavior. (We have not indicated this modulating factor in the diagram).
6. We include multiple possible causal relationships between Compact activities and disease. The first is the result of increasing per-capita water usage, and we emphasize that the relationship between water quantity interventions and health indicators such as diarrhea are not supported by the current literature. The second is by reducing disease transmission pathways resulting from urban wastewater overflows as well as those from land application of sewage sludge at As-Samra. Finally, the Water Smart Homes activity could result in reduced pathogen exposure via improvements in hygiene behaviors as well as reduced contamination in household storage.
7. We have not made a distinction here between overall energy savings for the utility and energy savings per unit volume of water delivered. We expect unit costs to decline, but overall system utilization - and thus, energy consumed, and operating cost incurred - could actually increase.
8. Though we have not seen significant data yet, we anticipate that the changes in water quality at the tap will be minimal, since there appear to be few documented instances of fecal contamination exceeding the WHO standard in the Zarqa system.
9. Improved water quality at the tap will result in increased per-capita use of utility water only if user perceptions of utility water improve in tandem. We note that the water quality benefits are likely to be difficult to detect (since pathogen detection in utility water is already so low), so a corresponding change in customer perceptions is also of low probability.
10. We note that the expansion of the As-Samra WWTP and the rehabilitation of the Wastewater Network may add to utility operating costs considerably, perhaps more than the associated increase in wastewater treatment revenues.

Table 2.1 Relationship between IE design components and the main expected economic benefits of the Compact (questions that are at least partly related to Component 1 are shaded in gray).

Economic impact question	Data collection components required
1. What is the economic value of increases in water consumption and reliability due to the intervention?	Components 1 (household/enterprise surveys) and 3 (utility monitoring)
2. What is the economic value of consumer savings from reduced vendor and tanker water consumption? (**secondary substitution effect**)	Component 1 (household/enterprise surveys; water vendor surveys)
3. What are the health benefits stemming from changes in water quality and consumption?	Component 1 (household surveys)
4. What is the value of avoided contamination of irrigated areas stemming from wastewater investments?	Component 2 (water balance, farm survey)
5. What are the net cost savings (in terms of expenditures on wastewater management) to consumers without sewerage of connecting to the wastewater network?	Component 1 (household/enterprise surveys)
6. What is the value of land reclaimed from septic / latrine for newly-connected wastewater network consumers?	Component 1 (household/enterprise surveys; land survey)
7. Are there utility cost savings from reduced maintenance of network infrastructure?	Component 3 (utility monitoring)
8. What is the economic value of substitution of additional blended KTR water for freshwater in irrigation? (**primary substitution effect**)	Components 1 (household/enterprise surveys), 2 (water balance, farm survey) and 3 (utility monitoring)
9. What is the value of new irrigation stemming from Compact investments?	Component 2 (water balance, farm survey and remote sensed data)
10. What is the value of citrus and other high value crops that are preserved due to increased water availability for irrigation?	Component 2 (water balance, farm survey and remote sensed data)
11. What are the time savings and productivity gains from improved urban water supply in Zarqa?	Component 1 (household/enterprise surveys)
12. What are the non-health aesthetic (quantity) benefits of improved urban water supply in Zarqa?	Component 1 (household/enterprise surveys)
13. What are the impacts on utility performance (namely cost recovery)?	Component 3 (utility monitoring)
14. Are there increases in property values in Zarqa separate from the value of reclaimed land?	Component 1 (household/enterprise surveys; land survey)

3 DESIGN OF IE COMPONENT 1: IMPACTS OF INFRASTRUCTURE IMPROVEMENTS ON OUTCOMES IN ZARQA

3.1 General description

The primary evaluation objective for the first IE component, which focuses on measurement of outcomes in Zarqa, is to determine how outcomes (Y_i^1) experienced by i individual and commercial/industrial sector enterprise units affected by the Compact's investments compare to what those individuals would have experienced investments not been made (Y_i^0). This latter counterfactual cannot be observed, and we require other methods for measuring it with a minimum of bias. Our evaluation strategy hinges on exploiting variation in the intensity of "exposure to treatment" to the Compact improvements. In particular, the IE aims to establish a causal relationship between program interventions and a variety of observable project-related social and economic outcomes, by comparing the changes experienced over time by beneficiaries (the treatment group) to those experienced by a similar set of non-beneficiaries (the control group).[4]

In simple terms, we can think of household and enterprise exposure to treatment as corresponding to the classes identified below in Table 3.1. The crux of an evaluation of the effects of these investments would be to compare areas in categories A (most intense treatment) with those in D (unaffected units), to detect combined impacts, and to compare areas in D with B, and D with C, to determine the separate impacts of the wastewater and water network investments, respectively. Our design exploited information on the layout of the different Compact projects to create such a sample.

Table 3.1. Definition of sample arms

Water / Wastewater	Treated	Control

[4] Because our overall design utilizes such a treatment and control strategy to identify impacts, we refer to it as an IE. However, in Table B.1 and elsewhere we use MCC's terminology to distinguish between two types of evaluation components that make up this overall design, impact and performance (per USAID's Evaluation Policy from January 2011), as follows. **Impact evaluation** is a study that measures the changes in income and/or other aspects of well-being that are attributable to a defined intervention. Impact evaluations require a credible and rigorously defined counterfactual, which estimates what would have happened to the beneficiaries absent the project. **Performance evaluation** is a study that seeks to answer descriptive questions, such as: what were the objectives of a particular project or program, what the project or program has achieved; how it has been implemented; how it is perceived and valued; whether expected results are occurring and are sustainable; and other questions that are pertinent to program design, management and operational decision making.

Treated	A. Both improvements	B. Wastewater network only
Control	C. Water network only	D. No improvements

3.2 Component 1 evaluation questions

The Jordan Compact aims to affect a number of key outcomes that contribute to reduced poverty and increased economic productivity in households and enterprises across Zarqa. The major evaluation questions related to these impacts are the following:

1. **Impacts on water consumption**: Does the WNP change the quantity of water consumed at the household (HH) and enterprise (E) levels (reduced leaks, increased reliability)?
2. **Impacts on environmental quality**: Does the WNP alter the quality of water consumed at the HH / E levels? Does the WWNP reduce the risk of disease from exposure to untreated wastewater?
3. **Impacts on expenditure**: Does the WNP affect time and money expenditure on water ('secondary' substitution effect)? Does the WWNP change consumer expenditure on wastewater management and disease prevention and treatment?
4. **Impacts on income**: Does the WNP change HH / E income?
5. **Impacts on asset value**: Does the WNP / WWNP affect property/asset values?
6. **Overall impacts on economic welfare**: What is the net economic value of changes in quantity and quality of water consumed?

Besides considering each of these evaluation questions, the final IE design was developed to study the mechanisms by which they are produced, through tracking as many intermediate impacts or contributing factors to them as possible (in collaboration with the M&E activities of the MCA-J), as shown in the IE logic. In addition, we aim to assess the distribution or incidence of these impacts on particular groups (e.g., within the household, to women or men, children or adults; or across households, to the bottom quartile of the income distribution or to the upper 3 quartiles) as far as possible, noting that statistical power may be limiting for detection of heterogeneous impacts.

3.3 Sample design methodology

Due to the non-random nature of the intervention, reliance on the use of a representative sampling procedure in Zarqa would force the evaluation to rely on ex-post control for systematic differences across treated and control areas, and therefore create a substantial risk of bias in estimation impacts, and loss of statistical power. Instead, the IE attempts to reduce the threats of confounding due to differential selection into treatment and control zones by employing ex-ante Propensity Score Matching (PSM) to identify pairs of intervention and control areas matched on a set of observable pre-intervention characteristics (Rosenbaum and Rubin 1983). Then, at the time of final analysis, we will employ difference-in-differences (DiD) to reduce the threats posed by unobservable differences between affected units that do not vary over time. Finally, regression analysis will further allow us to control for observed factors other than treatment status that are correlated with outcomes, thereby improving the precision of treatment estimates as well as indicating whether the quasi-experimental control achieved by the matching approach was successful (and adjusting them to the extent possible).

3.3.1 Selection of sample areas for household and enterprise surveys: Propensity Score Matching (PSM)

To select the zones for the household and enterprise surveys, we implemented ordinary propensity score matching (PSM) *ex ante*, i.e. prior to data collection. This procedure first entailed predicting the selection into the various treatment groups using logistic regression, where treatment status – defined in our case using GIS maps of the investments in the Jordan Compact – is regressed on a vector of observed pre-intervention characteristics *X* of those areas:

$$T_i = \beta X_i + \varepsilon_i, \quad (1)$$

where $T_i = 1$ if unit *i* is assigned to treatment and 0 otherwise, and ε_i is an error term. We conducted PSMat the block level (the unit used by the Department of Statistics for Census sampling), using 11 variables recorded or calculated from the Census. In an effort to address the issue of spillovers, we included untreated areas both in and outside of the Zarqa water and wastewater network, by running these regressions including nearby neighborhoods from Amman. For the classification of areas into the groups A-D (as described in Table 3.1), we merged shapefiles indicating the boundaries of blocks (obtained from DoS) with shapefiles showing the locations of the WNP and WWNP infrastructure works (obtained by project implementers and the MCA-J). Blocks treated by each project or by both projects were then identified using GIS functions that identify intersections of lines (for pipe works) and polygons (Census blocks). The pre-intervention data used for this first stage were then drawn from sources that were: a) available for our purposes; and b) derived from representative samples drawn at fairly fine geographical scale – namely block-level Census, income and expenditure survey data from Jordan's Department of Statistics (DoS). The asset lists available from DoS data sources were used to create a wealth index using principal components analysis, retaining the first principal component.

The results for the first stage logit are shown in Table 3.2. These regressions indicate that areas selected to receive the water network improvements tend to have fewer buildings but more households per block (i.e., higher density) than areas not receiving any Compact improvements in Zarqa, and slightly higher wealth. (Column 1). Households in these WNP areas tend to have lived at their current residence for fewer years, are more likely to be Jordanian, and are less likely to have married household heads. The areas selected for WWNP improvements in Zarqa have the opposite characteristics: lower density, more non-Jordanians, older and lower wealth residents, and a higher proportion of married household heads (Column 2). Finally, areas selected for both WWNP and WNP improvements tend to look more like WWNP areas than WNP areas (Column 3). Finally, compared to untreated zones in Amman, all three areas in Zarqa tend to be lower density, higher percentages of married household heads, and have lower wealth (Column 4-6).

Following logit estimation, we obtained the propensity score (or predicted probability of participation) for each geographic unit (Census block) in the sample:

$$p(x) = \Pr[T = 1 | X = x] = \frac{e^{\beta x}}{1+e^{\beta x}} \quad (2)$$

We then matched areas with similar propensity scores (i.e., blocks that appeared equally likely to have received specific exposures to the intervention, conditional on these observable factors) to ensure comparability across control and treated areas. We used 1-1 nearest neighbor matching with replacement and imposed a caliper of 0.02 around every match. For example, if a treatment block had a propensity score of 0.64, it could only be matched with a counterfactual block with a propensity score between 0.62 and 0.66, and the nearest match satisfying that constraint would be assigned. In most cases the nearest neighbor was retained as the match, but in a few cases the caliper rule superseded the nearest neighbor rule in order to ensure a good match for each treatment block.[5]

Finally, to verify the comparability of the matched samples, we conducted balance tests and checked the overlap in support between treatment and control groups (Tables 3.3 and 3.4 summarize the results, showing the balance across the different types of Census blocks before and after matching, respectively). The final map of sample areas located in Zarqa is shown in Figure 3.1; we do not show the map for Amman given the difficulty of reading the map on a broader geographic scale.

There are three principle threats to the validity of estimates obtained using PSM in this way (Rosenbaum and Rubin 1983). The first is that unobserved differences between treatment and comparison may lead to biased estimates of impact when these differences are correlated with treatment outcomes (violation of the so-called Conditional Independence Assumption). Such unobserved differences may encompass for example preferences among decision-makers for a particular zone that is not reflected in the formal prioritization algorithm, or systematic differences in the preferences for improved water supply among beneficiaries of different zones. Unfortunately, we could not fully address this threat, because the parameters used in the prioritization algorithm that was used to select treatment areas for the water network improvements were only available for the treatment zones themselves, and the criteria for selection of expansion of areas treated by the wastewater network expansion were not fully transparent. To the extent that such factors were accounted for in the observable (primarily socio-economic and demographic) characteristics of the affected zones, these would be reflected in our approach.

The second threat emerges when the common support region is narrow such that the universe of treated and control areas are difficult to compare (Dehejia and Wahba 2002). This was not a major problem in our case, and we had a sufficient number of matches from which to choose. Finally, the third important threat, which is more generally applicable to a variety of estimators for IE, emerges from violation of the Stable Unit Treatment Value Assumption (SUTVA), which requires that treatment does not indirectly affect untreated units (i.e. no spillovers). Our strategy of sampling from neighboring blocks in Amman was utilized specifically to deal with this issue.

[5] In a few cases, SI also conducted manual matches to reduce the number of control blocks required in our sample. SI wanted to minimize the pool of control blocks so as to not threaten the statistical power of the evaluation. For example, if a treatment block had a propensity score of 0.3456 and the nearest neighbor match was 0.3457 and a control block matched to another treatment block from a different treatment arm had a propensity score of 0.3460, we would manually choose the latter block to match with the current treatment block. The rationale for this will be discussed in more detail further below.

Table 3.2. Results of first-stage logit model for matching

Variable	ZARQA CONTROLS						AMMAN CONTROLS					
	Area A Treat WNP		Area B Treat WWNP		Area C Treat Both		Area A Treat WNP		Area B Treat WWNP		Area C Treat Both	
Population density ('000/hA)	0.00	(0.000)	-0.00***	(0.000)	-0.00***	(0.001)	0.00***	(0.000)	-0.00***	(0.001)	-0.00***	(0.001)
# buildings in block	-0.01***	(0.003)	0.03***	(0.007)	0.02***	(0.007)	0.02***	(0.003)	0.02***	(0.006)	0.02***	(0.006)
# households in block ('00)	0.01***	(0.002)	-0.00	(0.004)	-0.01***	(0.004)	-0.01***	(0.001)	-0.00	(0.003)	-0.01**	(0.003)
Male head of household (%)	-0.48	(0.842)	3.03	(2.019)	-3.29	(2.037)	-2.04**	(0.883)	3.27*	(1.879)	-0.11	(1.808)
Handicap (%)	1.41**	(0.691)	-2.12	(1.736)	-0.44	(1.759)	2.53***	(0.870)	1.09	(1.674)	-0.85	(1.525)
Non-Jordanian (%)	-2.51***	(0.491)	2.93***	(1.089)	4.07***	(1.283)	-3.68***	(0.636)	1.57*	(0.846)	1.66	(1.255)
Average residency (yrs.)	-0.02**	(0.010)	0.07***	(0.024)	-0.08***	(0.021)	0.11***	(0.010)	0.09***	(0.022)	-0.02	(0.019)
Head > Secondary educ. (%)	0.40	(0.312)	-1.43**	(0.683)	0.82	(0.641)	1.29***	(0.334)	-0.26	(0.664)	0.60	(0.590)
Marital status – head (%)	-2.77***	(0.730)	4.47**	(1.906)	5.01***	(1.762)	2.11***	(0.792)	2.46	(1.669)	1.81	(1.581)
Paid employee – head (%)	0.26	(0.319)	-1.64**	(0.714)	-1.02	(0.669)	0.36	(0.336)	-0.26	(0.686)	-0.13	(0.585)
Wealth index	0.14***	(0.043)	-0.25**	(0.102)	0.06	(0.085)	-0.23***	(0.041)	-0.20**	(0.094)	-0.04	(0.079)
Constant	2.48***	(0.749)	-7.42***	(1.949)	-0.44	(1.764)	-3.00***	(0.771)	-7.64***	(1.808)	-1.77	(1.566)
N	1822		623		612		1907		542		531	
Pseudo-R^2	0.073		0.385		0.318		0.226		0.293		0.193	

*Notes: Coefficients and standard errors (in parentheses) are reported. Statistically meaningful differences are indicated by the following: *** indicates p<0.01; ** p<0.05; *; p<0.1. The sample size varies, because units in other treatment groups are omitted from the regressions used for matching the group in question. The wealth index is expressed as the first principal component derived from principal components analysis of the following list of assets: washing machine, solar heater, microwave, private car, mobile phone, computer, and an internet connection.*

Table 3.3. Pre-matching descriptive statistics for Census blocks

	COMPARISON WITH ZARQA CONTROLS				**COMPARISON WITH AMMAN CONTROLS**			
Variable	Area A Both (N=104)	Area B WWNP only (N=115)	Area C WNP only (N=524)	Area D Controls (N=1303)	Area A Both (N=104)	Area B WWNP only (N=115)	Area C WNP only (N=524)	Area D Controls (N=1386)
Wealth index	-0.54***	-1.13	-0.77***	-1.21	-0.54***	-1.13***	-0.77***	0.37
Marital status – head (%)	91.0***	90.8***	87.2	88.2	91.0***	90.8***	87.2	87.7
Male head of household (%)	91.6***	92.4***	89.3***	90.3	91.6***	92.4***	89.3	89.8
Head > Secondary educ. (%)	45.3***	36.8	42.8***	38.1	45.3***	36.8***	42.8***	53.4
Average residency (yrs.)	14.2***	16.7	16.7**	16.2	14.2***	16.7***	16.7***	13.0
Non-Jordanian (%)	6.2*	7.7	4.9***	8.4	6.2	7.7	4.9***	7.6
# buildings in block	39.0	49.1***	34.3***	39.5	39.0***	49.1***	34.3***	30.6
Population density (per hA)	66.6***	72.2***	266.1**	238.4	66.6***	72.2***	266.1***	177.4
Paid employee – head (%)	78.6*	78.6*	79.7	80.6	78.6*	78.6*	79.7***	76.5
# households in block	70.6***	89.8*	85.3	83.1	70.6***	89.8	85.3***	92.3
Handicap (%)	5.6	5.6	6.2	5.9	5.6**	5.6**	6.2***	4.8

*Notes: Statistically meaningful differences are indicated by the following: *** indicates p<0.01; ** p<0.05; *; p<0.1.*

Table 3.4. Post-matching descriptive statistics for selected Census blocks

	COMPARISON WITH ZARQA CONTROLS						COMPARISON WITH AMMAN CONTROLS					
Variable	Area A Both	Area A Controls	Area B WWNP only	Area B Controls	Area C WNP only	Area C Controls	Area A Both	Area A Controls	Area B WWNP only	Area B Controls	Area C WNP only	Area C Controls
Wealth index	-0.25	-0.66	-0.94	-1.04	-1.08	-1.09	-0.25	-0.66	-0.94	-1.04	-1.08	-1.09
Marital status – head (%)	89.1	89.3	89.5	87.7	88.4	88.3	89.1	89.3	89.5	87.7	88.4	88.3
Male head of household (%)	90.1	89.8	90.1	90.3	90.2	90.1	90.1	89.8	90.1	90.3	90.2	90.1
Head > Secondary educ. (%)	51.4	47.2	40.0	38.3	39.3	38.6	51.4	47.2	40.0	38.3	39.3	38.6
Average residency (yrs.)	15.9	15.9	16.7	17.2	16.3	16.7	15.9	15.9	16.7	17.2	16.3	16.7
Non-Jordanian (%)	4.1	4.3	3.7	4.7	5.1	5.0	4.1	4.3	3.7	4.7	5.1	5.0
# buildings in block	35.1	37.6	38.1**	45.6	36.1	36.0	35.1	37.6	38.1**	45.6	36.1	36.0
Population density (per hA)	98.4	118.2	113.5	160.2	278.6	251.7	98.4	118.2	113.5	160.2	278.6	251.7
Paid employee – head (%)	80.3	77.8	81.5	81.4	80.9	80.3	80.3	77.8	81.5	81.4	80.9	80.3
# households in block	79.3	77.0	83.7*	96.2	81.6	83.6	79.3	77.0	83.7*	96.2	81.6	83.6
Handicap (%)	4.5	5.2	5.7	6.7	6.2	6.2	4.5	5.2	5.7	6.7	6.2	6.2

*Notes: Statistically meaningful differences are indicated by the following: *** indicates p<0.01; ** p<0.05; *; p<0.1. Matching was conducted using 1-1 nearest neighbor matching with replacement. The first stage was specified as a logit model using all of the characteristics shown above.*

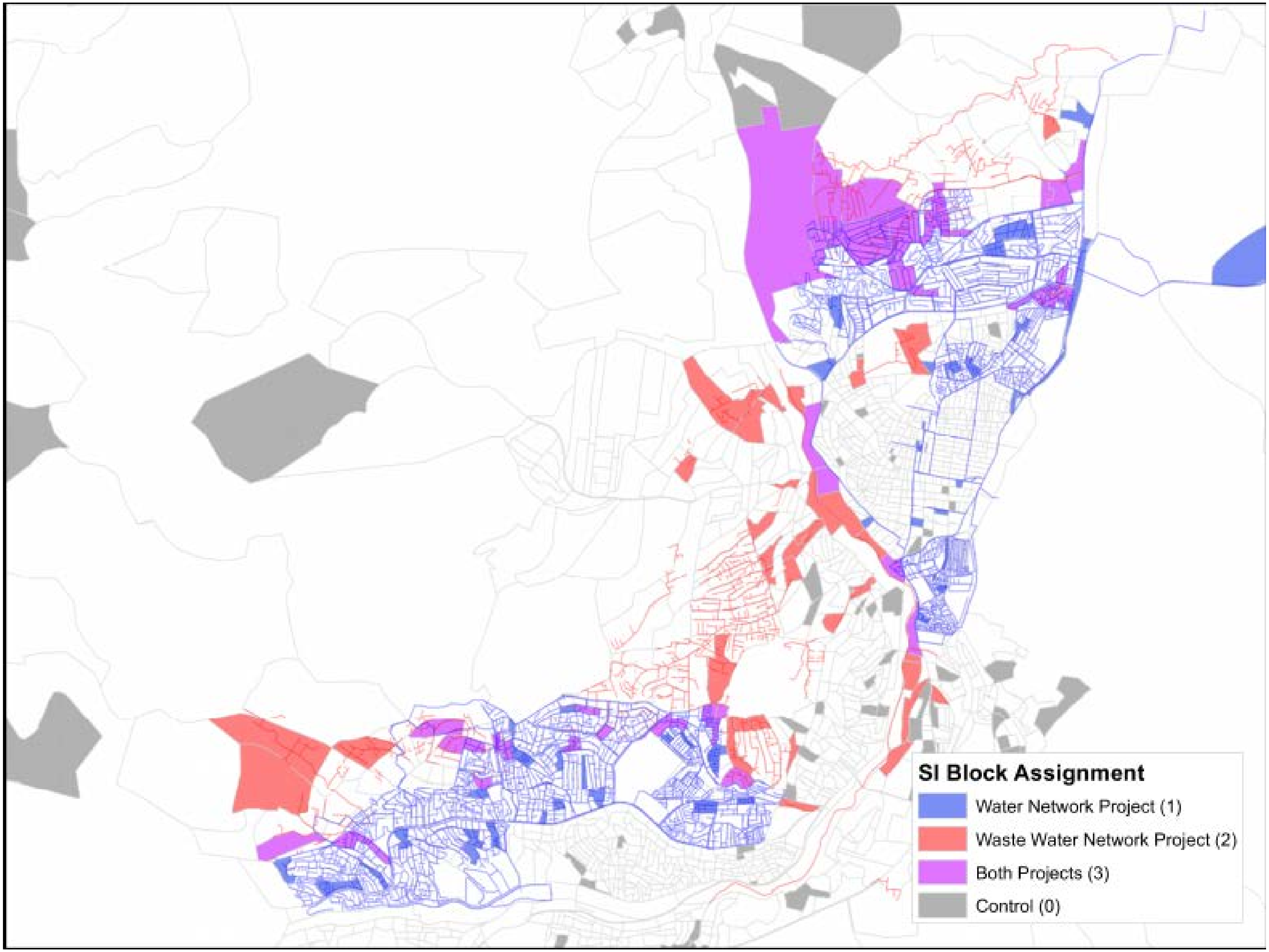

Figure 3.1. Map of treatment and control blocks and infrastructure works, with rehabilitated water pipes in blue and new wastewater networks shown in red (Note that some areas are off the map and therefore not shown, e.g. all controls in Amman, and some in Zarqa).

3.3.2 Sample construction for the household survey

In the EDR, it was estimated that detecting 10% effects of the Compact's infrastructure improvements on urban households would require a sample size of 3,440 households divided equally across all 5 arms (WNP, WWNP, both, Zarqa control and Amman control) for six important outcomes: i) hours of water supply; ii) water consumption; iii) water bills; iv) spending on and v) quantity purchased of shop water; and vi) total monthly expenditures on water. Detectable differences in treating water from the public network as well as self-reported health expenditures because of water consumption would be less likely with a sample size of 3,440. Reasonable power to detect such differences would require sample sizes beyond the capability of this evaluation. The same is true for smaller changes of a magnitude of around 5%.

Assuming groups of equal sizes, a sample size of 2,500 was first calculated to be sufficient for detecting 10% differences across treatment and control groups if the sample were comprised of 4 groups (roughly 625 households per group). However, in consideration of the likelihood of Zarqa-wide spillovers, the final sample size was increased to include an additional group from peri-urban zones in Amman (roughly 625 households). Finally, the evaluation sample size target included a 10% buffer for attrition over the life of the study; 313 households were thus added to the final sample, such that approximately 3,440 randomly-selected households were requested of DoS at baseline. Note that the additional sample size to protect against attrition does not represent an acceptable level of attrition, but was purely included to maintain statistical power over the outcomes of interest, and, importantly for internal validity, must be balanced across sample groups.

As noted above and confirmed by the calculations shown in Table 3.3, a random selection of households from blocks classified into the five sample categories would clearly have led to a sample plagued by significant differences across groups. We therefore constructed the sample by asking the DoS to implement a random sampling method at the block level to spread the sample evenly across all blocks retained using the PSM procedure described above, in other words selecting 11 households from each of 325 blocks (or 3575 households in all).

We note here, however, that one problematic aspect of the matching procedure described in the previous section was that finding "high quality" matches for treatment blocks required a somewhat larger number of control blocks being selected compared to the number of blocks retained in each of the three treatment arms. This was due to the systematic differences between WNP areas and WWNP areas, which required sampling of different controls that were matched specifically to each of these distinct groups. To minimize the loss of statistical power, we manually retained control blocks that matched well with blocks in several treatment arms. Even so, the need for extra control units created imbalances in the target sample sizes for each arm, as summarized in Table 3.5. More control households were especially needed from Amman, given that the quality of matches between treated Zarqa blocks and control areas in Amman was lower on average than with control areas in Zarqa. We came to the final target allocation of 3,465 households shown in Table 3.5 knowing that the imbalances and departure from sample sizes of 687.5 per arm would lead to some loss of statistical power for the cross-arm comparisons. Specifically, differences on several of the indicators described above would have to be closer to 15% to find clear differences between *specific* groups and their specific controls. On the other hand, differences across the *pooled* samples of treated vs. control units would remain detectable at well below the original levels.

Table 3.5. Proposed and targeted distribution of households across the five sample arms

	Proposed distribution of households (from the EDR)			*Final target distribution of households*	
	N	Attrition	Total	Blocks	Households
Zarqa Wastewater	625	62.5	687.5	43	473
Zarqa Water	625	62.5	687.5	49	539
Zarqa Both	625	62.5	687.5	43	473
Zarqa Control	625	62.5	687.5	82	902
Amman Control	625	62.5	687.5	108	1188
Total	3125	312.5	3440	325	3575

Notes: The final target sample sizes for each arm are not evenly distributed, due to the need for additional controls that match well to the different types of treated areas (see discussion above). High quality matches for treated areas in Zarqa were more difficult to identify from Amman; hence that sub-sample is particularly large.

3.3.3 Sample construction for the enterprise survey

For the enterprise survey, it was originally proposed, admittedly without any statistical rationale, that a minimum sample of about 275 enterprises be selected using snowball sampling methods, acknowledging that sample size would be reassessed following baseline. Since there were no data available to conduct power calculations and there are no known rigorous evaluations that assess the impacts of water infrastructure on enterprises, this proposed sample size was mainly for the purpose of gathering information without incurring excessive costs. It was suggested that power calculations with very transparent assumptions about (unknown) effect sizes could be conducted following baseline, which we do towards the end of this report.

Given this objective, SI first randomly drew 175 formal enterprises from referrals provided as part of the household survey, stratified across the sample treatment arms. SI also randomly drew 175 formal and 25 informal enterprises from a 2011 DoS-administered national enterprise survey and from the sampling frame for a recent labor and unemployment survey, all of which were drawn from the same census blocks as the household survey sample. Fifty additional enterprises were selected as replacements from the DoS enterprise listing, making a full sample of 425 eligible enterprises split across the above treatment and control arms. These replacements were included based on the experience of the household survey, which revealed that DoS only completed surveys with 3359 households (lower than the original target number of 3440).

DoS and SI also faced additional challenges in tracking down the household business referrals (both formal and informal); these were found to often a) be missing critical identifying and address information, b) be duplicates, or c) presented other problems that made locating them difficult. For this reason, the teams eventually fell back on a DoS listing of formal and informal enterprises conducted in 2011 to supplement the household referral sample. Within the final DoS listing of enterprises, only 25 were ultimately categorized as informal businesses. We chose to sample all 25 of these informal enterprises in an effort to maximize inclusion of informal business activity in our study, but note that the small number of informal enterprises will render it difficult to come to statistical conclusions about these types of businesses.

3.3.4 Development and testing of survey instruments

Once the sampling frame for the IE is specified and baseline and post-intervention data are collected, the major part of our evaluation analysis will be conducted by analyzing data on intermediate outcomes and welfare measures obtained from two separate panel surveys – required to carry out the DiD analysis – of households and enterprises within the sample zones identified by the PSM procedure. In this conceptualization, intermediate outcomes are physical or behavioral changes that can be theoretically linked to eventual changes in social welfare, as depicted in the IE logic diagram (Figure 2.1). Collecting and analyzing these data will produce a fine-grained understanding of the mechanisms that have or have not led to real changes in well-being. Both the household and the enterprise survey use panel survey designs to measure outcomes before and after the MCC interventions.

Household surveys. Household-level surveys were developed to collect information on household demographics; water sourcing (including network, tanker and shop water), pumping, storage, and use behaviors; preferences and satisfaction with water supply and sewer service; water quality measured at the tap and in in-house storage containers (chlorine residual, salinity, turbidity, and *E. coli* or *thermo-tolerant coliform* counts); coping and health costs related to intermittent water supply and poor water quality; and expenditures (as recorded in water bills, as well as on other household items), income, and other socio-economic characteristics. The household survey instrument (see Annex A) included 13 modules and took approximately 1.5 hours to complete.

Enterprise surveys. The enterprise surveys focused on enterprise characteristics, production inputs and outputs, costs and revenues, and assess constraints with regards to using water as an input to production. In addition, for assessing impacts on Zarqa's important informal sector, we rely on the informal production activities carried out by households selected into our sample, supplemented by a sampling strategy that selects enterprises from two groups: a) a snowball approach that begins with referrals by sample households to "businesses" (both formal and informal) that they use in their neighborhood; and b) sampling from the overall sampling frame for enterprises included in DoS enterprise surveys. The instrument for the enterprise survey (see Annex B) had 10 modules and took approximately 1 hour to complete.

The household and enterprise survey instruments, developed based on well-tested existing instruments previously applied by members of the SI team in studies in other countries, underwent forward and backward translation to ensure the accuracy and precision of survey language. Challenging and additional questions were thoroughly piloted in focus groups with men and women, and through training activities with enumerators. Finally, pre-tests were conducted for both surveys in non-sample areas of Zarqa prior to launch of the survey. Both surveys employed computer-assisted personal interviewing (CAPI) using tablets.

3.3.5 Risks and mitigation strategies

This section describes a variety of risks associated with the evaluation activities of Component 1 that were also discussed in the EDR. We summarize the most important risks (Table 3.6), and discuss them in greater detail below, focusing on our strategies for mitigating them.

Violation of CIA assumption of PSM. A major risk to the IE of the water and wastewater network improvements is related to the assumption that the expected outcomes of treatment and comparison groups are independent of the treatment assignment, conditional on control for baseline observable characteristics through use of PSM. There are two distinct concerns. First, given that a range of variables that were specified with varying degrees of subjectivity (for example preferences of decision makers in WAJ-Zarqa vs. population density measures) were used in prioritizing areas for water network improvements, there may be concerns about the statistical models we used for our sample construction. We did include as many factors as possible in the estimation of propensity scores, to limit the nature of such threats. In the results section of this report, we discuss how analysis of the baseline data does nonetheless reveal slightly greater imbalances that we would expect if the CIA assumption were fully valid. This makes the use of reliable longitudinal data for DiD analysis all the more critical.

The second concern relates to the overlap between treated and control units. As can be seen in Figure 3.1, there is significant spatial clustering of WNP, WWNP, and control areas. To the extent that outcomes are highly correlated within such clusters, statistical power will be diminished. For example, households and enterprises in treated neighborhoods that are located near one another will also experience similar water reliability (e.g., hours of supply per week), while those further away will tend to experience different water supply conditions. These spatial correlation patterns will decrease the amount of variation in the data that is useable for inference about differences between near and far away areas. In addition, systematic differences related to different locations may compromise our ability to make valid comparisons. Related to the latter is the fact that our matching algorithm relies on data sources that are somewhat dated (from the 2004 Census).[6] To best preserve flexibility to manage such threats, our power calculations at baseline were deliberately conservative, in order to allow us to drop poor matches *ex post* of the data collection activities (i.e., at the time of analysis), or increase the precision of our estimates. As we will discuss further below, however, the baseline survey data raise new concerns about power.

Lack of statistical power. Another distinct concern pertains to the potential lack of statistical power to reliably measure impacts of the MCC investments (pre-baseline power calculations can be found in the EDR). Many of the micro-level (household and enterprise) outcomes of interest are heterogeneous across units and over time. Some of the variables that are subject to change are also likely to be uncommon in our sample. Developing a power analysis for an IE with such a wide range of impacts as this one is a significant challenge, and we limited ourselves to conservative assumptions related to the key outcomes for which we had pre-intervention data (i.e. household-level outcomes such as expenditures on non-network water, and

[6] One major reason to believe that census blocks today do not necessarily represent the same census blocks from 2004 is due to population growth and changes in refugee populations in Zarqa and Amman governorates. In 2004 the majority of refugees in Jordan were Iraqi. By 2014 there was a large number of Syrian refugees, who may or may not have settled in ways similar to the Iraqi refugees.

hours of water supply delivered to households). It should not be surprising that the analyses of some of the outcomes we will aim to measure will be inconclusive.

Outcomes for which power calculations were not possible prior to baseline data collection given the non-existence of data were: Enterprise-level outcomes (formal and informal sector); sewer overflow hazard rates; water quality violations in storage containers (bacterial or other); household income and expenditure; vendor/tanker water sales, costs and revenues, and land values. In addition, at the time of planning, some of the outcomes for which much greater sample sizes seemed necessary were: Customer complaints or sewer overflow hazards; diarrheal disease prevalence; and water quality violations at the tap (bacterial or other).

Table 3.6. Categorization of threats to identification of impacts, and mitigation strategy

Type of risk	Description	Mitigation strategy
Violation of CIA assumption of PSM	Imperfect control for factors that affect assignment into treatment	-Conduct balance tests at baseline, re-assess and discuss balance in light of 2015 Census data -Test for systematic differences following baseline -Oversample at baseline to allow for *ex post* adjustments using 2015 Census data -Obtain better controls by leveraging differences in timing of exposure to treatment (using seasonal or high frequency data collection)
Lack of statistical power	Some/all outcomes are too small to be detected given sample size specified in the IE	-Use conservative assumptions in power calculations -Specify upfront the types of changes in intermediate or other outcomes that are unlikely to be detectable (e.g., changes in diarrheal disease)
Planned time horizon of IE	Impacts occur over a longer time horizon than the 5-year (planned) IE period	-Measure intermediate outcomes to obtain understanding of potential change mechanisms -Build local capacity for continuing IE beyond 5 years -Encourage MCC to support longer-term IE
Confounding	Outcome variables may be affected by time-varying factors that are not related to treatment (e.g. Disi water)	-Statistical control for confounders using DiD -Integration with other IE components -Measure intermediate outcomes to obtain understanding of change mechanisms
Spillovers / general equilibrium impacts	Control units are affected by treatment	-Incorporate a second control group from peri-urban Amman -Control for proximity/intensity of exposure to improvements, or implement *ex post* GPSM using additional hours of water supply and increased sewer connections as the "treatment" variables
Other important considerations	1) Attrition in sample 2) Limited sample size restricts ability to detect treatment effect	1) Power calculations allow for 15% loss to follow-up 2) Measures in multiple sectors; main effect heterogeneity of concern is on poor and can motivate re-assessment of sample

	heterogeneity 3) Confounding of effect of infrastructure improvement by O&M behavior of WAJ-Zarqa	sizes post baseline. 3) Conduct semi-structure interviews and focus groups with WAJ-Zarqa officials to determine if treated areas receive more attention in maintenance

Short (planned) time horizon of IE. As currently contracted, the IE will end in the fall of 2017. Water infrastructure projects are designed to deliver services and associated benefits for 20 to 50 years, not 5 years, so some of the most important social welfare benefits – effects on income, employment, etc. – of these investments may not be felt by that time (Whittington et al. 2008). However, many of the intermediate outcomes of the investments are likely to begin much earlier. Changes such as hours of supply, consumption of water, savings on expenditures, new connections to the sewer network, and perhaps property values (which capitalize a time series of benefits) will likely change during the currently-scheduled time horizon for the IE (North and Griffin 1993, Yusuf and Koundouri 2005). In addition to this, the inclusion of a second control group from peri-urban Amman to allow testing for spillover effects also offers more flexibility for tracking long-term differences in impacts in Zarqa relative to other locations in Jordan that did not receive MCC investment (using DiD methods as in Galiani et al. (2005)). We understand that MCC may continue supporting monitoring and evaluation activities beyond the currently planned IE duration of 5 years, and encourage such planning.

Confounding. Besides violations of the CIA assumptions, another systematic source of bias in our estimates of impact could emerge from confounding by time-varying factors affecting treated and untreated comparison units differentially. One obvious potential confounder is Disi water, to the extent that such additional water reaches treated and untreated areas in different amounts. There are now direct connections between this new supply and the Zarqa system, and the supply of Disi water to Amman indirectly affects Zarqa by reducing Amman's demand for other sources that currently serve both Zarqa and Amman. Whether these changes are increasing supply to all areas of Zarqa at the same time, and by the same amount, is currently unknown. There may be other changes from similar water supply projects in the quantity of water supplied to treated and untreated areas. Another major confounder is the restructuring of WAJ-Zarqa under a management contract with Miyahuna. This and other institutional changes to water utilities and water sector governance, as promoted through USAID's Institutional Support and Strengthening Program (ISSP), challenge the ability of the IE to isolate the effects of the Compact investments.[7]

Our approach for dealing with these kinds of confounders will utilize three strategies: 1) statistical control (including such time-varying factors explicitly in the DiD estimation through the term X_{ijt} in equation 3 above); 2) integration of the results obtained through the Zarqa-based surveys with those from the other data collection components (especially Component 3) to assess the relative contribution of Disi to Zarqa's water balance over time – as discussed further later in this chapter; and 3) careful measurement and

[7] We are aware that some make the argument that the institutional changes were partly attributable to the Compact, but this claim cannot be rigorously tested, since the Institutional Strengthening Program sponsored by USAID was not contingent on the MCC/MCA-J Compact.

attribution of changes in intermediary variables hypothesized to lead to the outcomes of compact investments. The latter will provide evidence on the mechanisms of change: modified intermediary variables that can be more convincingly linked to compact investments (for example changes in NRW in treated zones rather than increased water supply overall, as would occur due to water augmentation from Disi) will provide support that the changes are in fact attributable to MCC program investments.

Spillovers / general equilibrium impacts. Spillovers are a third important source of potential bias in treatment estimates. If these are positive (this is the more likely case through which untreated areas benefit from investments), then estimates of impact will be biased downward. On the other hand, if these are negative, then estimates of impact will be biased upward. In addition to these biases in detected impacts, the spillovers themselves will have been ignored by the IE. Building on the strategies for dealing with time-varying confounding, our approach for dealing with these kinds of confounders hinges on two specific adjustments: 1) sampling in zones more distant from the target areas for the Compact, notably neighborhoods adding a control group from peri-urban Amman that is close to Zarqa but governed by a different water utility and infrastructure system; and if necessary, 2) use of *ex post* generalized propensity score matching (GPSM) (Hirano and Imbens 2004) to consider intensity of the water network treatment based on the additional hours of water supply or changes in the percentage of sewer connections in various survey blocks. The intuition here is that households in different areas will be differentially exposed to improvements in water supply and wastewater service, and that these differences should be correlated with differences in the outcomes of interest. While the former approach is subject to additional concerns over differential confounding and lack of comparability between treated and untreated zones due to unobservables, our reliance on DiD and measurement of a rich set of covariates in sample areas will allow mitigation of such threats. The latter approach may help to link hours of supply or additional wastewater connections more convincingly to benefits such that a more complete picture of impacts can be obtained.

3.3.6 Timeframe and implementation of Component 1

Originally planned for spring 2013, SI experienced significant delays in its baseline data collection activity schedule due to a number of complications associated with organizing the fieldwork, including prolonged and fruitful discussions with project partners over the details of the final evaluation design, and challenges in finalizing survey instruments and sampling protocols given the realities of conduction such fieldwork in Jordan. Ultimately, the baseline household survey was conducted between April and May 2014, and the baseline enterprise survey was conducted between November 2014 and January 2015. Though the wastewater network project construction had already begun by early 2013, no new households had been connected at the time of baseline data collection activities. Water network rehabilitation construction began in the fall of 2013; though new connections and the physical effects will not be felt for some time still, it is possible that the commencement of network construction could have brought about specific behavioral responses among households and enterprises (in particular those related to asset accumulation and investment decisions of firms, households, and the water vending industry) in anticipation of new infrastructure investments. Given that awareness-raising activities were ongoing and actual works projects were underway in some areas prior to baseline survey activities, it is likely that expectations have already begun to affect behavior, which can potentially lead to an underestimation of the economic impacts of the Compact.

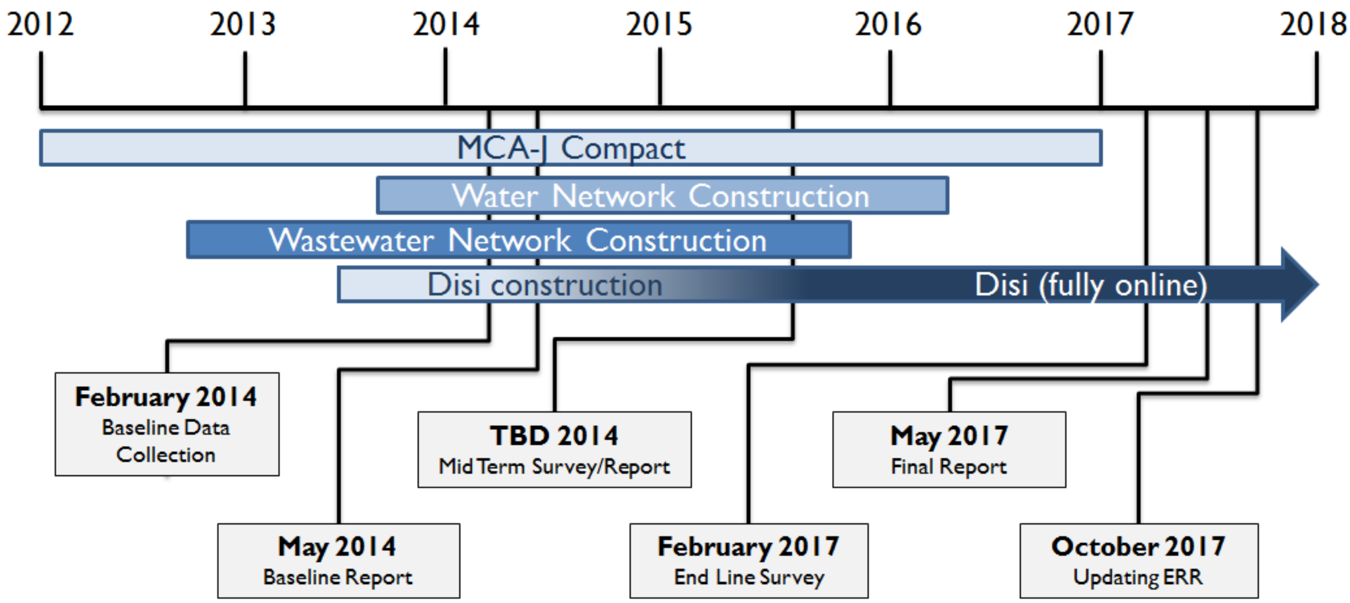

Figure 3.2. Water / Wastewater Network Projects and Evaluation Component 1 Timeline

4 DATA SOURCES AND OUTCOME DEFINITIONS

The primary data sources discussed in this report are the Component 1 Baseline Household Survey and Baseline Enterprise Survey, conducted by the Jordanian Department of Statistics (DoS) on behalf of Social Impact (SI). These surveys serve as the baseline for the longitudinal panel study of households and enterprises in designated treatment and control areas of Zarqa (with an additional control group in Amman). The household survey was programmed by SI, using Open Data Kit (ODK) electronic software. The enterprise survey was programmed directly by DoS using Java software.

Table 4.1. Data Sources

Data source	Population of interest and unit of analysis	Temporality (cross-sectional, longitudinal, panel)	Coverage (representative, non-representative,)	Type (Qualitative, quantitative)	Target Sample size
Household survey	All households in treatment and control areas	Panel (only baseline collected so far)	Non-representative, but subject to possible reweighting	Quantitative	3,465 households
Enterprise survey	All enterprises in treatment and control areas	Panel (only baseline collected so far)	Non-representative, but subject to possible reweighting	Quantitative	425 enterprises

Baseline data collection for the household survey was conducted between April and May 2014. Ultimately, DoS reported that enumerators visited a total of 3,655 households from the sample (1198 in Amman and 2457 in Zarqa) (Table 2). Questionnaires were recorded as complete by DoS for 95% of the total sample; 1.2% of households were vacant at the time of visit, while 0.5% of households were closed, and 2.8% of households refused to complete the survey. There were inconsistencies between the final DoS completion rates, and the completion rates in the final data set provided to and approved by SI (see Table 4.2). Specifically, the DoS reporting rates for "closed", "empty" or "refusing" households do not fully explain the difference between the number of households and enterprises visited and the number of surveys completed as recorded in the final dataset. For the household survey the visits minus unsurveyed units is 3492 (compared to 3359 completed interviews in the dataset), while the enterprise survey difference is 15 (425 visits minus 69 who were not surveyed yields 356 enterprises, in contrast to the 341 in the final dataset).

In addition to the survey, enumerators collected water samples from up to two different sources from a randomly-selected subset of households for water quality testing purposes. Due to budget constraints, SI decided to only conduct water testing in Zarqa (as opposed to control areas in Amman), of a total of about 450 water samples collected *at the household level*. The sample for water sample collection was determined at the block level; 60 Zarqa blocks from the household sampling frame were randomly chosen from among all Zarqa blocks included in the larger sample, with up to four households per block randomly selected to provide water samples. Each household could provide up to two types of water samples. By the completion of household survey field work, 457 total samples were collected: 426 of these were household samples and the remaining 31 were quality control samples (blanks, spikes, or replicates). These water samples were analyzed for *total coliforms* and *E. coli* using the IDEXX Colilert Quantitray 2000 method (Westbrook, ME, USA), and 30 of 31 QA/QC samples satisfied quality control criteria.

Baseline data collection for the enterprise survey was conducted between November 13, 2014 and January 27, 2015. Enumerators visited a total of 425 enterprises from the sample (76 in Amman and 349 in Zarqa) (Table 4.2). Questionnaires were recorded as complete for 356 of these enterprises by DoS; 2.0% of enterprises could not be located, while 12.3% of enterprises were closed, and 5.9% of enterprises refused to complete the survey. Again, there were some inconsistencies between these completion rates and those apparent in the final dataset (Table 4.2).

Table 4.2. Survey Outcomes

	From final datasets			**From DoS reporting**		
Province	**# of households visited**	**# completed questionnaires**	**% completed questionnaires**	**Closed** (from DoS)	**Empty** (from DoS)	**Refused** (from DoS)
Household survey						
Amman	1198	1100	92%	9	17	41
Al Zarqa	2457	2256	92%	10	27	59
Total	**3655**	**3359**	**92%**	**19**	**44**	**100**
Enterprise survey						
Amman	76	63	83%	4	0	8
Al Zarqa	349	278	80%	38	7	12
Total	**425**	**341**	**80%**	**42**	**7**	**20**

5 RESULTS FROM BASELINE DATA COLLECTION ACTIVITIES

As noted in the sampling section of this report, baseline data collection yielded useable survey data for 3359 households spread across Census blocks located in Zarqa (2259 households from 216 Census blocks) and Amman (1100 households from 108 Census blocks). The enterprise sample (n=345) was similarly split across Zarqa (281 enterprises from 53 Census blocks) and Amman (64 enterprises from 35 Census blocks). In this section, we provide a description of the demographic and socio-economic make-up of the household survey sample, and describe water sourcing, water and sanitation behaviors, and water-related health outcomes. We then describe the enterprise sample, discussing the types of businesses interviewed, their water sourcing and sanitation practices, and the role of water in their production.

5.1 Baseline descriptive statistics

5.1.1 Household Survey

The average household size in the sample is 4.9 people, with 0.4 children per household under the age of 5, and 3.0 adults over the age of 18 years (Table 5.1). Most households are Jordanian (93%); 85% have a male head of household. These basic demographic variables are the same in both the Zarqa and Amman sub-samples. Education levels in the sample are generally high: 91% of respondent had no trouble reading a written newspaper article, and enumerators judged that 87% had no trouble understanding the survey. The average years of education among all household adults was reported to be 10.6.

Turning to basic socio-economic variables, seventy-three percent of respondents own their residences, 55% of which are apartments or flats. The average residence size is 247 m^2 and has 4.2 rooms (and 1.5 bathrooms), with little difference seen between Zarqa and the larger sample including respondents from Amman. On average, households report spending slightly more (450 JD/month) than they earn (426 JD/month) on a monthly basis (1 JD = US$1.41). The Zarqa sample has slightly lower income and expenditures (409 JD/month and 429 JD/month, respectively). The percentage of households who are National Assistance Fund (NAF) recipients is low, at 2.7%. Only 3.9% of households have a home business, but 20% report accessing a loan in the prior year. Finally, ownership of household durable goods and vehicles varies: nearly all (98%) own a washing machine, but slightly fewer than half of households own a computer (45%) and at least one vehicle (45%), respectively. Twenty percent own at least one air conditioner.

About half of sample households report hearing water, sanitation or hygiene (WASH)-related promotion messages from sources outside the household, and 71% believe that diarrheal diseases can be prevented by WASH practices. Households have relatively low confidence in the quality of water obtained from the utility network, rating the safety of that water to be 5.1 on a 10 point scale ranging from not safe at all to completely safe. Households however consider their own drinking water (which may be treated or obtained from other sources) to be 8.2. An average of 0.15 members per household report having a case of diarrhea in the prior 2 weeks (representing an incidence of about 2.9%), and 0.048 report suffering from some other water-related illness (1.0% incidence).

Table 5.1. Household survey descriptive statistics – Demographic, socio-economic, and health status variables

	Overall sample			Zarqa sample		
Variable	**N**	**Mean**	**(SD)**	**N**	**Mean**	**(SD)**
Demographic variables						
Household Size	3359	4.91	(2.05)	2259	4.91	(2.03)
# of Children < 5 yrs	3359	0.42	(0.70)	2259	0.41	(0.69)
# of adults >=18 yrs	3359	3.02	(1.41)	2259	3.02	(1.40)
Female head of household	3359	0.15	(0.36)	2259	0.14	(0.35)
Age of head of household (in yrs)	3322	50.1	(14.0)	2247	49.9	(13.9)
Number of disabled HH members	3359	0.070	(0.30)	2259	0.064	(0.28)
Jordanian	3359	0.93	(0.25)	2259	0.93	(0.26)
Resident of Zarqa	3359	0.67	(0.47)	n.a.	n.a.	
Socio-economic status						
Respondent is literate	3359	0.91	(0.29)	2259	0.91	(0.29)
Average years of adult education	3359	10.6	(3.45)	2259	10.6	(3.33)
Respondent understood survey well (enumerator rating)	3359	0.87	(0.33)	2259	0.87	(0.34)
Homeowner	3359	0.73	(0.44)	2259	0.74	(0.44)
Area of home (m^2)	3359	247	(2843)	1909	250	(2934)
Home is an apartment/flat	3359	0.55	(0.50)	2259	0.56	(0.50)
# rooms	3358	4.24	(3.57)	2258	4.19	(4.27)
# bathrooms	3358	1.53	(0.65)	2258	1.49	(0.62)
Total expenditure (JD/month)	3272	450	(341)	2191	429	(297)
Total income (JD/month)	3214	426	(351)	2152	409	(303)
NAF recipient	3351	0.027	(0.16)	2253	0.026	(0.16)
Own washer	3359	0.98	(0.14)	2259	0.98	(0.13)
Own computer	3359	0.45	(0.50)	2259	0.44	(0.50)
Own air conditioner	3359	0.20	(0.40)	2259	0.22	(0.41)
Own vehicle	3359	0.45	(0.50)	2259	0.43	(0.450)
Have a home business	3359	0.039	(0.19)	2259	0.037	(0.19)
Took out a loan in the past year	3359	0.20	(0.40)	2259	0.21	(0.41)
Enumerator rating of wealth	3359	2.78	(0.86)	2259	2.75	(0.84)
Health perceptions						
Remember hearing water/sanitation message	3359	0.48	(0.50)	2259	0.42	(0.49)
Believe diarrhea can be prevented	3319	0.71	(0.46)	2240	0.70	(0.46)
Perceived safety of utility water (0=not at all, 10=completely)	3359	5.12	(3.22)	2259	4.88	(3.26)
Perceived safety of home drinking water (0=not at all, 10=completely)	3359	8.18	(2.00)	2259	8.11	(2.05)
Anthropometrics and health measures						
Mean upper arm circumference	707	10.2	(6.46)	461	9.94	(6.43)
Skinfold thickness	707	4.73	(5.58)	461	4.58	(5.88)

# of HH members w/diarrhea, past 2 wks.	3359	0.15	(0.53)	2259	0.15	(0.56)
# of HH members w/other water-related illness	3348	0.048	(0.21)	2253	0.050	(0.22)

The household connection rates to utility water and sewer are 97% and 79% respectively, and similar in the Zarqa sub-sample and the overall samples (Table 5.2). This is somewhat surprising given that the Zarqa sub-sample included areas that were supposedly not connected to the sewer network. These connection rates are indicative of two realities of this system and the partial success of our matching strategy: 1) First, some households do not have private sewer connections even when a sewer line is nearby, due to high connection costs; 2) Second, some households in wastewater expansion areas do already have access to sewer connections; and 3) Our matching strategy was designed to minimize differences (such as this one) across treated and control areas. A large percentage (39%) of households share water meters with others, and only 39% agreed to show a prior water bill during enumerator visits. For the 75.9% of households who produced a bill or otherwise provided self-reported estimates of network water consumption, the average monthly amount used 7.7 m^3 (see Figure 5.1 left panel for the distribution of these monthly amounts in the sample).

Table 5.2. Household survey descriptive statistics – Water and wastewater status and behaviors

	Overall sample			Zarqa sample		
Variable	**N**	**Mean**	**(SD)**	**N**	**Mean**	**(SD)**
Water sources and reliability						
Subscribed to utility water	3359	0.97	(0.18)	2259	0.96	(0.18)
Subscribed to utility wastewater services	3245	0.79	(0.41)	2179	0.81	(0.39)
Share a water meter	3174	0.39	(0.49)	2128	0.39	(0.49)
Able to show a water bill	3219	0.39	(0.49)	2160	0.41	(0.49)
Report using water from any non-network source	3359	0.44	(0.50)	2259	0.42	(0.49)
Use water from a borehole or a well	3359	0.035	(0.18)	2259	0.020	(0.14)
Use water from water shops	3359	0.38	(0.49)	2259	0.37	(0.48)
Use water from tankers	3359	0.036	(0.19)	2259	0.042	(0.20)
Amount of network water used (m^3/month)	2550	7.73	(6.34)	1749	7.64	(6.14)
Amount of non-network water used (m^3/month)	3359	0.38	(1.33)	2259	0.33	(1.25)
Change sources in other seasons	3350	0.054	(0.23)	2253	0.069	(0.25)
Change amounts of water used in other seasons	3350	0.34	(0.48)	2253	0.35	(0.48)
Reported experiencing water shortage	3343	0.23	(0.42)	2247	0.29	(0.45)
Days in average month receiving water	3156	9.17	(7.09)	2138	8.33	(6.34)
Had complaints about WAJ service	3245	0.19	(0.39)	2179	0.23	(0.42)
Water & hygiene behaviors						
Treats water currently	3359	0.35	(0.48)	2259	0.38	(0.49)
Had treated water on hand	3359	0.34	(0.47)	2259	0.36	(0.48)
If treating, stores treated water 1 day or less	1300	0.90	(0.30)	918	0.91	(0.29)
Keeps storage containers covered	3312	0.97	(0.17)	2240	0.97	(0.16)
Keeps storage containers elevated	3177	0.68	(0.47)	2148	0.61	(0.49)
Had soap on hand	3135	0.88	(0.33)	2103	0.86	(0.35)
Sanitation status & behaviors						
Has a private toilet	3359	0.76	(0.43)	2259	0.77	(0.42)
Toilet not connected to sewer system	3359	0.24	(0.43)	2259	0.23	(0.42)

Estimated WAJ sewer connection fee (JD)	927	117	(111)	650	117	(112)
Sanitation situation: 1=Excellent, 5=very poor	3218	2.83	(1.17)	2191	2.90	(1.16)

Households use a variety of non-network water sources (44% use at least one other source), but shop water is by far the most frequent alternative, with 38% of households purchasing it. The quantity of water taken from non-network sources (0.38 m^3/month) is far lower than the average network consumption, however (see Figure 5.1 right panel for the distribution of these monthly amounts in the sample among households using non-network sources). While 34% of households change their water consumption patterns in the alternative (winter) season, very few households (<1%) change their sourcing practices. Twenty-three percent of households report experiencing water shortages, with a somewhat higher proportion (29%) experiencing shortages in Zarqa. Households report receiving water 9.2 days per month on average (and only 8.3 days per month in Zarqa). Nineteen percent of households report complaints against utility services (23% in Zarqa).

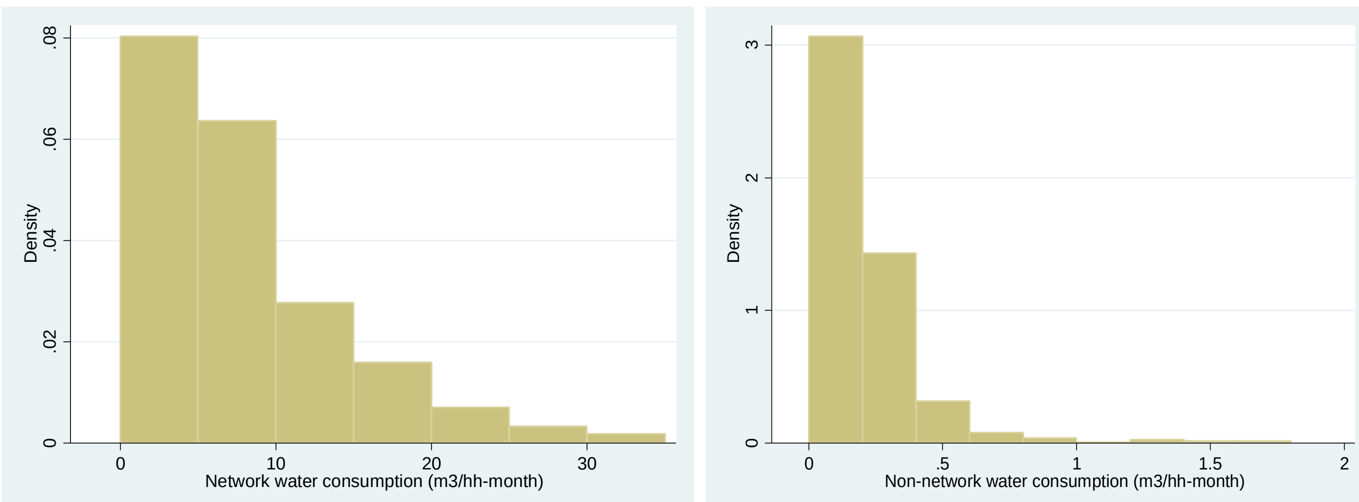

Figure 5.1. Distribution of monthly (left) network water amounts used and (right) non-network water amounts used
(Note the differences in the scale for the x-axis)

Water and hygiene behaviors were variable in the sample. Thirty-five percent of households report treating water in house at the time of the survey, and most of these (34%) were able to show enumerators a sample of treated water during the visit. Most of these households (90%) consume treated water less than one day after treatment. Nearly all (97%) households covered all of their water containers, though somewhat fewer (68%, and 61% in Zarqa) kept these containers elevated above ground level. A large majority of households (88%) also had soap at the time of the visits.

There was some variation in the quality of the household samples that were tested. Emphasizing that these samples were collected from storage containers and that their quality is likely highly dependent on storage and handling of water in the home, we found little evidence of *E. coli* contamination problems. *E. coli* was below detection (1 colony-forming unit per 100 mL sample) for *all* tap water samples collected from household storage tanks. Three of 91 samples that had been sourced from water shops and subsequently stored at home showed modest E. coli contamination (7, 28, and 54 CFU/100mL, respectively).

Unsurprisingly, there was somewhat higher prevalence of total coliforms in stored water: these were detected in ~10% of the tap water samples. Nonetheless, the counts were low: All samples except one had less than 100 CFU/100mL. Finally, likely due to the fact that such water is stored for longer periods of time, we detected total coliforms in over 70% of shop water samples, with 29% of samples having more than 100 CFU/100mL and 11% more than 1000 CFU/100mL. In general, these results suggest that water quality problems are not widespread in this sample, are mainly the consequence of water storage, and are unlikely to lead to significant diarrheal disease.

Seventy-six percent of households in the sample have a private toilet (others have a toilet shared with other households), though 24% of toilets used by households are unconnected to the utility network (consistent with the 79% reported sewerage rates). Most households did not know the cost of a sewer connection, but those that did (n=927) reported these to be 117 JD on average. On a 5-point scale, households rated their sanitation situation to be 2.83 (closer to acceptable (3) than good (2)).

Using the extensive data collected in the baseline survey, we are able to estimate water- and wastewater-related expenditures, and compute coping costs in different categories (e.g., time spent collecting water from alternative sources, or in-house water treatment costs). The various categories of coping costs are summarized in Table 5.3, and data on these variables follow in Table 5.4 (the distribution of these costs is shown in Figure 5.2). As shown, households spend about 18 JD/quarter on network water and sewer; the amount is only slightly lower (19.1) in Zarqa. The largest category of water-related coping costs is for purchase of non-network water (roughly 6.5 JD/month); this significant expenditure is not unexpected – after all it was a key component of the economic case for the Compact (Albert et al. 2013) – but is striking given the much lower amounts of non-network water that are consumed. Households spend another 5 JD/month on water collection, treatment, storage, and on repairs to household infrastructure, such that overall water-related coping costs are 11.5 JD/month (and 11.8 JD/month in Zarqa).

Wastewater-related coping costs are primarily in pit emptying (3.2 JD/month) and toilet infrastructure costs (6.5 JD/month). Time costs for use of toilets are low, given that most households have a private toilet. The wastewater-related coping costs in the full sample and in Zarqa are similar (17.5 JD/month vs. 17.2 JD/month).

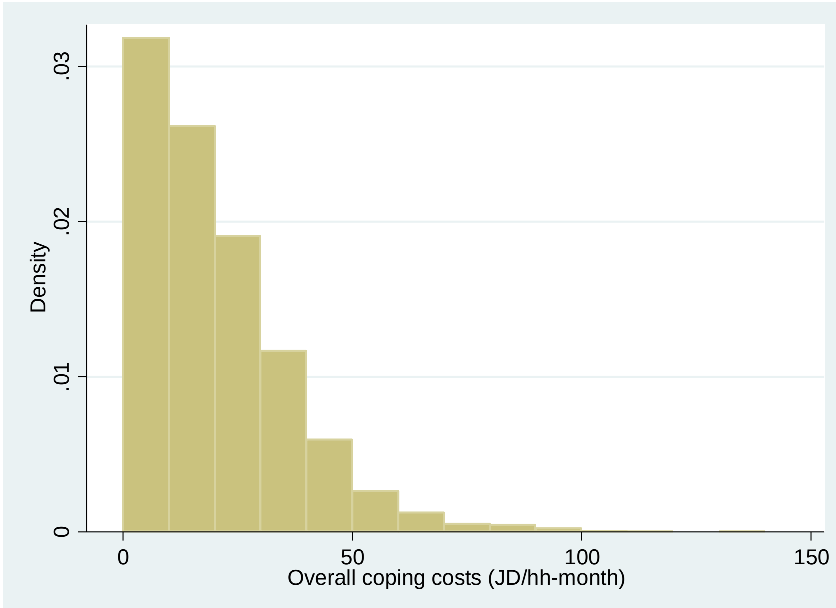

Figure 5.2. Distribution of total water and wastewater-related coping costs among survey households

Table 5.3. Definitions of water- and wastewater related coping cost variables

Coping Cost Category	Data Source	Assumptions	Formula
Non-network water expenditures	• Monthly non-network water purchases (p_i) for all non-network sources (i)	None	$\sum_{i=1}^{I} p_i$
Water collection costs	• Estimates of water collection time (min/trip) (t_i) for non-network water sources (i) • Estimates of monthly quantities (m^3) (q_i) used of non-network water sources (i)	• One trip/week and four weeks/month • Value of time is average wage (w)	$\sum_{i=1}^{I}(t_i/60) * w * 4$
Water treatment costs	• Estimates of monthly treatment costs (c) • Estimates of equipment costs (e)	• 5 year lifespan of equipment • 5% discount rate	$c + \frac{e}{1.05^{5*12}}$
Water storage costs	• Estimates of monthly costs of cleaning storage containers (S)	NONE	S
Expenses on in-house water repairs	• Estimates of yearly repair costs (r) • Estimates of time spent on repairs (s)	• Value of time is average wage (w)	$\frac{r+s*w}{12}$
Toilet cleaning costs	• Estimates of monthly time costs of cleaning toilets (c)	None	c
Toilet infrastructure costs	• Estimates of costs to replace toilet (r) • Reported connection fees to WAJ-wastewater (w)	• Average lifespan of toilet is 20 years • 5% discount rate	$(0.08 * (r + w))/12$
Time spent on trips to toilet	• Estimates of time (minutes/trip) spent walking to toilet for households with shared toilets (t)	• Value of time = average wage (w) • 3 trips/day per household member. • Household size=h.	$\left(\frac{t}{60}\right) * w * 3h$
Pit emptying costs	• Cost of emptying pit (p) • Frequency of emptying pit in months (m)	None	$\frac{p}{12} * m$

Table 5.4. Water and wastewater-related expenses, and coping costs (all in JD/month)

	Overall sample			Zarqa sample		
Variable	**N**	**Mean**	**(SD)**	**N**	**Mean**	**(SD)**
Expenses on network water	2851	6.16	(8.20)	1928	5.80	(8.06)
Water-related expenses / coping costs						
Expenses on non-network water	3359	6.51	(12.63)	2259	6.48	(12.93)
Value of collection time	3359	0.18	(0.37)	2259	0.18	(0.37)
Water treatment costs	3359	1.97	(4.19)	2259	2.15	(4.29)
Expenses on in-house water repairs	3359	1.46	(4.31)	2259	1.52	(4.33)
Storage costs	3359	1.38	(3.48)	2259	1.43	(3.64)
Total water-related coping costs	**3359**	**11.5**	**(13.79)**	**2259**	**11.8**	**(14.1)**
Wastewater-related expenses / coping costs						
Expenses on pit emptying	3359	3.20	(6.35)	2259	3.00	(6.25)
Toilet infrastructure costs	3359	6.47	(3.46)	2259	6.50	(3.48)
Time costs for using sanitation	3359	0.22	(1.43)	2259	0.11	(0.99)
Total wastewater-related coping costs	**3359**	**9.88**	**(7.33)**	**2259**	**9.61**	**(9.08)**

Notes: When billing information for network water from the prior period was not available, network water expenses were estimated using self-reports from the most recent quarter. If households reported water bill amounts, those were used. If households reported quantities but not bill amounts, the bill was estimated using the known water tariff structure in Zarqa.

5.1.2 Enterprise Survey

Table 5.5 provides an overview of firm composition, obstacles, and growth. The average enterprise in our sample has 5 employees, most of whom are male and full-time workers. The average full-time skilled worker for the businesses in our sample is 30 years old. The majority of business owners are male, and the average business owner in our sample is 46 years old with 15 years of experience. Respondents in our sample cited inflation and price instability, electricity costs, and water supply costs as the largest obstacles to growth for their enterprises. Very few enterprises in our sample reported making any investments in their business or taking out a loan in the past year.

Table 5.6 contains detailed descriptive statistics of water and wastewater usage and costs for enterprises. We observe that firms use a combination of piped connections, water tankers, and water shops for their business needs. Figure 5.3 shows the distribution of water sources that enterprises reported as their “primary water source.” We observe a fairly even distribution of private (company) piped connections, tankers and water shops as enterprise primary water sources. Some enterprises share connections with others or with households. Enterprises spend an average of 57 JD/month for their total water usage, and only 10% of enterprises treat any water, either for drinking or production purposes.

While most enterprises do not report having to shut down due to water shortages, enterprises with piped connections report receiving an average of 7.9 days of piped water during a summer month and only 42.8 hours in a normal week. Figure 5.4 displays the primary reasons that enterprises do not have piped connections. Nearly 50% of enterprises report that high connection (not tariff) costs prevent them from using piped water, and almost 90% of unconnected enterprises list this as the main obstacle to their use of network water.

Table 5.5. Enterprise survey descriptive statistics – employee data, business owner characteristics, obstacles to growth, assets, and costs

	Full sample			Zarqa sample		
Variable	**N**	**Mean**	**(St. Dev)**	**N**	**Mean**	**(St. Dev.)**
Firm characteristics						
Sole proprietorship	345	0.87	(0.34)	281	0.87	(0.33)
General partnership company	345	0.084	(0.28)	281	0.068	(0.25)
% with government shareholder	345	0.0	(0.0)	281	0.0	(0.0)
Business operates year round	342	0.99	(0.11)	278	0.99	(0.12)
Employee data						
Total employees	341	5.09	(11.0)	277	5.09	(11.4)
Total male employees	341	4.29	(7.89)	277	4.19	(7.86)
Total skilled full-time employees	341	2.19	(8.96)	277	2.32	(9.88)
Total unskilled full-time employees	341	1.70	(4.24)	277	1.66	(3.83)
Total skilled part-time employees	341	0.19	(1.17)	277	0.13	(0.87)
Total unskilled part-time employees	341	0.30	(2.17)	277	0.31	(2.32)
Total unpaid workers	341	0.71	(0.58)	277	0.66	(0.57)
Avg. age of full-time, skilled workers	155	30.0	(7.44)	118	30.2	(7.93)
Business Owner Overview						
Years of owner experience	341	15.3	(10.1)	277	14.9	(10.2)
Business owner's age	341	46.3	(12.3)	277	46.3	(12.1)
Business owner's gender (1=female)	343	0.079	(0.27)	279	0.082	(0.28)
Business owner's total monthly income	151	666	(629)	124	599	(470)
Obstacles to growth (1=Not at all; 5=Very big)						
Obstacle to growth - cost of electrical service	341	3.84	(0.97)	278	3.88	(0.92)
Obstacle to growth - water quality and reliability	341	3.06	(1.19)	278	2.91	(1.15)
Obstacle to growth - cost of water supply	341	3.65	(1.15)	278	3.55	(1.16)
Obstacle to growth - insufficient demand	341	2.62	(1.08)	278	2.62	(1.06)
Obstacle to growth - inflation and price instability	341	4.35	(0.92)	278	4.27	(0.96)
Enterprise Assets						
Estimated market value of property's land ('000 JD)	197	43.9	(61.4)	150	42.8	(66.8)
Estimated market value of property's buildings/ structures ('000 JD)	250	54.9	(203)	197	56.4	(224)
Cost of setting up firm to where it is now ('000 JD)	294	76.8	(201)	239	77.5	(214)
Made any investments in business, last yr	341	0.01	(0.09)	278	0.01	(0.10)
Business' total sales last month ('000 JD)	271	8.56	(26.5)	230	9.35	(28.1)
Enterprise has a checking or savings account	340	0.26	(0.44)	277	0.24	(0.43)
Took a loan during 2014	341	0.02	(0.13)	278	0.01	(0.12)
Monthly enterprise costs						
Paid labor ('000 JD)	240	1.79	(4.12)	195	1.84	(4.38)
Services (JD)	38	1.23	(2.63)	38	1.23	(2.63)
Land/building rent (JD)	278	0.97	(3.17)	224	0.89	(2.33)
Electricity (JD)	325	0.44	(1.47)	262	0.44	(1.53)

Table 5.6. Enterprise water and wastewater use practices and characteristics

	Full sample			Zarqa sample		
Variable	**N**	**Mean**	**(St. Dev)**	**N**	**Mean**	**(St. Dev.)**
Water usage and behaviors						
Use private piped water	341	0.30	(0.46)	278	0.28	(0.45)
Use shared piped water	341	0.18	(0.39)	278	0.19	(0.39)
Use water tanker	341	0.26	(0.44)	278	0.27	(0.44)
Use water shops	341	0.43	(0.50)	278	0.45	(0.50)
Use other source of water	341	0.06	(0.24)	278	0.06	(0.24)
Monthly cost of using private piped water	101	74.2	(122)	77	74.4	(129)
Average cost of total water per month	341	57.2	(114)	277	58.1	(118)
Treats water	341	0.09	(0.29)	277	0.10	(0.30)
Firm stores water	341	0.72	(0.45)	277	0.70	(0.46)
Total amount of water that firm is currently storing (m^3)	341	3.62	(6.97)	277	3.36	(5.87)
Water shortage and piped water characteristics						
Business has reduced or stopped work due to water shortage	341	0.01	(0.11)	278	0.01	(0.10)
Private connection pre-dates business start	162	0.81	(0.39)	129	0.79	(0.41)
Connection costs for business	26	273	(197)	24	283	(202)
Price of 1 m^3 of water	123	1.29	(0.81)	96	1.33	(0.88)
Days of piped water per month (days/month)	163	7.88	(4.93)	130	8.45	(5.07)
On days w/ water, continuity of supply (hrs/day)	163	17.0	(7.72)	130	18.3	(7.05)
Hours of water in normal week (hours/week)	163	42.8	(33.6)	130	46.7	(33.9)
Wastewater characteristics						
Business has a wastewater management system	341	0.69	(0.46)	278	0.68	(0.47)
Wastewater system pre-dates business start	232	0.86	(0.35)	188	0.85	(0.36)
Wastewater is connected to sewer	234	0.93	(0.25)	190	0.92	(0.28)
Monthly cost of sending wastewater to sewer	218	7.22	(13.2)	174	6.21	(11.0)
Cost of installing sewer connection	218	141	(370)	174	133	(382)
Wastewater goes to septic tank/field	234	0.06	(0.24)	190	0.07	(0.26)
Monthly cost of sending wastewater to septic tank / field	14	13.1	(20.0)	14	13.1	(20.0)
Cost of installing septic tank	14	326	(419)	14	326	(419)

Table 5.6 also provides important information about wastewater behaviors and costs. Most (69%) of the enterprises in our sample have a wastewater management system. The majority of these firms report that their wastewater goes to a sewer connection. The average cost of using the WAJ sewer connection is 7.2 JD/month with a connection fee of 141 JD. Interestingly, the wastewater costs for enterprises that use the WAJ sewer connection are significantly lower than the costs of firms that use septic tanks for their wastewater.

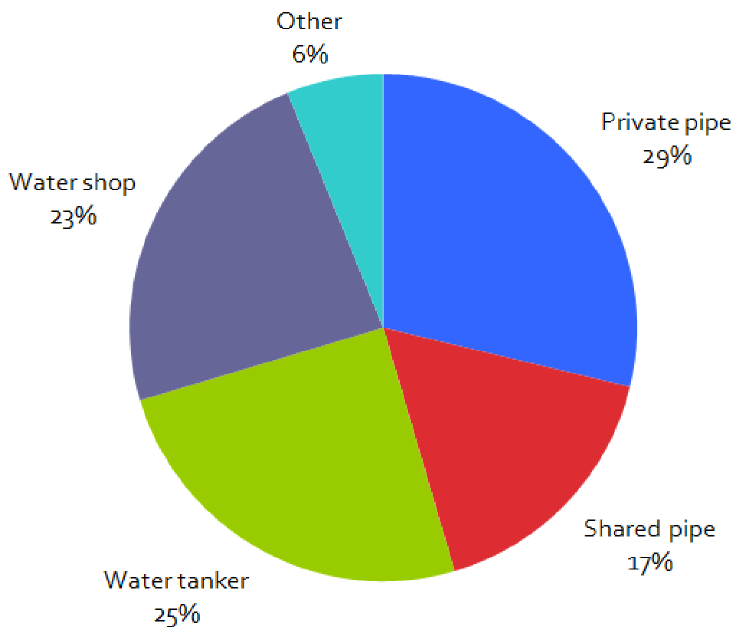

Figure 5.3. Distribution of main water sources used by surveyed enterprises (n=341)

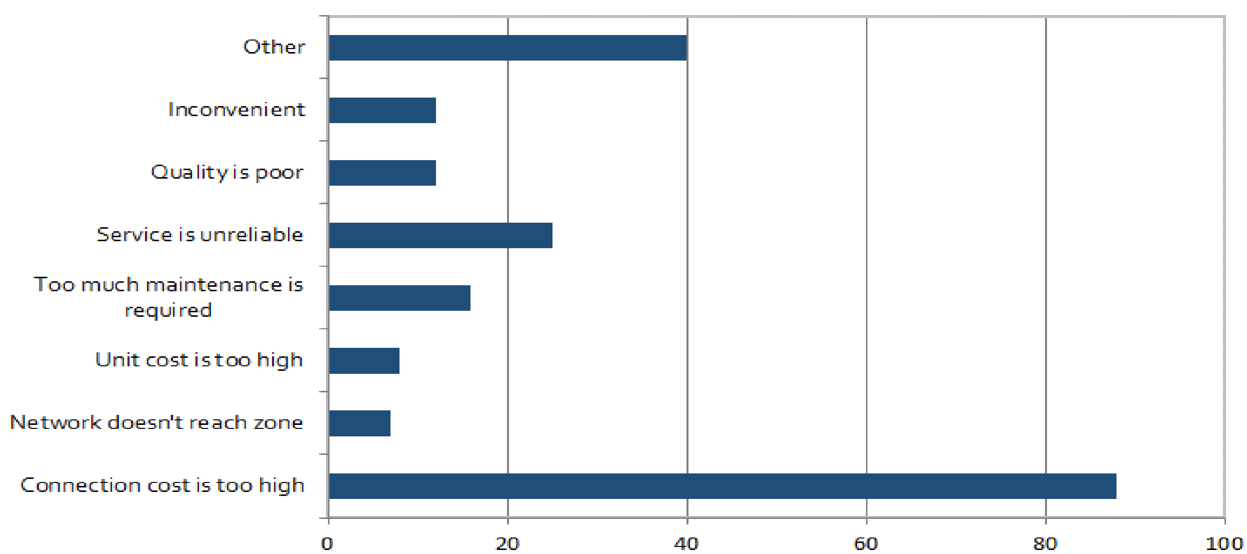

Figure 5.4. Reasons why enterprises do not have piped water (n=178)

5.2 External validity of the baseline samples

Noting that the sampling design does not purport to create a representative sample as described in the EDR and above, it is nonetheless useful to compare our descriptive statistics with those obtained in other representative surveys previously conducted in Jordan. Such a comparison provides a basis for interpreting the results we obtain, since accurate estimation of the overall population-wide impacts of the Compact may require re-weighting of our sample-estimated impact measurements, if the sample is found to not represent Zarqa well overall.

5.2.1 Household survey

Regarding the household survey, we compare our Zarqa sample with the representative sample surveyed in the 2009 Department of Statistics Water Survey (DoS, 2010). In that survey, the average household size was found to be 5.4 people, or slightly more than the 4.9 in our survey. Monthly expenditure and income were 318 and 306 JD, respectively, which converted to 2014USD corresponds to 391 and 376 JD (compared to 429 and 409 JD, respectively, in our sample).[8] Sixty percent of households lived in apartments (compared to 56% in our sample), and 12% are female-headed (14% in our sample).

With regards to water and sewer services, the 2009 Water Survey found that 97% had WAJ water connections, and that 85% were connected to the sewer network (similar to the 96% and 81% that we found). Fifteen percent of households reported health effects from poor water supply (compared to about 19% in our sample). Water sourcing in the 2009 Water survey indicated 97% using piped water, 34% using shops, 10% using water vendors, and about 6% using other sources (compared with 96%, 37%, 4%, and 5%, respectively in our sample). The average number of days per month of water supply reported in that survey was 9.5, slightly higher than the 8.3 in our sample.

Thus, our sample does not appear fully representative (it is slightly higher income, households have fewer members, and tanker/vendor use appears lower[9]), but the differences with the 2009 representative sample are minor, and it seems likely that our results will be representative of overall impacts. We will of course explore the implications of using sample weights to re-estimate impacts. One potentially important issue for MCC and MCA-J however is that the sample, being fairly representative, includes very few NAF recipients, and so will not be helpful for measuring the effects of the infrastructure components of the Water Smart Homes Activity.

5.2.2 Enterprise survey

Similarly, for assessment of the representativeness of our sample of enterprises, we can compare our survey to the World Bank Enterprise Surveys, the most recent of which was conducted in 2013 for Jordan (World Bank, 2013). World Bank enterprise surveys are not representative, as the sampling methodology stratifies on firm size. Specifically, these overweight large firms since the majority of firms are small or medium in size, whereas a disproportionate share of economic activity and growth occurs in large firms. Nonetheless, these Enterprise survey data can be reweighted to yield representative population means, for comparison with those of the Impact Evaluation sample.

[8] Using the inflation rates for Jordanian Dinar from 2009 to 2014 reported by the IMF, see: http://www.imf.org/external/pubs/ft/weo/2015/01/weodata/index.aspx

[9] It should be noted that tanker water use is highly seasonal, however, so this difference could simply be due to the time of year when the survey was conducted. This provides motivation for doing more frequent data collection.

Following this procedure, the years of owner experience reported in Table 5.5 above (14.9) is somewhat lower than the average over the weighted Zarqa sample from the World Bank survey (19.4 years), but the percentage of firms with women owners is similar (8% vs. 9%). As in the World Bank survey, almost no enterprises have government ownership, but many more in our sample are sole proprietorships (87% vs. 54% in the World Bank survey), with a correspondingly lower number of partnerships (7% vs. 36%). The firms in that survey are also much larger than those in our survey (53.9 full or part time workers versus 5.1 total workers in our sample). The proportion of unskilled workers in our sample (37%) is also higher than that in the World Bank survey (16%), as is the percentage of female workers (20% vs. 13%). Finally, many more enterprises in the World Bank survey have checking accounts (81% vs. 24%) and took loans in the past year (16% vs. 1%) than in our sample.

Turning to water sources, 26% of enterprises in Zarqa have piped water connections, compared to 28% in our sample. The weighted World Bank sample indicates 27% use tankers (vs. 27% in our sample), and 11% use other sources (vs. 6% in our sample). Altogether, we can therefore conclude that our sample is not representative of Zarqa enterprises, in that it contains a preponderance of small enterprises, but that water sources used are not that different from governorate-wide sourcing practices. This is unsurprising given the sampling approach we utilized and the fact that we did not see large differences in the household sample from the same blocks versus Zarqa as a whole. Perhaps most crucially for the enterprise component of the evaluation, the current sample will not allow us to measure impacts on large businesses, unless we expand the scope of the enterprise survey.

5.2.3 Overall summary

Overall, the household sample appears much more representative than the enterprise sample, when we compare descriptive statistics for the full sample. The enterprise survey sample only looks representative with respect to water sourcing practices (and not overall firm characteristics).

5.3. Internal validity risks and considerations

This section summarizes the main risks to internal validity of the impact evaluation as informed with analysis of the baseline data. We focus on three major threats at this time, namely measurement issues with specific and important variables, overall sample balance across treatment and control arms, and known confounders that may affect our evaluation moving forward.

5.3.1. Measurement issues

<u>Household survey</u>: The data completeness from the household survey generally looks good, but there are a few very important variables that raise concern. Perhaps most critically, there are problems in the measurement of expenses on network water. Many households refused or were unable to share a recent bill for enumerators to record these costs. In total, only 1102 households provided bills. To estimate costs for households who did not provide bills, we first used self-reported bill amounts, for 1235 additional households. Finally, some respondents only provided estimated quantities. For these households, we applied the tariff structure to self-reported volumes consumed during the last billing period for those

whose expenses were still missing (n=220). Thus, 510 households still have missing expenses at baseline (out of 3,067 households subscribed to WAJ services). We also note that households that estimated both quantities and expenses (n=861) had statistically different values when we compared self-reported estimates to the expenses imputed from the tariff structure. We believe the estimated expenses may be more reliable than estimated amounts because these are a) likely more salient to households, and b) closer to the actual bill amounts for the sub-sample that produced bills.

In addition to missing data on network expenses, it seems likely that households underreported some of their use of non-network sources of water. While the percentage of households using shop water seems realistic, tanker water seems underreported (only about 4% of households report using this source). This could be a seasonal issue, as our survey occurred during a specific window in time. Nonetheless, few households reported additional use of tanker water in other seasons. This missing data suggest a need for monitoring of water sourcing on a more continuous basis.

<u>Enterprise survey</u>: As in the household survey, we have concerns over missing data for some important variables in the enterprise survey. Wages for workers in our sample are not realistic and have likely been misreported, leading to erroneous estimates of labor costs. Enumerators likely coded the unit of time for the wage (day, month, or hour) incorrectly, and it is difficult to decipher from the data the actual wages. Finally, we observe an underreporting of certain financial or cost estimates. For example, we observe a preponderance of missing data for business owner's monthly income, estimates of market value for business property, estimates of last month's sales, and monthly business cost estimates, as well as unit costs for network water and connection fees for piped water supply. We can likely recover the latter using utility data. The follow-up survey training will need to emphasize the importance of obtaining complete data, though we know from the fieldwork that part of the underreporting is due to concerns related to divulging sensitive cost or revenue information.

5.3.2. Sample balance

<u>Household survey</u>: As described in more detail in the sampling section of this report, we matched the samples using 10 Census variables, and achieved good balance on these characteristics at the block level (Table 5.7). We then tested balance within the matched treated and control blocks for 105 variables measured in the household survey, covering demographics; socio-economics; health; water status, sources and water-related behaviors; sanitation status; and perceptions of water and sanitation situation. By chance, we would thus expect about 10 variables to be unbalanced at the 10% significance level, and our imbalances exceed these expected amounts, especially between Amman and Zarqa (as might be expected due to the differences in these two governorates). Table 5.7 summarizes these results, and identifies the main variables of concern for internal validity. We do emphasize, however, that difference-in-difference analysis does provide protection against imbalances, so long as common trends prior to intervention can be established. We intend to use the 2009 Water Survey to establish such trends.

Table 5.7. Summary of balance checks across households and enterprises sampled from matched treatment and control blocks in the household survey

	Zarqa controls			Amman controls		
	WNP only	WWNP only	Both	WNP only	WWNP only	Both
Household survey						
# of different block-level variables; Census (p<0.1)	0 / 10	2 / 10	0 / 10	0 / 10	0 / 10	0 / 10
# of different variables; baseline hh survey (p<0.1)	40 / 106	33 / 106	41 / 106	41 / 106	31 / 106	41 / 106
Chief variables of concern for particular group	-Consumption -Income -Assets - Water quality perception -Use water shops -Water continuity -Water shortage -Sewer connection -Private toilet -Sanitation situation	-Own house -Consumption -Income -Water quality perceptions -Meter sharing -Complaints -Use shops -Treats water -Sanitation situation	-Consumption -Water quality perceptions -Diarrhea prev. -Meter sharing -Use shops -Water continuity -Water shortage	-Own house -Consumption -Income -Assets -Water quality perception -Sewer connection -Use shops -Water continuity -Water shortage -Private toilet -Sanitation situation	-Own house -Consumption -Income -Water quality perception -Meter sharing -Use shops -Complaints -Treats water -Sanitation situation	-Consumption -Water quality perception -Diarrhea prev. -Meter sharing -Use shops -Water continuity -Water shortage
Enterprise survey						
# of different variables; baseline firm survey (p<0.1)	11 / 110	14 / 110	14 / 110	10 / 110	13 / 110	17 / 110
Chief variables of concern for particular group	-Water continuity -Infrastructure an obstacle -Access to credit	-# unskilled workers -Female owner -Cost of water -Water supply an obstacle -Land value -Cost of materials	-Owner experience -Cost of water -Water continuity -Sewer connection -Bank account -Costs	-Labor skills an obstacle -Cost of materials	-Water continuity -Sanitation situation -Sewer connection -Total sales	-# unskilled workers -Owner education -Water cost -Water continuity -Infrastructure an obstacle -Total sales -Access to credit

Enterprise survey: We also tested balance within the matched treated and control blocks for 110 variables measured in the enterprise survey, covering firm characteristics; revenues and costs; sources and water-related issues facing the firm; sanitation status; and perceptions of business constraints. By chance, we would thus expect about 11 variables to be unbalanced at the 10% significance level, and our imbalances only slightly exceed these expected amounts, and are slightly larger between Amman control and Zarqa treatment units (again as might be expected due to the differences in these two governorates). Table 5.7 again summarizes these results, and identifies the main variables of concern.

5.3.3. Known confounders

There are three important known confounders that may pose significant problems for inferences from this evaluation. These are a) the completion of the Disi project and the water it contributes to Zarqa; b) the installation of a management contract for Miyahuna (the operating company in Amman) at the utility in Zarqa; and c) the evolving refugee situation in Zarqa. We discuss the threats from each of these qualitatively, and briefly, below.

First, the Disi Project promises to provide substantial inputs of new water to urban areas in Jordan, and it is currently unclear to the Evaluation Team exactly how much of this water is being sent (directly or indirectly to Zarqa). It is our hope that we can get a handle on inputs from Disi through the water balance exercise in Component 2 of the Evaluation. In theory, if Disi water is being evenly distributed throughout the Zarqa Governorate, the use of DiD to difference out this input will protect us against this confounder. However, it is possible that the value of Disi will be higher in Compact areas if more of the water makes it to beneficiaries living in WNP improvement zones. In addition to this, the comparison with control blocks in Amman may be compromised by Disi, given that Amman is likely receiving relatively more water from that new source than Zarqa. The first effect (synergies between Disi and the Compact) will lead to overestimates of the value of the MCC investments, while the second effect will likely lead to downward bias in those estimates. The water balance exercise is key to quantitative understanding of these offsetting effects

Second, the new management contract in Zarqa (with Miyahuna) creates the potential for useful reform and efficiency gains in utility operation, but may also create disruptions, if costs increase (WAJ-Zarqa is getting a new building), water rates change, or if utility policy evolves in a way that changes water consumption. Here again, it is important to understand this threat in the broader context of what SI proposed in the EDR, which includes (in Component 3) detailed utility tracking. This utility tracking has a comparative component (with other utilities throughout Jordan), as well as a more spatially-differentiated component, which focuses on performance measures within sub-zones of the network in Zarqa. This evolution in management only serves to emphasize and remind us of the need for devoting additional attention to Component 3 of the evaluation.

Finally, the refugee situation in Zarqa offers a different type of challenge. Specifically, there is a growing demand for water in Zarqa due to population growth which was not fully anticipated at the time of the Compact development. This growth in demand likely makes the water savings in the Compact more valuable, but it also makes the impacts harder to detect. Besides making simple comparisons on a per household basis (where greater numbers of households may be sharing water than before), the evaluation should therefore work to understand the larger picture of population and demand growth, and might include a cross-sectional or repeated cross-sectional survey to quantitatively understand the nature of changing demands due to this transient population.

5.3.4. Summary

The threats to the internal validity of the impact evaluation that are assessed here suggest that the greatest concerns from the baseline survey pertain to the household survey sample, and stem from a) measurement problems with a few key variables (namely subscriber numbers and the dynamics of water sourcing); and b) imbalances between treatment and control units. Many of these imbalances, such as those arising from time-invariant differences) will be mitigated through the difference-in-difference design of the study, but additional testing of differential trends is warranted, perhaps using the 2009 DoS Water Survey. We can also utilize ex-post matching of households and multivariate regression methods to further mitigate these threats. There are also major confounders, of which the evaluation and its clients must be aware as they interpret our findings.

5.4. Issues related to the power of the IE

Finally, we close with a discussion of issues related to the statistical power of the evaluation. We discuss this issue with regards to the household and enterprise survey separately.

5.4.1. Household survey

Power of the baseline sample. As mentioned above, the quality of the matches obtained from the *ex ante* PSM approach used in the design of our sample implies some loss of power for detecting differential impacts across the various treatment arms of the intervention, relative to our initial baseline power calculations (presented in the EDR). Specifically, the three treatment groups are smaller than the original target sample size, while the control groups are larger.

As shown in Table 5.8, additional power was lost because DoS delivered data for 97% of the targeted baseline sample size, likely because it was not clear that a procedure for replacing households who were unavailable to be surveyed should be implemented at the time of this survey (a problem that was subsequently addressed prior to the baseline enterprise survey). This *initial loss of sample size* is less than the 10% attrition expected in our initial power calculations, but the buffer to protect against this attrition has been reduced to about 7%.

These issues together imply that we may not have sufficient power to detect 10% differences between the WWNP, WNP, and both WWNP and WNP groups, for the following variables of interest: hours of water supply, water consumption from the network, and water quantity purchased from shops. We will be able to detect such differences between groups treated by the WNP project and those not treated by it (since the former includes two treatment groups: Zarqa

WNP and Zarqa both). We will also be able to use utility data to address the lack of power with respect to water consumption from the network, since we now have access to quarterly consumption of *all households* within our sample blocks.

Table 5.8. Proposed, targeted, and final distribution of households across the five sample arms

	Proposed distribution of households (from EDR)			Final target distribution of hhs		Final distribution of hhs		Deviation from target
	N	Attrition	Total	Blocks	Households	Blocks	Households	
Zarqa Wastewater	625	62.5	687.5	43	473	43	456	-3.6%
Zarqa Water	625	62.5	687.5	49	539	48	493	-8.5%
Zarqa Both	625	62.5	687.5	43	473	43	450	-4.9%
Zarqa Control	625	62.5	687.5	82	902	81	845	-6.3%
Amman Control	625	62.5	687.5	108	1188	108	1098	-7.6%
Total	3125	312.5	3440	325	3575	323	3359	-6.0%

Notes: Two of the originally sampled blocks were dropped from the sample due to their low population or very high refusal rates; this explains why the final sample only includes households from 323 blocks.

We would likely also recover power if we: a) implemented higher frequency data collection (as proposed to the MCC and MCA-J); b) replaced the households that were initially lost (at the time of the next round of data collection – this is planned); c) increased sample sizes within the blocks beyond the original sizes. The first two of these options should clearly be done, particularly if we conduct a seasonal survey later this year to address some of the other data quality issues we identified in the baseline data. The proposed seasonal data collection in September of this year would still be prior to many benefits accruing in treated areas, so we might even still be able to add these replacement households to the full DiD estimation. The third option requires additional discussion, since the gains in power would come at greater relative cost owing to the diminishing gains in power given the existing numbers of sample blocks.

Re-calculation of power using the new baseline data. Re-doing the power calculations using our new baseline data (rather than that from the 2009 DoS Water Survey) and a somewhat augmented set of outcomes produces the results summarized in Table 5.9. As shown, the treatment effects that we will be able to detect, based on these data, are typically larger than 10%. For many key variables, these remain lower than 20% (these are highlighted in bold), but several exceed 20%, which is somewhat concerning. In part, this stems from the fact that the data are much more variable than they were in the 2009 Water Survey. It may be that additional data cleaning and work with the outliers in the data will help improve power, but these results reinforce the need for additional high frequency data collection, better programming controls (to limit problems with outliers), and more intensive real-time monitoring of data quality by Social Impact and DoS, to improve understanding of the water and wastewater expense dynamics among these households.

5.4.2. Enterprise survey

At the time of the design of the IE, there were two major challenges related to assessing the statistical power of the enterprise survey: a) a lack of pre-intervention data that were relevant for measuring water-related impacts on firms; and b) a lack of literature on the effects on enterprises that could be anticipated from investments such as the Compact. The first of these challenges has been addressed through collection of data from a representative sample of enterprises *within* our survey zones, but the second persists. As a result, the power calculations presented here should be considered highly speculative. We have assumed a 20% treatment effect in column A, without a great deal of justification.

Table 5.9. Revised power calculations, based on baseline survey data

	A. Sample size required to identify a 10% treatment effect	B. Size of treatment effect detectable with current sample size (p=0.1)
Water Network Project Outcomes		
Monthly expenses on tanker water (JD)	2062	**-19.4%**
Monthly expenses on shop water (JD)	1592	**-17.1%**
Shop water consumption (L/month)	1740	**-17.6%**
Hours of network supply per day	144	**5.1%**
Network consumption (m^3/month)	1660	**17.5%**
WAJ bill amount (JD/month)	4557	28.4%
Monthly costs of in-house water treatment (JD)	10746	-49.7%
Monthly cost of water-related illness (JD)	17357	-56.0%
Total monthly spending on water (JD)	3626	-25.7%
Monthly water-related coping costs (JD)	2318	-27.0%
Wastewater Network Project Outcomes		
Connected to sewer (%)	1129	**13.9%**
Pit-emptying costs	18414	-57.8%
Monthly sanitation coping costs	2065	-**19.7%**

Notes: Standard deviations, means, and intra-cluster correlation statistics were obtained using the full baseline data, pruning outliers that were more than 3 standard deviations away from the mean. The treatment effect in column B is for the relevant treatment block assuming that we will replace missing households from the baseline (i.e. WNP = 550 households; WWNP = 484 households). We assume the probabilities of type 1 and type 2 errors are 10%, and 20%, respectively.

As shown in Table 5.10, identifying impacts with respect to sourcing of water and to many wastewater management outcomes will be difficult, and it seems unlikely that a large enough sample could be assembled to answer a question of impact on sourcing at the enterprise level. Changes in water costs, perceptions of water as an obstacle or of the sanitation situation in the vicinity of the enterprise, and in revenues, on the other hand, appear more likely to be detectable, even with this small sample.

Table 5.10. Enterprise survey power calculations, based on baseline survey data

	A. Sample size required to identify a 20% treatment effect	B. Size of treatment effect detectable with current sample size (p=0.1)
Water Network Project Outcomes		
Use private piped water	1702	86.5%
Use any piped water	1091	69.7%
Use tanker water	2013	71.1%
Use shop water	842	67.9%
Water is an obstacle to growth	98	21.1%
Water cost is an obstacle to growth	74	18.3%
Log of monthly water cost	106	21.7%
Wastewater Network Project Outcomes		
Neighborhood sanitation rating	68	20.4%
Have wastewater management system	247	35.3%
Connected to WAJ sewer	602	57.3%
Payment for sewer (JD/month)	1192	85.5%
General outcomes		
Log of monthly revenues (JD)	26	12.7%

Notes: Standard deviations, means, and intra-cluster correlation statistics were obtained using the full baseline data, pruning outliers that were more than 3 standard deviations away from the mean. The treatment effect in column B is for the relevant treatment block with the sample sizes from the baseline (i.e. WNP = 90 enterprises; WWNP = 65 enterprises). We assume the probabilities of type 1 and type 2 errors are 10%, and 20%, respectively.

6 ADMINISTRATIVE

6.1 Institutional Review Board clearance

SI houses an internal IRB that is used to review and approve the study before any data collection activity commences. Upon addressing final comments received from MCC and local stakeholders, the evaluation team submitted both the household and enterprise survey to SI's Internal IRB as a terminal step in the survey development process. In the cases of both survey instruments, internal IRB approval was granted within two weeks of submission, whereby approval documentation was submitted to MCC prior to formal survey implementation.

Participation in a local Jordanian IRB was not required; however, the Implementing Entity Terms of Reference ensured that DoS would assume responsibility for receiving IRB approval from a local institution. Key contacts at the local IRB, as well as within DoS, provided extensive guidance on local context, survey timing, instrument design, and potential question sensitivities. Any identified issues were addressed by SI prior to the commencement of field work.

6.2 Data Access, Privacy and Documentation Plan

Given that DoS, a government entity, has have access to potentially sensitive respondent data, SI requires DoS to maintain an encrypted server to house data, accessible only by critical personnel approved by SI. As stated and agreed to by DoS, the Implementing Entity Terms of Reference states:

> All datasets and other data collected shall be the sole and exclusive property of MCA-J, and DoS is not authorized to use the data or derivatives of the data for its own purposes in any form without the express written consent of MCA-J. DoS will be required to securely store copies of all datasets on their own premises during the entire life of the project.

Data from both electronic surveys was entered directly into tablets and stored in a data cloud, with all confidential data encrypted and only accessed with a password. Household survey data was stored in a secure, encrypted online platform managed by SurveyCTO, a data management cloud service affiliated with Open Data Kit. Once saved online, household data was only accessible by members of the SI team. Enterprise data was stored using the web platform Java, a renowned data management system, in which the Jordanian Department of Statistics underwent extensive training. The server housing enterprise data was accessible only to senior members of the DoS field management team, with data sets downloaded and sent to SI directly from the DoS technician assigned to the enterprise survey activity. Tablets used by enumerators were password protected, and select data managers within DoS had access to data (for spot verification purposes) for both surveys prior to server upload.

While the MCC requires an identifiable dataset for their records, all personally identifiable data will be removed prior to data reporting. SI will anonymize data in accordance with MCC guidelines in preparation for final data set submission to MCC's public database.

6.3 Evaluation Team roles and responsibilities

The SI evaluation team has several key personnel that will work together to design and implement the IE, analyze the data, and produce final reports. Team composition is detailed in Table 6.1 as follows:

Table 6.1. Evaluation Team Roles and Responsibilities

Position	Responsibilities
Senior Analyst/ Water Specialist **Marc Jeuland**	Dr. Jeuland serves as the technical and methodological lead. He is heavily involved in the evolution of the proposed IE design throughout consultations with MCC DC staff and MCA-Jordan. Dr. Jeuland leads the IE design and ERR activities, manages any changes to the design required during the implementation process and provides guidance to data analysis, consulting with the Senior Network Engineer, Dr. Albert, as necessary. He contributes to written sections of evaluation reports, and other project deliverables, including serving as lead author of the final IE report.
Statistician/ **Sampling Expert** **Nathan Cutler**	Mr. Cutler advises on statistical and sampling issues. He is responsible for designing and implementing the sampling framework being implemented in the study. He oversees the technical aspects of the propensity score matching and survey sampling design, and consults directly with MCC and DoS staff regarding instrument pre-testing, enumerator training, and piloting.
Research Assistant **Jenny Orgill**	Ms. Orgill supports the technical aspects of the IE, including conducting data cleaning and analysis, as well as providing contributions to deliverables.
Research Assistant **Sabreen Alikhan**	Ms. Alikhan supports the IE team through project coordination, data collection, data cleaning and analysis, contributions to deliverables, and day-to-day management of the overall IE.

References

Albert, J., S. Alikhan, L. Etter, M. Jeuland and A. Wyatt (2013). Impact Evaluation Design Report: Jordan Compact - Water Sector. Washington, DC., Social Impact.

Dehejia, R. H. and S. Wahba (2002). "Propensity score-matching methods for nonexperimental causal studies." Review of Economics and statistics **84**(1): 151-161.

FAO (2015). Aquastat Database. U. N. F. a. A. Organization. Available at: http://www.fao.org/nr/water/aquastat/main/index.stm.

Hashemite Kingdom of Jordan (2008). Water for Life – Jordan's Water Strategy 2008 – 2022. Amman.

Hashemite Kingdom of Jordan (2009). Jordan's Second National Communication to the United Nations Framework Convention on Climate Change (UNFCCC). Amman.

Hirano, K. and G. W. Imbens (2004). "The propensity score with continuous treatments." Applied Bayesian modeling and causal inference from incomplete-data perspectives **226164**: 73-84.

MCC (2009). MCC and Jordan Sign U.S. Government Grant to Develop Poverty Reduction Program. Retrived online (August, 23, 2012): http://www.mcc.gov/pages/press/release/release-060809-jordansignsgrant.

North, J. H. and C. C. Griffin (1993). "Water source as a housing characteristic: Hedonic property valuation and willingness to pay for water." Water Resources Research **29**(7): 1923-1929.

Rosenbaum, P. R. and D. B. Rubin (1983). "The central role of the propensity score in observational studies for causal effects." Biometrika **70**(1): 41-55.

USAID (2011). Institutional Support and Strengthening Program – Institutional Assessment Report. Washington, DC.

Whittington, D., W. M. Hanemann, C. Sadoff and M. Jeuland (2008). "The Challenge of Improving Water and Sanitation Services in Less Developed Countries." Foundations and Trends in Microeconomics **4**(6-7): 469-609.

Yusuf, A. A. and P. Koundouri (2005). "Willingness to pay for water and location bias in hedonic price analysis: evidence from the Indonesian housing market." Environment and Development Economics **10**(6): 821-836.

Annex 1: Reporting of indicators from the MCC/MCA-J Monitoring & Evaluation Indicator Tracking Table (ITT)

MCC and MCA-J requested the help of the Evaluator in providing a number of key statistics for the ITT that is part of the Jordan Compact M&E plan. These indicators and the relevant values from the baseline survey are reported in Table A1 below. We note here that these statistics are drawn from a non-representative (evaluation) sample in Zarqa, but that we do not see significant differences between this sample and the representative sample drawn by the Department of Statistics in 2009, across a number of key variables. For details, see Section 5.2 of this report.

Table A1. Summary of ITT Indicators from the Baseline Household Survey (April/May 2014)

Indicator	**Value in April/May 2014**
Residential water consumption, network and non-network (L/capita-day)[1]	54.1
Use of tanker water (L/capita-day)	Overall: 3.1 Users only: 81.3
Use of shop water (L/capita-day)	Overall: 0.4 Users only: 1.4
Incidence of diarrhea (past 2 weeks)	Overall: 3.1% Under 5 yrs: 7.8% Over 5 yrs: 0.87%
% of water utility customers who think duration OR pressure of supply is "bad" or "not very good"[2]	30.0%
Customer dissatisfaction with water quality[3] Network water safety is "bad" Network water safety is "bad" or "not very good" Network water taste/smell is "bad" Perceived safety of network water (10-point scale) Household treats drinking water because it is unclean	 23.7% 48.6% 24.6% 4.9 27.0%
% of households cleaning large water storage tanks[4]	At least weekly: 20.8% At least monthly: 29.4%

Notes:

[1] This indicator represents the sum of monthly network (7.64 m^3) and non-network (0.33 m^3) water consumption at the household level, divided by household size (4.91) and 30 days.

[2] If respondent indicated that any of these two things were "bad" or "not very good", the household is included in this percentage.

[3] Dissatisfaction with water quality was measured in several ways, 5 of which are reported here.

[4] We measured the frequency of this cleaning in the household survey. The percentage reported here is for at least weekly cleaning and at least monthly cleaning.

Annex 2: MCC/MCA-J Comments on Draft Report and Evaluator Responses

Date: November 14, 2015

Social Impact thanks the MCC and MCA-J for their thoughtful feedback on our baseline report. We summarize changes and responses to the main comments below, which were corrected as noted in the final baseline report. In addition to these responses and changes, we improved 1) the referencing of sources cited in the document, and 2) updated the sample information following determination that the dataset we had been using was not the complete DoS dataset for the baseline household survey in particular.

List of comments and changes (Changes or responses noted in red)

- MCC Comment, Table 3.2: "This table is not described anywhere in the report and is confusing for readers. Please add in a description of the table and from where these statistics come."
  - Response: Actually, Table 3.2 is introduced in Section 3.3.1, under equation 1. But we have discussed this in more detail to respond to the comment.
- MCA-J Comment, Section 3.3.2: "Does this mean 10% is acceptable? And will not affect power of evaluation?"
  - Response: We have clarified that there is no definition of "acceptable" but rather that this was for protection against non-differential attrition and loss of sample size: "Finally, the evaluation sample size target included a 10% buffer for attrition over the life of the study; 313 households were thus added to the final sample, such that approximately 3,440 randomly-selected households were requested of DoS at baseline. Note that the additional sample size to protect against attrition does not represent an acceptable level of attrition, but was purely included to maintain statistical power over the outcomes of interest, and, importantly for internal validity, must be balanced across sample groups."
- MCA-J comment 1, Section 3.3.3: "Selection was from the enterprises sample frame at DOS, right? This sample frame was updated on 2011"
  - Response: Yes, see next paragraph.
- MCA-J comments on enterprise sample in Section 3.3.3.
  - Response: We have clarified several issues as suggested.
- MCC Comment, Table 3.5: "What was the justification for the large sample of Amman control? Was this an artifact of field work?"
  - Response: This is discussed in the paragraph above the table, but a note was also added to the Table. It has to do with the tradeoff between the quality of matches (and thus the need for more control units given differences across sample arms).
- MCC Comment, below Table 3.5: "Is there a benefit to other forms of statistical power? The inference that should be drawn from the technical description is not clear. Ultimately, does this improve the studies ability to detect effects across and between groups? Or not? Or is it more nuanced?"
  - Response: We are not sure we understand the comment, as we describe implications for cross-arm comparisons (15% differences being detectable) as well as overall treated vs. control comparisons in the sentences that follow the comment.

- MCC Comment, Section 3.3.3: "What are the implications for follow up sample size?"
  - Response: This is discussed later in the report, in Section 5.4.
- MCC Comment 2, Section 3.3.3: "Do we need to incorporate an alternate listing exercise to get at informal businesses? Why was a registration approach taken for informals, when by definition they are likely to be unregistered?"
  - Response: We do not follow the comment; there was no registration approach but rather the listing was from household referrals, few of which actually bore fruit. Due to this, we supplemented the household referrals with enterprises identified as informal through a previously administered DoS enterprise survey.
- MCA-J Comment, Section 3.3.5: "This is a condition in MCC Compact, I mean this management contract is a result of MCC intervention."
  - Response: We have added a footnote. It is possible that this is true, but the claim cannot be rigorously verified: "We are aware that some make the argument that the institutional changes were partly attributable to the Compact, but this claim cannot be rigorously tested, since the Institutional Strengthening Program sponsored by USAID was not contingent on the MCC/MCA-J Compact."
- MCC Comment 1, Section 3.3.5: "It would be useful to expand on this technical point with examples that will be assessable to a lay-audience"
  - Response: We provide an example now: "For example, households and enterprises in treated neighborhoods that are located near one another will also experience similar water reliability (e.g., hours of supply per week), while those further away will tend to experience different water supply conditions. These spatial correlation patterns will decrease the amount of variation in the data that is useable for inference about differences between near and far away areas."
- MCC Comment 1, Table 3.6: "Could any adjustments be made with the data from the 2015 census?
  - Response: Perhaps; some new ideas are now listed in the mitigation strategy, in particular: "re-assess and discuss balance in light of 2015 Census data" and "Oversample at baseline to allow for ex post adjustments using 2015 Census data."
- MCC Comment 2, Table 3.6: "Are these measured as controls? Or intermediate outcomes?"
  - Response: We do not understand the comment.
- MCC Comment 3, Table 3.6: "How will this be measured?"
  - Response: We have added this information – "using additional hours of water supply as the "treatment" variable"
- MCC Comment 2, Section 3.3.5: "Did you redo your power calculations for households based on the baseline data?"
  - Response: Yes, please see section 5.4.
- MCC Comment 3, Section 3.3.5: "Please provide greater detail on this strategy and define GPSM"
  - Response: We have added details on this and definitions: "2) use of ex post generalized propensity score matching (GPSM) (Hirano and Imbens 2004) to consider intensity of the water network treatment based on the additional hours of water supply or changes in the percentage of sewer connections in various survey blocks. The intuition here is that households in different areas will be differentially exposed to improvements in

water supply and wastewater service, and that these differences should be correlated with differences in the outcomes of interest."

- MCC Comment, Section 3.3.6: "This makes it sound like only DoS and MCC were responsible for the delays. I would encourage SI to use language that is more general as SI also played a role in the delays and avoids placing blame as both institutions are more likely to post a response letter refuting this when the language is so strong."
  - Response: This was not the intent. We revised the sentence as follows: "Originally planned for spring 2013, SI experienced significant delays in its baseline data collection activity schedule due to a number of complications associated with organizing the fieldwork, including prolonged and fruitful discussions with project partners over the details of the final evaluation design, and challenges in finalizing survey instruments and sampling protocols given the realities of conducting such fieldwork in Jordan."
- MCC Comment 1, Section 4.1: "Please provide a description of these inconsistencies and whether they have implications for measurement in the IE"
  - Response: Information was added as requested.
- MCC Comment 2, Section 4.1: "Please provide a description of these inconsistencies and whether they have implications for measurement in the IE"
  - Response: Details have been updated and added based on the final survey retrieved by Jenny Orgill in September 2015: "There were inconsistencies between the final DoS completion rates, and the completion rates in the final data set provided to SI (see Table 4.2). Specifically, the DoS reporting rates for "closed", "empty" or "refusing" households do not fully explain the difference between the number of households and enterprises visited and the number of surveys completed as recorded in the final dataset. For the household survey the visits minus unsurveyed units is 3492 (compared to 3359 completed interviews in the dataset), while the enterprise survey difference is 15 (425 visits minus 69 unsurveyed yields 356 enterprises, in contrast to the 341 in the final dataset)." Implications are considered in Section 5.4 (power implications).
- MCC Comment, Section 5 intro: "What was the attrition rate planned for? Does the smaller sample cause any power issues?"
  - Response: See discussion in Section 5.4.
- MCA-J Comment, Section 5.1.1: "Including treatment and control areas? Is there any cases within control areas with ww service? Is there any comparison between treatment and control areas for key variables? As we are talking about baseline."
  - Response: We have clarifies that indeed the connection rates are not so different in treatment and control areas (interestingly enough): "These connection rates are indicative of two realities of this system and the partial success of our matching strategy: 1) First, some households do not have private sewer connections even when a sewer line is nearby, due to high connection costs; 2) Second, some households in wastewater expansion areas do already have access to sewer connections; and 3) Our matching strategy was designed to minimize differences (such as this one) across treated and control areas. For the comparison between treatment and control areas, please look at the sample balance tables later in the chapter.

- MCC Comment 1, Section 5.1.1.: "Do we have a sense of how this compares to the census figures? Any way to get a sense of a) quality and b) how representative this sample is of the rest of the country?"
  - Response: See discussion in Section 5.2. We don't think comparisons should be made to the rest of the country, but rather to the rest of Zarqa.
- MCC Comment 2, Section 5.1.1: "Do we have rates on how often households treat their water and how that relates to their confidence in quality?"
  - Response: Information on water treatment is included in Table 5.2, which deals more specifically with water /wastewater status and behaviors.
- MCC Comment 3, Section 5.1.1: "What are the implications of these statistics for water borne diseases?"
  - Response: We have added some discussion of the findings on water quality.
- MCA-J Comment, Section 5.1.2: "I would suggest to have the range for these age and experience variables as the SD is high comparing to other variables."
  - Response: The standard deviation is reported in table 5.5.
- MCC Comment, Table 5.4: "How does this relate to what was expected?"
  - Response: The only data on which we had prior information was purchase of non-network water, and the results for this are not unexpected, as commented on p.31: "The largest category of water-related coping costs is for purchase of non-network water (roughly 6.5 JD/month); this significant expenditure is not unexpected – after all it was a key component of the economic case for the Compact (Albert et al. 2013) – but is striking given the much lower amounts of non-network water that are consumed."
- MCC Comment, Section 5.1.2: "Was this the enterprise sample you were aiming for? How do informal and formal businesses compare?"
  - Response: This issue was discussed in Section 4.
- MCC Comment, Section 5.2.1: "I would like greater detail here or for this question to be addressed earlier. Specifically, how is the sample different? What are the characteristics? Does this change the external validity of the results?"
  - Response: We prefer to leave this discussion here, following presentation of the d-stats, because external validity / generalizability is only one of several issues that merit attention following description of the sample. We have however some details on the main differences to the paragraph: "Thus, our sample does not appear fully representative (it is slightly higher income, households have fewer members, and tanker/vendor use appears lower), but the differences with the 2009 representative sample are minor, and it seems likely that our results will be representative of overall impacts."
- MCC Comment, Section 5.2.2: "Does this alter the learning that can be gleaned from the sample?"
  - Response: We have explained in more detail: "Perhaps most crucially for the enterprise component of the evaluation, the current sample will not allow us to measure impacts on large businesses, unless we expand the scope of the enterprise survey."
- MCC Comment, Section 5.3.1: "This seems like a critical issue on the firm survey. What is the plan to address this in following surveys?"

- Response: We have added a line indicating that more training for enumerators may help, but acknowledging that some of these data are sensitive: "The follow-up survey training will need to emphasize the importance of obtaining complete data, though we know from the fieldwork that part of the underreporting is due to concerns related to divulging sensitive cost or revenue information."

- MCC Comment 2, Section 5.3.2: "How does this alter the analysiss? Are we concerned about internal validity issues?"
    - Response: Yes, as discussed in this section.
- MCA-J Comment, Section 5.3.3: "Could the new building for WAJ be a confounder? Or it will be part of the utility assessment? "
    - Response: We have added mention of this along with general utility changes.
- MCC Comment 1, Section 5.4.1: "What is the follow up plan? What was the projected attrition rate and buffer built in?"
    - Response: We have clarified sample sizes and the extent of the buffer, which is now about 7% (compared to the original 10% projection).
- MCC Comment 2, Section 5.4.1: "Are there additional data quality issues that this brings to the fore? Are there additional mitigation issues that could be taken with dos?"
    - Response: Yes, we have expanded on this: "It may be that additional data cleaning and work with the outliers in the data will help improve power, but these results reinforce the need for additional high frequency data collection, better programming controls (to limit problems with outliers), and more intensive real-time monitoring of data quality by Social Impact and DoS, to improve understanding of the water and wastewater expense dynamics among these households."