IMPACT EVALUATION DESIGN REPORT

JORDAN COMPACT – WATER SECTOR

Submitted to the Millennium Challenge Corporation

by

Social Impact, Inc.

Revised: December 2017

ADVANCING DEVELOPMENT EFFECTIVENESS

IMPACT EVALUATION | PERFORMANCE EVALUATION | STRATEGY, PERFORMANCE & CAPACITY BUILDING

IMPACT EVALUATION DESIGN REPORT

Jordan Compact – Water Sector

December 2017

PREPARED BY

Jeff Albert[1]
Sabreen Alikhan[2]
Luca Etter[2]
Daniel Hudner[2]
Marc Jeuland[3]
Mateusz Pucilowski[2]
Alison Smith[2]
Alan Wyatt[4]

[1] Aquaya Institute
[2] Social Impact
[3] Duke University
[4] RTI

TABLE OF CONTENTS

Acronym List

BCC	Behavior Change Communication
BOD	Biological Oxygen Demand
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
lpcd	Liters per Capita per Day
M&E	Monitoring and Evaluation
MCA	Millennium Challenge Account
MCA-J	Millennium Challenge Account – Jordan
MCC	Millennium Challenge Corporation
MCM	Million Cubic Meters
mg/L	Milligrams per Liter
MLD	Million Liters per Day
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
TSS	Total Suspended Solids
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
ZC	Zarqa Carrier

EXECUTIVE SUMMARY

The Millennium Challenge Corporation (MCC) signed a five-year, $275 million Compact with the Government of Jordan (GOJ) to reduce poverty and increase income in Zarqa Governorate through increases in the supply of water available to households and enterprises through improvements in the efficiency of water delivery, the extension of wastewater collection, and the expansion of wastewater treatment. The Compact entered into force in December 2011, commencing the five-year implementation period scheduled to end in December 2016. The MCC Jordan Compact includes three inter-linked projects:

(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.[5]

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

Social Impact (SI) has been contracted by the MCC to measure the impact of the Compact activities on economic and social outcomes. This Impact Evaluation (IE) design report lays out how the SI team aims to establish a *causal* relationship between program interventions and observed changes in household availability and consumption of different sources of water, household income, household expenditure and household health indicators. It also details our strategy for measuring potential impacts on other sectors (agriculture, utility financial performance, and local enterprises) should these occur in parallel to, or instead of, the expected

[5] Note that the main text of this report does not present a detailed evaluation design for the infrastructure component of the WSH activity, given that was planned on a somewhat different timeframe from that of the other components. A proposal for evaluating this infrastructure component appears in Annex F.

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 will make every possible attempt to measure the impact of the three inter-linked projects separately, in order for MCC to better understand which component(s) of their investment led to specific changes in outcomes. A comparison of the different impacts will further allow for conclusions about the relative cost-effectiveness of each intervention. It must however be noted that because of the complementarities between these different projects, there are very important limitations in the extent to which the IE design will be able to disentangle the separate impacts of the Water Network Project (WNP), the Wastewater Network Project (WWNP), the As-Samra Expansion Project (AEP). Since all three projects are being implemented in Zarqa at the same time, often in overlapping locations, there will be complicated interactions among them which may make it difficult to identify the incremental impacts of each individual intervention. In addition, a significant element contributing to the economic logic of the Compact investments – which we discuss more thoroughly throughout this report – is an assumed water efficiency improvement that would stem from substitution of conventional freshwater currently used in irrigated areas in the Jordan Valley with an expanded supply of treated wastewater collected in Zarqa. The full extent of this assumed substitution effect in fact relies on both the water and wastewater network investments and not on the AEP. It is further assumed that the conventional freshwater saved by this substitution would be made available for higher value uses by municipal and industrial users, thereby improving economic outcomes.

For the purposes of presentation, we have grouped similar and complementary data collection activities into three components, which are described in detail in Section E of this report. This presentation is not intended to imply that any of these three specific components are non-essential; indeed we make the case that all are necessary if the goal is to adequately measure and reduce the risks of misattribution in the overall Compact impact. This point is made following our presentation of the IE logic and a subsequent discussion of the overarching evaluation framework unifying the three components, prior to presenting the details of those components.

Overview of the impact evaluation logic

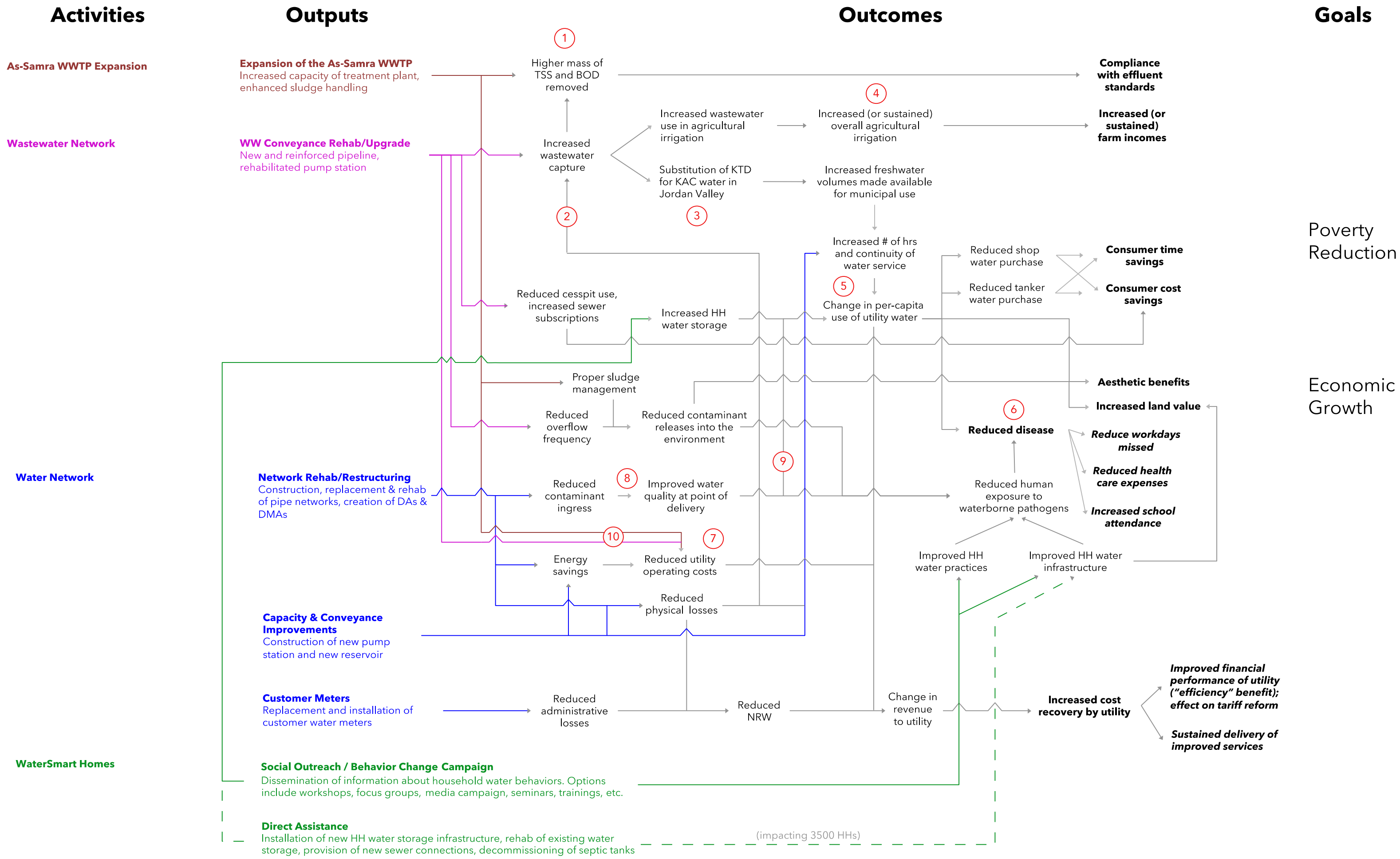
As emphasized in pre-project feasibility studies and economic analyses of the Compact investments, the economic case for the MCC investments rests on a complex and interrelated set of hypothesized changes. The linkages between the various components and intermediate and final outcomes, respectively, are depicted in the IE logic shown in Figure ES.1. It is important to note that Figure ES.1 does not directly follow the categorization of impacts promulgated in previous descriptions of Compact impacts (e.g. 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 in quantifying them. (For example, effects on enterprises and/or on property values were omitted from the analyses – see Section D of this report.)

2) The purposes of the IE logic are a) to trace the relationships between projects, intermediate outputs, and final outcomes, b) to illustrate the overlapping relationships between project activities and desired outcomes, and c) to draw attention to the underlying assumptions.

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 our IE design. 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.

Figure ES.1. Impact Evaluation Logic Diagram



Key:
- Colors correspond to Activity
- Normal text corresponds to intermediate outcomes
- **Bold** corresponds to longer term outcomes
- ***Bold italic*** corresponds to secondary or derivative long term outcomes
- Numbers correspond to relationships that include caveats included below

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 ES.1), 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 (including water tanker/vendor surveys and household surveys with populations that newly settled in Zarqa during Compact implementation) conducted in both intervention and control areas of the Zarqa/Ruseifa conurbation (as well as some similar control areas selected from the Amman Governorate);
2. **Component 2**: Water balance analysis conducted through enhanced data collection of longitudinal inflows to and outflows from the Zarqa and Amman water networks, the As-Samra treatment facility, the King Talal Reservoir, and the complex irrigation network of the Jordan Valley. In parallel: farmer surveys in the Jordan Valley to estimate the magnitude and economic impacts of changes in the availability and utilization of conventional freshwater and blended King Talal Reservoir (KTR) water (which is a mixture of natural runoff and recycled wastewater that is reused for any purpose following treatment to discharge standards).
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.

In this report, these components have been described separately for the purposes of presentation because they consist of different data collection activities that are easier to understand as discrete components. Yet we urge readers to consider how these components fit together for the measurement of the specific and general impacts of the Compact. Recognizing that the mapping of impacts to components of our design proposal is not straightforward, we have developed Table ES.1 to facilitate understanding of how they fit together to answer specific impact questions.

Component 1: Impacts of infrastructure improvements on outcomes in Zarqa (WNP and WWNP)

Since the water and infrastructure projects take place at the same time but in different (yet sometimes overlapping) areas of Zarqa, the main challenge for this component of the IE is to identify areas that are affected by the different activities and compare them with areas that are unaffected. In essence, the simultaneous implementation of the water and wastewater activities means that we are observing four different types of areas in Zarqa where infrastructure and behavioral changes could occur over time: (1) areas that receive both infrastructure improvements; (2) areas that receive the water network or the (3) wastewater network improvements only; and (4) areas that receive no infrastructure improvements. Moreover, the timing of the different activities means that some locations move from one phase to another over the course of the implementation, meaning that the evaluation has to take both geography and timing into account when identifying treatment and comparison areas. We use initial project implementation plans to identify preliminary treatment and comparison areas, which will be amended as implementation plans are finalized.

To measure the effects of the water and wastewater network projects on households and enterprises (Element A of Component 1), the primary approach will be to implement ordinary or generalized propensity score matching (PSM) in combination with difference-in-differences (DiD) and regression analysis. PSM (or GPSM) will be used to predict selection into the various treatment groups (or, if more appropriate, some measure of continuous treatment intensity), using pre-intervention characteristics of those areas. We will then match areas that have similar propensity scores (i.e., that appear equally likely to have received specific exposures to the intervention, based on observable characteristics) to ensure comparability across controls and differentially treated areas, and will conduct balance tests. The DiD design will, in turn, allow us to reduce the threats posed by unobservable differences between affected units that do not vary over time. Households will be surveyed at baseline, during different seasons (summer and winter) during implementation, and finally during the summer after the Compact closes. Next, regression analysis will further allow us to control for factors other than treatment status that may be related to outcomes, thereby increasing 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). Finally, in an effort to address the issue of spillovers, and in order to test the extent to which households in Zarqa are affected by the awareness-raising activities of the WSH campaign, we will aim to include control areas outside of the Zarqa water and wastewater network – for example, areas in Amman that are nearest to Zarqa.

Table ES.1 Relationship between IE design components and the main expected economic benefits of the Compact.

Economic impact question (included in *ex-ante* ERR analysis)	**Data collection components required**
1. What is the economic value of increases in water consumption, reliability, and behavior changes due to the intervention?	Components 1 (household/enterprise, and refugee 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)
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, and refugee 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)
Economic impact question (omitted from *ex-ante* ERR analysis)	
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)

In addition to use of the primary strategy described above, the proposal for Component 1 also includes additional data collection and analysis that is meant to provide greater insight on the impacts of the Compact, using surveys[6] focused on two key groups:

- Element B: A cross-sectional survey of water vendors (shops and tankers) timed shortly after the end of the Compact. The major purpose of this survey is to indicate: a) how this business has changed during the intervention period, and b) whether non-network water sales have been stable, declining, or increasing in time.
- Element C: A cross-sectional survey – complemented by semi-structured interviews – of new (mainly refugee) populations who arrived in Zarqa and Amman during the Compact. The justification for surveying this population is to better contextualize and allow adjustment of the IE estimates, given that population growth has been much higher than anticipated at the time of the design of the Compact. This higher rate of population growth increases water demand, influences nearly every impact channel put forth in the IE logic

[6] The budget for these two surveys was approved by MCC as part of a contract modification and extension in September 2017.

diagram, and may threaten achievement of the original Compact targets. It may also affect the internal validity of the IE, since settlement patterns for this population are not likely to be randomly distributed across the control and treatment arms of the sampled area for Component 1.

The two main challenges for the IE of the water/wastewater projects based on the results of Component 1 include a) the potential violation of the assumption that the expected outcomes of treatment and comparison groups are independent of the treatment assignment – a necessary condition to apply a PSM approach; and b) the inability of this component to properly account for the overall effects of the compact investments – namely those stemming from freshwater substitution from agriculture to urban use, or improvements in utility performance (the most obvious of the potential spillovers described above). To address the first of these challenges, we will use as many factors as possible in the estimation of the propensity score and also take advantage of differences in the timing of treatment exposure, perhaps using targeted areas as controls for areas treated early in the investment program. This latter approach is facilitated by our analysis of data from multiple waves of seasonal surveys. We also have deliberately used conservative estimates in our power calculations that would allow us to drop poor matches *ex post* of baseline data collection (i.e., at the time of analysis), or to increase the precision of our estimates. The second challenge will be partially mitigated by inclusion of control areas located outside of Zarqa; however the larger strategy to account for these relies on the integration of findings under Component 1 with those of Components 2 and 3 of the evaluation, as mentioned above.

Other challenges include the potential lack of statistical power due to the heterogeneity of micro-level outcomes across units and over time (mitigated by the fact that we have been able to conduct power calculations using pre-baseline data on several key variables of interest); and the possibility of non-project confounders such as the Disi water project and population increases. These will provide a large and new influx of water to the Zarqa governorate, on the one hand, and increase demand, on the other; both types of changes co-occur with implementation of the Compact and are likely to have spatially heterogeneous effects (and which we will address as far as possible through careful timing of baseline and other data collection activities, as well as the water balance analysis included in the integration of evidence and data from components 1-3).

Component 2: Impacts on irrigators downstream of As-Samra treatment plant (WNP; WWNP; and AEP)

Analyzing the impacts of the water, wastewater, and As-Samra expansion activities on the agricultural sector poses four principal challenges. First, because of the overlapping causal relationships between water and wastewater activities, respectively, and water made available for agriculture, the effects of the individual project activities on irrigators may not be separable. Second, changes in irrigation volumes must be analyzed in the context of multiple inputs and outputs into a complicated flow regime that connects water users in Zarqa to the production and transport of wastewater from sewer network subscribers (as well as the portions of Amman served by the As-Samra plant), and then in turn to agricultural end-users in the Jordan Valley (see Figure

E.4). Third, the data required for modeling the hydrology of the system in the King Talal Dam (KTD) and the agricultural zones depending on it appear to be limited and of questionable reliability, rendering the aforementioned water balance analysis challenging. Finally, there is no untreated comparison group that can be used to approximate the counterfactual that is Jordan Valley agricultural activities without the expansion.

Given these challenges, the evaluation team proposes two complementary components for the IE of the WWTP expansion. For one, we recommend detailed water balance calculation supported by development of a comprehensive dataset that includes careful measurement and/or defensible estimation of flows into the As-Samra-bound sewer network as well as inflows to and outflows from the Zarqa River upstream of the KTD. Second, we propose to use longitudinal surveys to exploit natural spatial and temporal variation in the quantities of water from different sources that are delivered to farm users in the Jordan Valley. By applying a Difference-in-Difference (DiD) methodology to isolate the effects of changes in the quantities of these different sources of water supplied to a representative sample of farms, we will estimate the treatment effect of an additional unit of water that we can demonstrate to have resulted from Compact activities (specifically the WWNP and WNP). In addition, accounting for the changes in the use of water sources, and of the expansion/contraction of irrigated areas over time, we will attempt to estimate the extent of the primary substitution that actually occurs.

Component 3: Impacts on Water Authority of Jordan (WAJ)-Zarqa Performance

One of the important challenges facing the IE of this program stems from the possibility that some of the benefits may not be directly reflected in welfare changes among households and enterprises in Zarqa, nor among the farmers who may receive additional flows of treated wastewater for their irrigation activities. Indeed, many of the benefits of the investments may be captured by the local water utility, the WAJ-Zarqa, or by other larger government institutions responsible for water delivery in Jordan, including the central WAJ, the Jordan Valley Authority (JVA), or the Ministry of Water and Irrigation (MWI). Benefits captured by these institutions could in turn lead to reductions in public debt in Jordan and free up capital for other productive economic activities nationwide.

For this reason, though we cannot estimate these types of benefits using traditional IE methods, we believe that the IE should at least include careful tracking of utility performance indicators. We therefore propose enhanced analysis of standard indicators of system-wide utility performance, coupled with local-scale measurements and engineering tests that correspond closely to areas monitored using the household and enterprise surveys. This analysis would encompass both existing indicators collected by the Millennium Challenge Account -Jordan (MCA-J)'s Monitoring and Evaluation (M&E) team, as well as a number of additional utility performance and water balance indicators consistent with typical norms for utility management / monitoring best practice. The engineering tests would then allow better understanding of the components of non-revenue water (NRW), one of the main avenues through which MCC expects benefits to accrue (through a reduction in NRW). Finally, these would be combined with an effort to collect and analyze available secondary geo-coded water quality data.

Integration of findings from components 1-3: If the IE is successful, the three components described above should provide the bulk of the information needed to assess both specific elements described in project feasibility and economic analyses (e.g., the primary substitution effect), as well as the larger impacts on development in Jordan, whether these flow primarily through impacts on households, or through more complicated and diffuse channels that may include changes in the cost recovery achieved at the WAJ-Zarqa, increased economic activity in the private (enterprise) sector, or even changes in productivity of irrigated agriculture. Together, these three components will provide estimates of both physical flows of water (in and out of households/enterprises in Zarqa, then to As-Samra and the King Talal Reservoir (KTR), and finally to the irrigated areas of the Jordan Valley) and of the economic value associated with those flows (indicated by changes in water consumption and demand, income, expenditures of time and money, health outcomes, wealth and asset ownership, net value added in the commercial, industrial and agriculture sectors, and/or improved cost recovery and the utility's financial balance sheet). Components 1-3 do not allow analysis of the impacts of the WSH infrastructure support provided to poor households in Zarqa; a design for evaluating such impacts is promulgated in Annex F.

Implementation and Challenges

For each of the three proposed IE components described above, this report details the preferred design option, followed by the main alternatives for that option that were considered. We highlight the required data collection activities, sample sizes or sampling considerations (as estimated using available data or by other means), and IE implementation milestones, as summarized in Table ES.2 below.[7] The main concern moving forward with the implementation of the IE constitutes the fact that implementation plans of most projects and activities are dynamic, and that precise timelines are difficult to anticipate. In the full report, we also highlight a number of other important considerations and data gaps that would need to be addressed in order for this IE to provide the most rigorous evaluation of Compact investments.

[7] The timeline for the IE that is assumed in this report is a duration of 5 years. We understand that this timeline may be extended in the future; however the existing parameters for the evaluation contract are limited to 5 years.

Table ES.2 Summary of Proposed IE Activities (*Italicized dates are planned*)

Component	Evaluation Methodology[a]	Estimated Sample Size	Estimated Timing
Component 1: Impacts of infrastructure improvements on urban households and enterprises in Zarqa (WNP and WWNP)	**Impact Evaluation**[b] Element A: HH / E survey, Sample construction & analysis using PSM + DiD[c] Element B: Water vendor industry analyses Element C: Cross-sectional survey of newly-settled households (refugees) in survey zones	Element A: 3440 households; 345 enterprises Element B: 500 vendors Element C: 1500 households	Element A: HH Baseline: Feb. 2014 HH Midline: Winter 2015; Summer 2016 HH Endline: Summer 2018 Enterprise: 2015, 2018 Element B: *Apr. 2017* Element C: *Apr. 2017*
Component 2: Impacts of Compact on irrigators downstream of As-Samra treatment plant (WNP; WWNP; and AEP)	**Impact Evaluation** Element A: Water balance modeling Element B: DiD methods comparing agricultural production at locations in the JV that do and do not receive reclaimed wastewater	Element A: n/a Element B: 550 farmers	Baseline Survey: Summer 2015 Follow-up surveys: Summer 2016, *2018*
Component 3: Impacts of Compact on NRW, and changes in relative performance of WAJ-Zarqa	Element A: **Performance Evaluation.** Augmented tracking of utility performance Element B: **Impact Evaluation.** Small number of meter tests in areas included in Component 1. Element C: **Impact/Performance Evaluation.** Other geo-coded data collection over areas included in Component 1 (and across Zarqa).	n/a	Ongoing data collection Meter testing: 2015, *2018*

[a] MCC distinguishes between two types of evaluations, 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.
[b] Element 1 is essential; elements in B and C will be conducted if funds allow, but note implications for IE questions (esp. Element B) and internal validity of IE (esp. Element C).
[c] PSM = Propensity score matching; DiD = Difference-in-differences.

MAIN REPORT

Background

The Hashemite Kingdom of Jordan is one of the four driest countries in the world facing a severe water scarcity challenge with declining per capita water resources as a result of population growth combined with decreasing water availability (Hashemite Kingdom of Jordan 2008). The scarcity of water in Jordan is the single most important constraint for the country's future growth and poverty alleviation prospects. It will 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), water availability in Jordan declined from 3600m^3/year in 1946 to 145m^3/year in 2009, well below the international water poverty line of 500m^3/year (Hashemite Kingdom of Jordan 2009). Moreover, the severe water scarcity is a major drain on the Government of Jordan (GOJ)'s fiscal resources with water sector costs 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 on economic activity and poverty are amplified in Zarqa Governorate, a large, dry, mostly urban governorate west of the capital Amman where nearly three in ten households 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 the result of irregular water availability with many households receiving piped water only one or two days per week, and only during a limited number of hours per day.

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. We also recognize that the MCC Compact investment was motivated in large part by the **primary substitution effect**, through which increases in the use of recycled wastewater[8] 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).

Compact Activities

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

---

[8] 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.

(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.[9]

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

Overview of Evaluation Objectives and Questions

The main objective of this Impact Evaluation (IE) is to determine whether or not the interventions of the Jordan Compact lead to changes in poverty and household income, primarily (though not exclusively) among beneficiaries living in the Zarqa governorate. In particular, the IE aims at establishing a causal relationship between program interventions and a variety of observable project-related social and economic outcomes (Table B.1), by comparing the changes experienced over time by beneficiaries (the treatment group) to those experienced by non-beneficiaries (the control group).[10] By carefully developing an IE design based on state-of-the-art methods used in program evaluation, and carefully identifying comparable treatment and control groups, we will minimize the potential for bias in our estimates of project effects.

In addition, through the conduct of three distinct data collection components developed to

[9] Note that the main text of this report does not present a detailed evaluation design for the infrastructure component of the WSH activity, given that was planned on a somewhat different timeframe from that of the other components. A proposal for evaluating this infrastructure component appears in Annex F.

[10] 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.

measure different types of changes, the IE will make every possible attempt to measure the impact of the three inter-linked MCC Compact projects – the WNP, the WWNP, and the AEP – separately. This will in turn allow MCC to better understand which component of its investment led to specific changes in outcomes. Still, there are several general impacts that stem from the collective set of investments rather than from separate projects (e.g., the economic value of the primary substitution effect), and measurement of these impacts will require integration of specific measures collected under the three proposed data collection components. Also, there are inherent and unavoidable limitations in the extent to which the IE design will be able to disentangle the separate contributions of the three Compact projects. Since all three interventions are being implemented in Zarqa at the same time, and will affect overlapping locations, there will be complicated interactions between them, which will make it difficult to identify the incremental impacts of each individual intervention. To make the overlap among interventions clear, the analytical portion of this report begins with a thorough discussion of the pre-project Economic Rate of Return (ERR) analysis followed by presentation of an updated project logic that was developed to guide the design of the IE. This discussion then serves to motivate the methodology section of this report, and clarifies why we do not ultimately group our IE questions by MCC project / activity, but rather by IE data collection components conducted in different places (Table B.1).

In addition to the challenges posed by overlap in interventions, it is very likely that there will be important spillovers from the Compact to areas not specifically targeted by it (both within as well as outside of Zarqa). The presence of spillovers creates a real threat that standard estimates of impact will be biased. Moreover, many of the effects of the Compact may not be felt for many years (beyond the 5-year evaluation period), and there are a number of other interventions being implemented by the GOJ and other donors in the water sector in Jordan and in Zarqa (e.g., Disi water supply; the USAID-funded ISSP water sector reform) that will affect and potentially confound the outcomes of interest. We discuss each of these issues in this report, and though our IE design cannot fully eliminate the threats they pose, we believe that it strikes a good balance between diminishing the severity of the threats they pose and maintaining the overall integrity of the design.

As with all 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.

Potential Contribution to Economic Development and Poverty Reduction Literature

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, as laid out in the literature review in the next section, there is 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 an unprecedented 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.

Table B.1. Main evaluation questions by Impact Evaluation Component, and general integrating evaluation questions

Component	Evaluation Question(s)	Methodology	General outcomes
Component 1: Impacts of infrastructure improvements on urban households and enterprises in Zarqa (WNP and WWNP)	• **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)? • **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? • **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? • **Impacts on income**: Does the WNP change HH / E income? • **Impacts on asset value**: Does the WNP / WWNP affect property/asset values? • **Overall impacts on welfare in Zarqa**: What is the net economic value of changes in quantity and quality of water consumed? [c]	**Impact Evaluation**[a] • Element A: HH / E survey, Sample construction & analysis using PSM + DiD[b] • Element B: Water vendor industry analyses • Element C: Refugee survey	• Water reliability and consumption • Sewerage coverage; sewage collection • Water quality & health • HH/E expenditure on water and wastewater management (time and money) • Shop and tanker sales • HH/E Income • Property Values
Component 2: Impacts on irrigators downstream of As-Samra treatment plant (WNP; WWNP; and AEP)	• **Impacts on water sourcing**: Does the combined WNP/WWNP/AEP result in increased irrigation with addition blended KTR water? Does the volume of irrigation using KAC freshwater correspondingly decrease? • **Impacts on farming costs**: Does the combined WNP / WWNP / AEP lead to changes in farm input costs? • **Impacts on farm output**: Does the combined WNP / WWNP / AEP lead to changes in the value of farm output in affected areas? • **Impacts on asset value**: Are farm values affected by the WNP / WWNP / AEP investments? • **Overall impacts on farm welfare**: What is the net economic value of changes in irrigation? • **Impacts on compliance**: Does the AEP result in increases in the quantity of wastewater that meets effluent standards prior to discharge into the environment?	**Impact Evaluation** • Element A: Water balance modeling • Element B: DiD methods comparing agricultural production at locations in the JV that do and do not receive reclaimed wastewater	• Water balance indicators • Farm water use (by source) • Irrigated area and production • Farming costs • Farm incomes • Property values • Irrigation water quality
Component 3: Impacts on WAJ-Zarqa	• **Impacts on utility cost recovery**: Does the net cost recovery of the utility improve due to the Compact, and is this related to service improvements? • **Operations and maintenance**: What is the impact of the Compact on the budget and execution of O&M?	**Performance / Impact Evaluation** Element A: Augmented utility tracking	• Utility performance (losses, pipe breaks, NRW, cost recovery indicators)

	• **Service improvements**: At the utility level, are there measurable changes in service delivery quality trends in Zarqa relative to those of other municipal utilities in Jordan?	Element B: Small # of meter tests Element C: Other data collection	
Integration of component s	• What reallocation of water is made possible by the Compact investments? ('**primary'** substitution effect)? • What is the comparative economic value of water consumption for different uses (domestic, commercial/ industrial, irrigation)? • What are the overall net benefits from the Compact, and what are its distributional consequences?[c]	Detailed water balance and distributional analysis	• Volumetric water allocations • Valuation of water quantities in different uses

[a] Element A is essential; elements in B and C will be added if it is determined that their scope is appropriate and their importance justified.
[b] PSM = Propensity score matching; DiD = Difference-in-differences.
[c] If the elements in B are not funded, this question cannot be fully answered.

Literature Review

The primary goals of the activities comprising MCA-J's water sector interventions are to reduce poverty through economic growth in Zarqa Governorate. These goals are to be achieved by increasing the supply of water available to households and enterprises through improvements in the efficiency of water delivery, the extension of wastewater collection and the expansion of wastewater treatment. The underlying project logic in particular assumes that improvements made to water infrastructure at the network and household level, as well as corresponding public outreach and household technical assistance, will improve socioeconomic and health indicators among Zarqa Governorate households and farmers. The limited availability of studies pertaining directly to the hypothesized linkages between water and development, and more specifically focusing on urban populations similar to MCA-J's target population, makes it difficult to rely on past evidence and experiences to make predictions that are directly relevant to this intervention. Likewise, there is scant literature on the impacts of large-scale water interventions in middle-income settings similar to those envisioned by MCA-J. Much of the literature on the impacts of water and sanitation improvements detailed below, therefore, pertains to populations lacking improved WASH services and to interventions targeted at the households.

From a broader network level, the literature holds that investments in urban water supply lead to lower input costs, to which firms using these improved services respond with expanded production and employment, reduced prices, and investment of savings in other economic activities. This is consistent with the theory behind the Water Network activity logic that increased water network quality and water access/availability is a driver of economic gains to the private sector. Gains at the enterprise and industry level ultimately translate into increased production and income at the national level. Further, water supply investment is likely to reap the greatest benefits where small distribution systems can be expanded without exceeding current production capacity to cover a broad geographic area servicing existing users in urban and peri-urban centers. Also, systems considered technically inefficient are the best candidates for investments to increase quality and quantity of water. Factors that should determine investment decisions are the volume of water used in production by existing enterprises, the likelihood of high-volume users locating to the area, the current price and quality of alternative supplies, and the size and location of the market for additional goods to be produced (Schwartz and Johnson 1992).

Recent literature using households as the unit of analysis offers some support to the Water Network Project's assumption that increased availability of network water can lead to significant cost savings (including reduced time and health costs), though most such studies pertain to rural areas in low-income developing countries. For example, a study on valuation of the time spent collecting water by households in rural Kenya found it to be approximately equivalent to the wage rate for unskilled labor (Whittington, Mu et al. 1990). Furthermore, time spent walking to a household's main water source was found to be a significant determinant of health among children under age 5 across Sub-Saharan African populations lacking access to piped water (Pickering and Davis 2012). A 15 minute decrease in one-way walk time to water source is associated with a 41% average relative reduction in diarrhea prevalence, improved anthropometric indicators of child nutritional status, and an 11% relative reduction in under-five child mortality. Evidence also indicates that increased water access can have positive economic benefits, particularly by increasing household savings and freeing up funds and time for other pursuits (Galiani, Gonzalez-Rozada et al. 2009). The extent of the economic benefits of improved water supplies depends on the characteristics of existing and improved sources, such as price, reliability, convenience, and quality (Cairncross and Kolsky 1997). A major determinant of benefit is the reliability of both existing and improved supplies. Also important is whether the improvements can be sustained over time while keeping costs to beneficiaries low; the suppressed water rates commonly levied throughout MENA, for example, are closely tied to water shortages and reduced performance of utilities over time (Bucknall, Kremer et al. 2007; Jeuland 2012).

Women in particular are often the primary beneficiaries of water supply interventions, as they stand to gain from the take-up of income-generating opportunities and education enabled from time savings due to increased access to water (Cairncross and Kolsky 1997). Such productivity benefits are likely lower in urban areas relative to rural settings; however even in Zarqa supply disruptions that require staying at home (to monitor taps), or that require travel to obtain water from shops or other out-of-home locations, may fall disproportionately on women. Devoto et al. (2011), however, found that while greater access to clean water saved households a significant amount of time from having to gather water from elsewhere in urban Morocco, this time was spent primarily on leisure and social activities. Also, little research has been conducted on the explicitly negative impacts of projects to increase water supply, such as the loss of water-vending work (private sector), and opportunities to socialize while gathering and hauling water.

Overall, there are mixed findings on the impact of increased water availability on improved health at the household level. One of the major challenges plaguing observational studies of the impacts of improved water and sanitation services on outcomes is that households with improved services tend to be systematically different from those without them (in terms of socio-economic status (SES), risk-altering behaviors, and unobserved preferences for health), rendering comparisons of those with and without access suspect. The literature on the effect of increased water volumes highlights the fact that multiple channels influence the incidence of diarrhea, and that many of these are unrelated to water access, such as hygiene practices, contamination problems related to in-house water storage, sanitation, and exposure to food-borne pathogens (Zwane and Kremer 2007; Waddington, Snilstveit et al. 2009). Early systematic reviews of water supply interventions suggested a weak link between improved household water quality and diarrheal disease control (Esrey, Potash et al. 1991). In one more recent study from South Africa, household water quantity did have a more important positive impact on health than water quality (Lewin, Stephens et al. 1997). Yet recent randomized trials and systematic reviews of evaluations of water supply improvements in less-developed countries do not generally support that result (Fewtrell, Kaufmann et al. 2005; Waddington, Snilstveit et al. 2009). Among the types of improvements considered by Fewtrell et al. (2005), for example, water supply improvements were least effective in reducing diarrhea on average, compared to hygiene, sanitation, and combined interventions. It is also generally unclear from such studies whether improved access to water (i.e. lower cost) leads to increased consumption. There is somewhat more positive evidence linking expanded piped water access to health, especially among populations with the highest mortality rates (Galiani, Gertler et al. 2005; Gamper-Rabindran, Khan et al. 2008). At a broad water network level, investments in water supply systems have also been found to reduce the incidence of water-borne disease (Jalan and Ravallion 2003).

Water quality improvements, on the other hand, have been much more strongly related to health benefits, though results vary considerably across studies (Fewtrell, Kaufmann et al. 2005; Waddington, Snilstveit et al. 2009). Cutler and Miller (2004) use the natural experiment arising from differential timing of the introduction of chlorination in large cities in the US to link water treatment to reduced rates of typhoid and cholera. A review of different types of water quality interventions by Clasen et al. (2006) found that household-level interventions were more effective for improving water quality and reducing diarrhea than community-level source improvements (e.g. improved wells, installation of handpumps, spring protection), a view that is supported by Waddington et al. (2009). Increased water use by families gaining piped water access in urban Morocco similarly did not change the incidence of waterborne illness, perhaps because transmission of diarrheal disease in target communities through drinking water was low, or otherwise because of poor maintenance and condition of network infrastructure (Devoto, Duflo et al. 2011). A study in rural Jordan highlights the difficulty in pinpointing the origin of diarrheal disease; in 35% of study cases, the etiologic agent could not be determined despite the wide prevalence of diarrheal disease, suggesting that diarrhea-causing pathogens circulate easily through the population, but not necessarily through water (Nimri, Elnasser et al. 2004). Esrey et al. (1991) similarly found that while better water quality had a significant impact on proximal hygiene indicators, effects on diarrheal incidence were modest. In general, the effectiveness of water quality interventions for preventing diarrhea may be more closely related to compliance with the intervention (behavior change) than the specific intervention type (Clasen, Schmidt et al. 2007).

With regards to improvements in sanitation, large-scale systematic reviews again suggest a potentially larger impact on health than for increased water availability. Unfortunately, most of the evidence on the effects of sanitation on health comes from rural studies that focus on the transition from open defecation to latrine use (as opposed to the shift from on-site sanitation to sewerage). The systematic reviews previously described find a roughly 30% reduction in diarrhea from introducing improved sanitation among rural households (Fewtrell, Kaufmann et al. 2005; Waddington, Snilstveit et al. 2009). Recent rigorous IEs find similar effects, but note that subsidies and persuasion are frequently required to increase adoption (Pattanayak, Yang et al. 2009; Patil and Pattanayak 2010; Pattanayak, Poulos et al. 2010). Whether the health and other effects of moving to sewered sanitation systems from on-site excreta disposal in densely populated urban areas where the Jordan Compact activities are targeted would be similar to the effects of moving from open defecation to on-site sanitation in rural areas is unknown. The question is of great importance, however; in Jordan for example, recent data suggests that only 88% of domestic wastewater is estimated to be collected in sewers, and only about half of this amount (47%) is treated in wastewater treatment plants (Jeuland 2012).

Another relevant finding in the literature on water and sanitation interventions is that individual interventions seem to deliver diarrheal disease reductions of 30-40%, but combined interventions offer little additional benefits (Fewtrell, Kaufmann et al. 2005; Waddington, Snilstveit et al. 2009).

This is puzzling because such interventions often target different channels of contamination, but it may result from coordination problems during implementation, compensating behaviors, or misattribution of benefits to individual interventions (Whittington, Jeuland et al. 2012). The quality of the underlying studies indicating this surprising lack of additive effect has also been questioned (Eisenberg, Scott et al. 2007). The evidence does not necessarily rule out additional benefit from combined interventions, but it does raise questions about whether the additional cost of integrated approaches is warranted on the basis of health gains alone.

The added benefit of education – relevant here because of the outreach component of the Water Smart Homes activity – on health outcomes for piped water interventions was explored by Jalan et al. (2003): results indicated that health gains at the household level largely bypass children in poor families, particularly when the mother's level of education is low. Low maternal education level was similarly found to be a significant risk factor for childhood diarrheal incidence among rural populations in northern Jordan (Nimri, Elnasser et al. 2004). In this case, combined water and sanitation improvements may be more effective than single interventions (Lewin, Stephens et al. 1997); further, Newman et al. (2002) conclude that investments in small community water systems had no major impact on water quality until combined with community-level education, strengthening the argument for the addition of an education component to large-scale water interventions. There are a growing number of studies showing that information provision can lead to beneficial changes in behavior, at least in the short term (Luoto, Levine et al. 2011; Bennear, Tarozzi et al. 2012; Hamoudi, Jeuland et al. 2012). This evidence bolsters the theory driving the planned education and behavior change strategies of the Water Smart Homes activity. On the subject of behavior change strategy, Mosler (2012) found that behavior change communication (BCC) interventions in the water and sanitation sector are most effective when they are framed as positive gains – "staying healthy" – rather than negative consequences – "preventing disease", but other studies have found contrast framing – combining positive and negative messaging, to be even more powerful (Luoto, Levine et al. 2011).Overall, preferred attributes of water should be used to promote water treatment, particularly as part of a cluster of health and hygiene behavior.

Another important expectation, and key outcome indicator, of the Water Network's infrastructure activity is reduced demand for and usage of water sold through the private sector. While the expected increase in demand for network water has negative economic implications for private suppliers, MCA-J's water interventions aim to impart economic gains on households (and enterprises) through cost savings resulting from increased access to network water. Literature and observational evidence supports the general assumption that improved water availability leads to greater water consumption, but the extent to which households are willing to substitute vended water for network water may be determined by a number of factors unrelated to structural interventions. A study of water vending and willingness to pay in urban Nigeria emphasizes the significance of perceptions of water quality: people were willing to pay water vendors over twice the operation and maintenance costs of piped water for what they perceived to be higher quality water, despite vast structural improvements made to the piped water system (Whittington, Lauria et al. 1991). Demand for water supply improvements among small enterprises in two Ugandan towns has similarly been found to be limited (Davis, Kang et al. 2001). Private water vendors may continue to claim a large market share unless improvements to the structure – as well as water quality – of the network are effectively advertised throughout the intervention's target population. Little research exists on the scale and magnitude of water vending activities in metropolitan areas of developing countries; qualitative data collected as part of this evaluation will serve to fill some of these gaps in order to inform the design of future water interventions in Jordan and throughout the Middle East.

Examined through the lens of the economic analysis justifying the Compact, success of the collective of projects included in the compact largely hinges on the effective substitution for irrigation of recycled wastewater for currently-used freshwater supplies. The conditions leading to this type of substitution on a wide scale are not well understood, and therefore are hard to predict for the Jordanian context. In Jordan, reuse of mixed water is already fairly well developed and enabled by a combination of heavy water subsidies and a lack of choice over irrigation water source. Kijne et al. (2003) offer an optimistic vision in which farmers in water-scarce countries will adopt new practices to improve water use efficiency to the extent that they are involved in the development of project strategy at the outset. Studies in the Middle East and India additionally indicate that acceptance of new irrigation methods and resources largely depend on farmers' attitudes toward production risk, perceptions of the risk, and potential profitability. In effect, a number of water-scarce countries, including Jordan, have made significant efforts at promoting wastewater reuse in agriculture, often with relatively limited success due to lower demand for recycled water that is high in salinity or of unknown microbiological quality (Jeuland 2012). To the best of our knowledge, there have been no IEs of wastewater reuse interventions in areas where such water has replaced freshwater supply.

One of the primary reasons why the IE literature on the effects of regional or urban water and sanitation infrastructures is so thin is that such studies are difficult to conduct in an experimental or quasi-experimental framework. Additionally, the literature cited in this review has noted important limitations that make it difficult to estimate the true impact of large-scale water interventions. First, due to the nature of the scale of intervention, specific impacts at the household level are difficult to attribute to large, multi-pronged activities that represent in fact a package of interventions. Specifically, observed differences in outcomes for beneficiaries of large-scale interventions may arise from a combination of factors, some of which are unrelated to the intervention itself, making the true impact more difficult to tease out. Moreover, a distinct feature of urban water interventions is the partial coverage of populations; in many cities, certain clusters of households are connected to pipes, and certain clusters receive water from a variety of sources, such as tanker and shop water (Lokshin and Yemtsov 2003). In addition to this, access to services may not always coincide with high quality and reliability, particularly in the long term (Zérah 1998). Second, indicators used in many IEs are heterogeneous and targeting of infrastructure interventions is a blunt instrument, so a traditional IE may have limited ability to capture certain phenomena surrounding water use. For example, a low-income elderly household for which tanker or shop water has always been prohibitively expensive would not expect to report any increase in female wage employment or income savings as a result of greater availability of piped water.

Summary of the Program Impact Logic

D.1. Discussion of the Economic Rate of Return Analysis

To better understand the goals of the IE, we have analyzed the relationship between 1) the expected impacts of the MCC-Jordan Compact as assumed in the Economic Rate of Return (ERR) calculations, and 2) the general IE logic as presented in Figure 4. It is important to note both the included and omitted categories of the ERR as detailed below, and to attempt to develop strategies for measuring the outcomes or their proximal indicators.

D.2. ERR Categorization of Benefits from the Compact

The ERR calculations were conducted separately for a) the water network investments and b) the wastewater network + As-Samra expansion investments.

Water network investments

Benefits were grouped as:

1. Water "efficiency" benefits, resulting from steep declines in non-revenue water (NRW), from 57% to 19% overall, following the network rehabilitation, which would decrease the cost of water supply relative to the alternative source for this extra water, i.e. the Disi project;
2. Consumer savings, from substitution of network water for tanker and shop water;
3. Health benefits for households, due to increased consumption of water.

It is useful to consider some of the key assumptions underlying these benefit categories, and to assess the degree to which they can be measured by the IE.

1. *Efficiency benefit***.** There are three key assumptions underlying this ERR calculation. First, it is assumed that none of the 38% (57-19%) of NRW that would be eliminated (and would thus consist of extra water delivered to beneficiaries) is actually currently consumed by households (i.e. all of this reduction is physical loss). Since some physical losses (perhaps 10%) are probably inevitable, the implication is that only about 9% of the 57% (of the NRW consists of administrative losses (consistent with the estimate of 7% in consultant reports). If this did not represent additional water made available to households, it would be inappropriate to consider that the cost of the water would be saved relative to the next best alternative (Disi water). It is also worth noting that a method we have applied to data from the Zarqa water system, outlined in Annex D, also indicates a level of NRW of 57%; however, this is composed of 33% physical losses and 24% administrative losses, not the consultant estimate of 50% physical losses, and 7% administrative losses. The latter set of figures implies that residents in Zarqa are actually using more water than previously measured, and that the potential gains from reducing physical leakage are likely to be smaller than anticipated.

Second, and perhaps more critically, it is assumed that the extra water that is produced for consumption in Zarqa by reducing NRW would have been sourced from the relatively more expensive Disi project in the absence of the investment (as shown in the alternative cost valuation in the ERR analysis). In reality, it is quite possible that this alternative cost is irrelevant. For example, if water would not have been supplied to Zarqa from Disi, then there would be no "efficiency" gain to speak of. In such circumstances, the MCC investments still might produce benefits in the form of reduced losses to households desiring additional water. The relevant economic measure would then simply be the willingness to pay for increased water supply. In the absence of supply augmentation from Disi, the latter reduced losses to households at least partially overlap with the consumer savings that are included as the second benefit item from the water network component. This is because the additional water that would be supplied to households due to reduced physical losses, now at lower cost to households, would be partially used to offset tanker and shop water purchases. This additional water might also provide other benefits as well, in terms of productivity due to greater quantity, reduced time costs associated with acquiring water, or other lifestyle benefits, etc.

Third, the cost of supply of Disi water incorporated into the ERR calculation is 1 JD/m^3, but this may be too low (Al-Salaymeh and Al-Salaymeh 2008). As far as we understand, this estimate was based on data from the engineering firm doing the initial feasibility analysis on the basis of estimated production of 100 million cubic meters per year, though other sources suggest the water yield will be lower, which would increase costs (Puri, Wong et al. 1999; Jasem, Shammout et al. 2011).

Implications for the IE: On a Zarqa-wide basis, the IE will not be able to definitively determine what would have happened to water supply in the absence of the MCC investments (this pure counterfactual does not exist). We can, and should, however, attempt to measure the demand for additional water at baseline and endline, perhaps using stated preference methodologies and by analyzing revealed demand for the various available sources, in order to provide a lower bound that corresponds to the costs of water rationing. Through the combination of sampling in treatment and control areas (both of which will be affected by Disi, but only the former of which will be affected by the water network improvements), and the integrating water balance analysis we propose to include in the evaluation, we can also observe how changes in overall water supply coincide with inflows of Disi water into Zarqa (assuming such data can be obtained from WAJ-Zarqa). Such water flow measurements should be carried out over the duration of the Compact, to will provide an indication as to whether the alternative cost (of additional Disi water) is really relevant in this case. Finally, to assess the accuracy of assumptions about NRW, it would be useful to contract an engineering firm to conduct some forensic auditing of leakage (night flow tests) in randomly selected portions of the system that overlap with our household and enterprise data collection activities, before and after the improvements are made.

2. *Consumer savings.* The ERR estimates of consumer savings are based on a complicated set of assumptions regarding the nature of demand for different sources of water in Zarqa, assumptions which cannot be fully validated without careful empirical study. Nonetheless, the logic behind these savings is clear and compelling. Households spend much more per unit of water for shop and tanker water than they do for network water. Therefore, if they substitute away from the former supplies, there will be significant savings. Assuming that the demand curves underlying the calculation are derived from sound econometric analysis, the only potential

problem arises from erroneous estimation of the substitutability of water supplies. For example, if there are important differences in quality between source (e.g. shop vs. network water), and if these have not been considered in the demand analysis, the substitution may be much less than expected.

Implications for the IE: We can measure the degree to which consumption of network, shop, and tanker water changes over time among households (and enterprises, though these were not included in the ERR analysis) as they are differentially exposed to increased delivery hours from the network, assuming that a sufficient number of meters are functional. This will allow us to assess whether the expected substitution patterns emerge in Zarqa. Note however, that our estimates may be biased downward if there are spillovers to unimproved areas, simply because there is increased water availability throughout the system that affects all zones serviced by the utility.

3. *Health benefits.* The health benefits calculation in the ERR is based on an assumption that increased consumption of water among poor households with higher disease rates will lead to significant savings in terms of health treatment costs and productivity losses. The estimation is based on comparison of high and low consumption groups from a cross-sectional study, and is inconsistent with the global experimental or quasi-experimental evidence on the health benefits of water supply augmentation, however. It does not consider water quality aspects.

Implications for the IE: We will measure self-reported health outcomes in our household survey. However, powering the study to measure health effects given the low incidence of diarrheal disease in Jordan would entail significant costs, even if such effects were thought to be revealed (based on other findings in the literature as well as our own power calculations based on data of the 2009 Jordan Water Survey presented in Annex C).

<u>Wastewater network + As-Samra expansion investments</u>

These benefits were grouped as:

1. Avoided contamination of agricultural lands with untreated wastewater due to insufficient treatment capacity at As-Samra;
2. Benefits to households from new connections to the sewer network, in the form of cost savings from new cesspit construction and routine pumping of cesspits;
3. Benefits to households from new connections to the sewer network, in the form of reclaimed land;
4. Savings on utility net maintenance costs to the utility in the form of increased wastewater tariff revenue that compensates for increased maintenance costs due to enlarged network;
5. Substitution benefit to urban areas receiving additional freshwater due to the supply of treated wastewater to agricultural producers in the Jordan Valley;
6. Incremental irrigation added value, due to increased water flows to agriculture; and
7. Agricultural benefits in the form of maintained citrus and other high value irrigation, as well as increased water supply to farmers.

It is again useful to consider some of the key assumptions underlying these benefit categories, and to assess the degree to which they can be measured by the IE.

1. *Avoided contamination benefit.* This estimate is based on the alternative cost of wastewater treatment, for the wastewater flows that would be generated as a result of the water supply improvements (greater consumption of water in Zarqa, and therefore production of wastewater) as well as population and demand growth over time. Without the wastewater project, the West Zarqa pumping station would not be able to pump flows in excess of 85000 m^3/day, an amount that is projected to be surpassed in 2014. The critical assumption is therefore that the alternative wastewater investments would be made, and that they are priced correctly; otherwise the avoided contamination benefit is instead the real damages averted from insufficient treatment.

Implications for the IE. This is not a traditional IE question since the counterfactual of contamination or alternative treatment cannot be observed. The amount of additional wastewater that would be produced in the absence of the sewer expansion to new areas is also unobservable. What the M&E and IE should track, however, is the point at which the sewage pumping capacity improvements become necessary to evacuate the new wastewater flows from Zarqa. The evacuation of this excess is made possible by the MCC wastewater investments.

2. Benefits to households in the form of cost savings. This benefit (actually a net cost) is the relative spending on current cesspits and pumping as compared with the wastewater connection and tariff. The inclusion of the wastewater tariff, however, appears to be an error in the ERR calculation since it is actually a transfer from consumers to the utility that should already be included in the cost of maintenance that appears in the project cost stream. It is assumed that all households in the serviced areas will connect, as mandated by law.

Implications for the IE. Despite the mandate to connect, households may choose to remain outside the network given the costs involved, at least in the short term. Household surveys should help us to determine connection rates and assess the balance of consumer savings and costs.

3. Benefits to households in the form of reclaimed land. This benefit is based on the value to property owners of land reclaimed from being a cesspit, and is based on a land cost of 35 JD per square meter and an affected area of 30 m^2 per land parcel.

Implications for the IE. The value of land reclaimed is likely to vary substantially over the affected zones and over time, but is something that our evaluation study could aim to assess. The estimate in the ERR may be too low, given that sewage connections likely provide other benefits in addition to simple land reclamation (aesthetic improvements over the entire property, home value reflecting access to reliable infrastructure and lower future costs for cesspit renovation or mitigating environmental damage, etc.). If we collect extensive property values in the IE, the various mechanisms for enhanced property value will not be something we can determine, but aggregate benefits could be determined using hedonic valuation models (including category 2 above and any aesthetic / productivity benefits from enhanced household water and wastewater infrastructure as detailed below).

4. *Savings in utility net maintenance costs.* This small benefit (1 million JD per year) comes from reduced O&M at rehabilitated pumping stations. Also, revenues from increased wastewater collection are assumed to cover increased costs of expanded sewerage.

5. *Primary substitution benefit: Freshwater transfer from irrigated agriculture to urban consumers.* This is a significant assumed benefit of the wastewater and As-Samra component of the project.

Implications for the IE: The IE will include a water balance analysis that will aim to measure how the balance of freshwater and blended KTR water (and therefore, recycled wastewater) delivery to the Jordan Valley changes over the course of the project, and to assess the implications of those changes for water supply augmentation in urban areas. To understand the economic benefits of these changes, the IE will include an integrating analysis – based on data collected through the household, enterprise, and agriculture surveys – that will update and compare the willingness to pay (WTP) for water in different uses. These estimates will update those cited in the ERR calculation (Haddadin, 2006). However, it should be recognized that it will be difficult to definitively ascribe the causality of this change to the Compact investments, since national policy is also prioritizing this type of agriculture to urban substitution. In other words, it will not be possible to observe whether the government of Jordan would not have found other ways to make this substitution (and what the costs of those alternative methods would have been) in the absence of the MCC Compact. Attributing the change in volumes of treated wastewater reaching Jordan Valley farmers to specific Compact activities will also prove challenging, given the fact that the water and wastewater projects both lead to increased volumes of collected wastewater. In addition, there is an array of confounders that must be incorporated into the water balance assessment, particularly with regards to the natural variability in the hydrological system upstream of the King Talal Reservoir.

There is also an important assumption in the ERR calculation that suggests that the expansion of the As-Samra treatment plant is necessary for this water substitution benefits to take place. We note that the expansion by itself will have no practical effect on wastewater volumes flowing from As-Samra into the lower Zarqa River and eventually into the King Talal Reservoir and to the farms of the Jordan River Valley. These increased volumes instead come from the combination of the water and wastewater network improvements. The expansion of the plant's treatment capacity will rather help ensure that the facility's effluent continues to meet international sewage effluent standards. In the absence of this expansion, the facility would continue to receive increases in sewage influent from its service areas due to increases in population, per-capita water use, and water deliveries that might result from the Disi project, as well as in wastewater capture that could result from the wastewater and water network activities of the Compact. If and when the influent volumes (and specific loadings of suspended solids and biochemical oxygen demand) exceed the treatment plant design capacity, the result will be increased concentrations of TSS (Total Suspended Solids) and BOD (Biological Oxygen Demand) in the treated effluent. The consequences of these effects are likely to be marginal for the purposes of agricultural reuse. Even if the As-Samra facility were to be completely shut down due to technical failure, sewage would still be routed downstream towards the KTD and the Jordan Valley (albeit with no TSS or BOD removal whatsoever and the violation of Jordanian effluent standards). We note that the earlier pond treatment system at As-Samra that preceded the current activated sludge facility operated well above its design capacity for decades. Furthermore, the current system is already operating above capacity for BOD (meaning that the mass of BOD entering the facility each day exceeds the mass for which the system was designed in order to achieve a target level of 30 mg/L in treated effluent).

We are aware that there has been an argument made that the maintenance of effluent quality standards will lead to agricultural benefits, but we urge caution on this point. The water quality parameters, which are of greatest importance with respect to agricultural reuse, are salinity, algae/phytoplankton, microbial pathogens, and suspended solids. The treatment plant expansion is targeted at suspended solids and BOD, which are the parameters used to measure treatment plant performance with the objective of protecting aquatic ecosystems. Increases in TSS and BOD that would occur in the absence of the treatment plant expansion are unlikely to lead to concurrent increases in the contaminants of importance for agriculture downstream. TSS tends to decline dramatically at the KTR from natural settling, while algae/phytoplankton tends to spike (regardless of changes at As-Samra). Meanwhile, there is natural die-off of much of the bacterial population due to UV exposure and other natural process between the As-Samra WWTP and the irrigation system in the Jordan Valley. Salinity, perhaps the contaminant of the greatest concern to downstream farmers, is not removed at all by the treatment processes at As-Samra, and actually increases as evaporation occurs at the King Talal Reservoir. Finally, the dilution processes that occurs as the treated effluent is mixed with naturally occurring streamflow in both the lower Zarqa and within the King Talal Reservoir leads to contaminant reductions that may well drown out any benefits from the increase in effluent quality that result from the As-Samra WWTP expansion.

6. *Incremental irrigation added value.* There will be additional water (net of substitution) due to the wastewater network and treatment plant expansions. This water is valued at \$0.33/m^3, and it is assumed that 75% of the water makes it to the farm level.

Implications for the IE. The water productivity assumption can be assessed through careful agricultural surveys which trace water source to production types and profits. In assessing this category of benefits (and the next one discussed immediately below), we will have to conduct pre- and post-project assessments of farm-level water availability, to determine whether additional water is actually used in irrigation or not, and which kinds of cropping transitions occur.

7. *Retained agriculture benefit.* This is a very large benefit in the ERR, and it rests on three important assumptions. First, it is assumed that without the As-Samra expansion project, untreated wastewater would flow to areas irrigated in the Jordan Valley and lead to their steady and complete conversion (over about 5 years starting in 2014) to low value agriculture. This assumption is inconsistent with our understanding of the operation of the facility and the fate and transport of pollutants between effluent discharge at the plant and distribution to Jordan Valley farmers (see our comment on Assumption 5 above). The estimated quantities of untreated wastewater are based on the amount of sewage collected from the system (due to population growth as well as expanded sewer networks) in excess of the treatment plant capacity. Second, it is uncertain whether citrus areas would be preserved in the way that was anticipated in the ERR

calculation (a separate benefit stream), given that highly saline treated wastewater may be unsuitable for citrus production. Third, it is assumed that the marginal value of water in irrigation is quite high ($0.33/m^3), and the value added per hectare in high cultivation areas is similarly high (3199 JD/hA). These are among the highest values for agriculture globally.

Implications for the IE. The second and third assumptions can be assessed through careful agricultural surveys which trace water source to production types and profits. The first assumption will be more difficult to assess as discussed in section C.

D.3. Other potential impacts omitted from the ERR

As shown in the IE logic diagram presented in Figure D.1, there are a number of potential benefits not readily incorporated into the calculation of the ERR, probably due mostly to data limitations. Specifically, these include:

1. Time savings and/or other productivity gains (household and enterprise-level) due to water supply and reliability improvements;
2. Non-health aesthetic gains of increased water supply and reliability, and sanitation;
3. Health benefits from improved water quality (due to sanitation and water project interventions), as opposed to increased quantity of water supplied to consumers (which are included but likely very modest if even detectable);
4. Long-term effects on utility performance that are not related to "efficiency" benefits, including effects on tariff reform (note that these may be negative);
5. Increased property value other than the reclaimed cesspit area (note that these likely encompass improvements 1-4); and
6. Potential for increasing land under irrigation. Though controversial, several stakeholders mentioned the possibility that the total land area under irrigation could be increased due to the additional water in the system.

D.4. Other considerations

During an initial scoping trip, the evaluation team became aware of some changes in the implementation of works relative to those described in the project feasibility studies. In particular, the scale of the water network project has evolved, as demonstrated by the creation of optional low-water use DMAs that will receive improvements only if funds remain sufficient, as well as the shifting status of several other DMAs. Thus, it is likely that the balance of costs and benefits may change relative to those initially anticipated. Similarly, there are specific complementary investments (beyond the installation of the West Zarqa Pumping Station) that are not included in the ERR calculation, most notably conveyance infrastructures for moving water to the Northern Jordan Valley.

D.5. Description of Program/Project Logic

In order to better understand the mechanisms by which the MCC investment program is supposed to foster change and the delivery of positive outcomes to beneficiaries in Zarqa, the Social Impact team spent time mapping the relationships between specific investments and outcomes. In doing so, the SI team was mindful and thankful of the significant attention that MCA-J and MCC staff paid to development of logical frameworks for proposed compact activities.

In order to develop the causal chains that can ultimately be tested in the IE, the team expanded the matrices in the MCC's M&E Plan into the IE Logic diagram, presented in Figure D.1 below, that illustrates all possible causal relationships that could be traced between Compact activities and measurable impacts. In the IE Logic diagram, program interventions are grouped by color, and outputs, outcomes, and goals are presented separately. Though this report does not address evaluation of the WSH activity, the IE logic diagram does include it in order to make that activity's relationship to the overall program clear. It is important to note that the causal relationships that are shown are only hypothesized; it is thus possible that the evaluation will not detect some of these relationships. We have also highlighted with annotations a number of the key assumptions that underlie these relationships. Furthermore, no attempt has been made to express the relative magnitude of each intervention with respect to the desired outcome, with the one exception of highlighting that the direct assistance component of the Water Smart Homes Activity is only expected to benefit 3500 households.

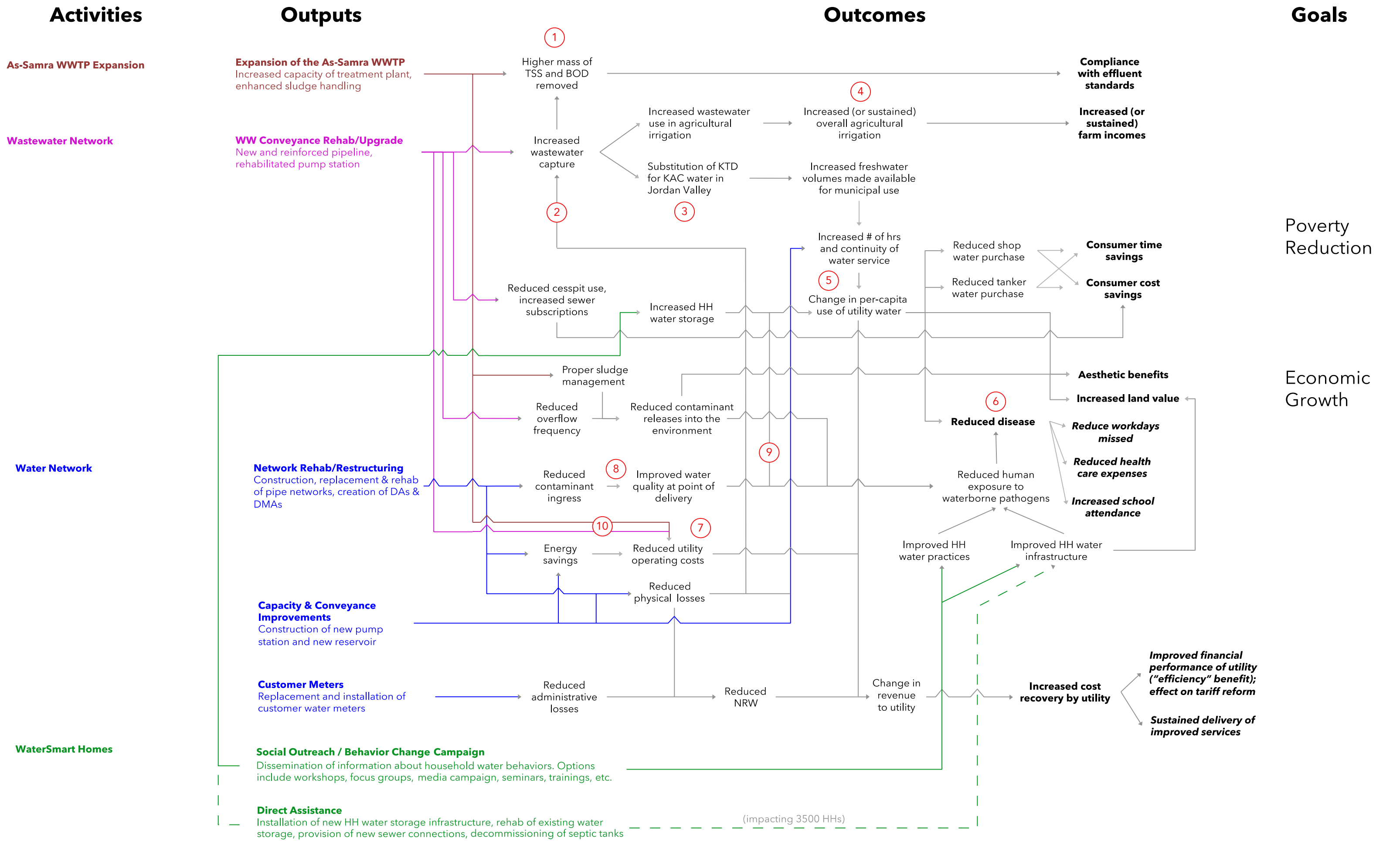
The Compact's Water Network Project is anticipated to result in increased water supply system capacity (through new reservoir construction), reduced energy expenditures per unit volume (through transition to a closed, gravity driven system), reduced administrative losses (through replacement, repair and provision of new customer meters), reduced physical losses, and reduced contaminant infiltration into the network. These effects in turn generate a cascade of projected benefits with respect to utility operations and financial performance as well as household benefits from increased frequency of water service. One noteworthy relationship we identify is the potential of the Water Network Activity to benefit farmers, insofar as reduced physical losses translate into increased wastewater capture. At the same time, we note that increases in system-wide drinking water delivery (as well as wastewater capture, conveyance, and treatment) could increase utility operating costs at rates that may exceed growth in revenues, although the feasibility analysis suggests this will not be the case.

The Wastewater Network Project generated environmental and public health benefits (in the form of reduced sewage overflows), consumer cost savings (in the form of eliminated expenditures from cesspit maintenance), and agricultural benefits (in the form of increased wastewater capture for treatment and conveyance to farmers downstream). The important "primary substitution effect" highlighted in the ERR calculation is situated in this causal chain.

The As-Samra Wastewater Treatment Plant Expansion Project will enable treatment of these additional wastewater flows (as well as those stemming from general population growth), allowing for compliance with effluent water quality standards to be maintained for these larger quantities of wastewater, and perhaps improving the management of sewage sludge. This treatment expansion may have positive effects on irrigation due to improved water quality (specifically, the avoidance of a transition to low-value agriculture in the valley), though we have significant doubts about the strength of this connection. We include this benefit out of recognition of its inclusion in the ERR calculation, but we dodo not expect that there will be observable water quality benefits in Jordan Valley irrigation water resulting from the plant's expansion (see section D.2 on the ERR calculation as well as the Methodology Section C below). We also note that the Wastewater Network and As-Samra Expansion Projects will both affect utility operating costs and thus warrant careful evaluation, since these may increase more than the additional revenues generated from connection fees and wastewater tariffs.

The Water Smart Homes Activity presents household water storage and management behavior changes that in turn result in possible health benefits from reduced pathogen exposure, economic benefits from increased storage and changes in water sourcing and consumption. We note the likely differences in magnitude of the different elements of the program (direct assistance vs. social outreach/BCC activities), and do not consider the WSH specifically in the remainder of this design report. We will present MCC with options for evaluating WSH in a separate document.

Figure D.1. IE Logic Diagram.



Key:
- Colors correspond to Activity
- Normal text corresponds to intermediate outcomes
- **Bold** corresponds to longer term outcomes
- ***Bold italic*** corresponds to secondary or derivative long term outcomes
- Numbers correspond to relationships that include caveats included below

Figure D.1 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.

METHODOLOGY

E.1. Summary of General Approach

In this chapter, we provide a proposal for a broad menu of evaluation activities, grouped as a set of evaluation **options** for measurements within each of three data collection **components**. We strongly urge the MCC to fund all three components, given that they each provide critical and complementary information for understanding the totality of the Compact's success (Table E.1). The different **options** proposed within a component would measure similar impacts (and thus should be considered mutually exclusive), though they have different strengths and weaknesses and entail different data collection intensity and cost. The **elements** (or distinct activities) within a component however are not mutually exclusive; in all cases, particular elements will add new information on impacts, although MCC must determine whether that additional information is worth the cost of data collection for that element. Within each component, we have described our preferred strategy as Option 1, for which we provide the most detailed information. Should MCC favor one of the less preferred options, some additional design work would be necessary.

The IE logic presented and discussed in the previous chapter highlights the complexity of the causal chain that leads from projects to impacts expected from the MCC investment program in Jordan. To properly track and account for the most important changes arising from this complexity, the IE must use a combination of several evaluation components consisting of different types of data collection methods. For the purposes of presentation, we offer three components within which similar and complementary activities are conducted, that would focus specifically on:

1. Water and wastewater network project impacts on households and enterprises in Zarqa / Ruseifa (including water tanker/vendor surveys and household surveys with new populations that settled in Zarqa during Compact implementation);
2. Water network, wastewater network, and As-Samra wastewater treatment plant (WWTP) expansion impacts on the quantity of blended KTR water (and therefore, recycled wastewater) reaching the Jordan Valley farmers, as well as the implications of this substitution that stem from differences in water quality; and
3. Compact impacts on system-wide water utility metrics (e.g., water quantities sold to consumers; or wastewater collected and routed to As-Samra) and financial performance.

We urge readers to consider how these components fit together for the measurement of the specific and general economic impacts of the Compact. Recognizing that the mapping of expected impacts to components of our design proposal is not straightforward, we have developed Table E.1 to facilitate understanding of how they fit together to answer a number of general impact questions. The specific evaluation questions that are related to each component are then summarized in the component-specific sub-sections of this chapter, and followed by a discussion of how the information collected will be used to answer a set of integrating evaluation questions about the value of the primary substitution effect (substitution of blended KTR water for

KAC freshwater in irrigation), the overall economic benefits of the Compact, and its distributional implications across sectors (domestic, commercial/industrial, agricultural, and for the financial situation of the public sector).

In addition, we note that the different IE data collection methods must be conducted at different times and in different locations as deemed appropriate for measuring the expected impacts. For example, the timing and geographic coverage of the wastewater network expansion differs from that for the water network rehabilitation, although there is some spatial overlap of those two activities. In another example: both water and wastewater infrastructure activities are expected to yield increased volumes of wastewater reaching the Jordan Valley to enable the primary substitution effect, but the timing of their respective contributions will vary. Water network activities will lead to lower physical system losses, which would mean an increase in volumes entering into the sewer system. At the same time, wastewater network activities will bring more subscribers onto the sewer network, also increasing sewage volumes. This chapter also includes a discussion of the timeline for these different data collection activities.

Table E.1. Relationship between IE design components and the main expected economic benefits of the Compact

Economic impact question (included in *ex-ante* ERR analysis)	**Data collection components required**
1. What is the economic value of increases in water consumption, reliability, and behavior changes due to the intervention?	Components 1 (household/enterprise, and refugee 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)
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, and refugee 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)
Economic impact question (omitted from *ex-ante* ERR analysis)	
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)

Finally, the complicated web of intervening and confounding variables affecting the water and wastewater management situation in Zarqa means that it will be difficult to establish definitive attribution in some cases, and means that it will also be challenging to separate the impacts of water, wastewater, and treatment plant improvements, given the complementarities existing among these projects. We discuss the major risks and threats to attribution under each of the components as well, and how we have sought to address them.

There is also a set of significant challenges related to definition of the impacts or outcomes of the project components, and to how those should be measured. Indicators I for all of the intermediate outputs and outcomes depicted in the IE Logic diagram cannot all feasibly be measured through the IE, which rather focuses primarily on ultimate welfare outcomes. Many of these intermediate outcomes are instead tracked through the routine monitoring and evaluation (M&E) activities of MCA-J, while for others we are unaware of any plans for tracking (refer to the memorandum transmitted to MCC and MCA-J, dated January 24, 2013, and included as Annex A, for more discussion of the M&E indicator tracking table). Furthermore, many of the potential project outcomes we have identified will occur in sectors that are not explicitly considered by the M&E monitoring, such as the commercial sector (formal and informal), property markets, or among specific subgroups in the population for which only limited M&E data collection activities are planned (e.g., NAF beneficiaries). As described in more detail throughout this chapter, our IE design does seek to fill some of these gaps identified in the M&E plan, and to provide a clear picture of the indicators we will aim to measure (see summary in Table E.2), though we note that such intermediate indicators do not always fit within the scope of standard impact evaluations where the standard for attribution is one of treatment and control.[11]

[11] Not including WSH-related indicators, which are considered in Annex F of the design report.

E.2. IE Component 1: Evaluation of Water and Wastewater Network Project Impacts in Zarqa

E.2.1. Preliminary considerations

The primary evaluation objective for the first IE component, which focuses on measurement of outcomes conducted 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 *had the investments not been made* (Y_i^0).[12] This latter counterfactual obviously cannot be observed, and we require other methods for measuring it with a minimum of bias. The approaches we are considering for evaluation of these changes hinge around exploiting variation in the intensity of "exposure to treatment" with these improvements. In simple terms, we can think of household and enterprise exposure to treatment as corresponding to the classes identified below in Table E.3:

Table E.3. Household and enterprise exposure to water and wastewater network improvements

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

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. Using existing plans and maps, we judge that the physical layout of the network improvements should allow such comparisons (Figure E.1).

[12] Despite the lack of evidence of impacts of piped water and sanitation infrastructure on private enterprises in the literature, we feel the need to include them in our IE design because: a) there has been very little rigorous research on this question; and b) small businesses (both formal and informal) appear to play an important role in the economy of the Zarqa Governorate. Such businesses seem likely to be constrained by the limited reliability of water supply, and the frequency of service interruptions.

Table E.2. Proposed IE measurements of outcomes and impacts from MCC investment program

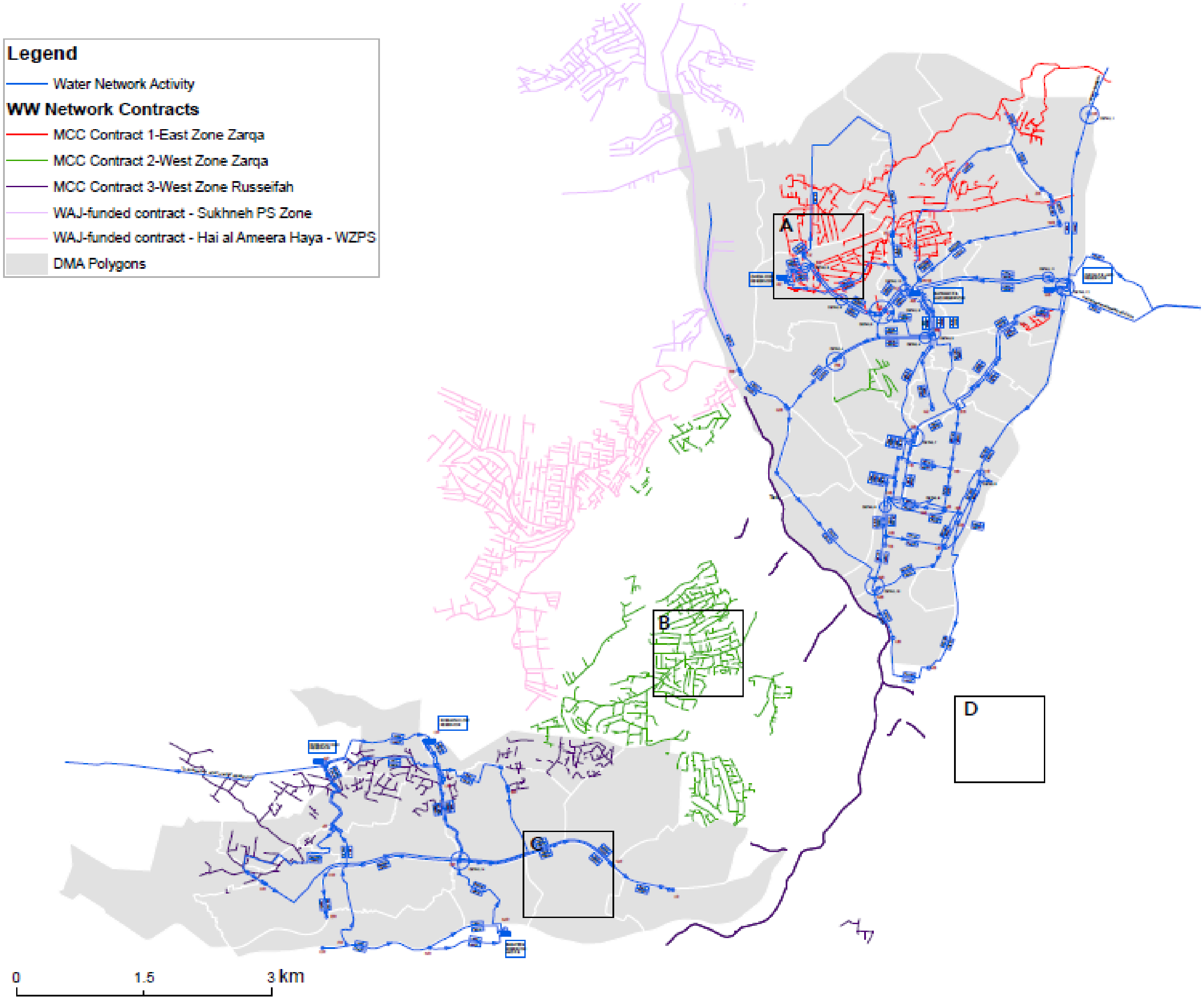
Outcome	Component / Element [a]	Indicators	Level of measurement	Comments
Intermediate outcomes				
Increased water service	-Component 1; Elements A/C -Component 3	-Hours of supply/week -Number, frequency & duration and reason for supply interruptions -Customer complaints about supply reliability	Utility, household[b] & enterprise	These parameters will vary considerably by season, so should be monitored at least on a quarterly if not monthly basis
Improved sewer service	-Component 1; Elements A/C -Component 3	-Number of sewer customers -Volume of wastewater flowing to As-Samra from Zarqa -Customer complaints about sewer failures/breakdowns -Sewer overflows / blockages	Utility, household & enterprise	
Improved water quality	-Component 1; Element A	-Chlorine residual at household -*E. coli or thermo-tolerant coliform* counts at household -Perceptions of network water quality -Customer complaints about water quality	Household & MOH	
Increased water consumption	-Component 1; Elements A/C -Component 3	-Metered consumption (hh and overall) -Quantity of shop/tanker water purchased	Utility, household and firm	
Reduced Non-Revenue Water (NRW)	-Component 3	-Apparent Losses -Real Losses -Total Losses -Pipe breaks / bursts / km of mains	Utility, network	Losses should be reported in several different formulations, as per IWA and PMU practice. Values of these parameters should be determined for different zones, distribution and sample DMAs in the network. Other data such as line length, average pressure and other factors will be needed for full interpretation of these indicators
Water balance (between Jordan Valley and urban areas of Jordan)	-Component 2; Element A	-Sewerage in zones served by the As-Samra plant -Δ in water volume (e.g., from Disi; population growth) flowing into zones served by As-Samra -Δ in water volume (e.g., from WNP and WWNP) flowing into zones served by As-Samra -Δ in withdrawals of treated effluent from the Zarqa River between As-Samra and the Jordan Valley -Local inflows to the KTD from other WWTPs and local runoff -Water releases from the KTD to the Jordan Valley -Water allocations from the KAC to the Jordan Valley	System-level	

Outcome	Component / Element [a]	Indicators	Level of measurement	Comments
Blended KTR water quality	-Component 2; Elements A/B	-TSS, BOD, and bacteria levels in As-Samra effluent -TDS, TSS, bacteria, and chlorophyll-A in KTD water	As-Samra and farm-level	
Farm-level water use and production	-Component 2; Element B	-Balance of blended KTR water and freshwater use in irrigation -Irrigated area -Farm selection of non-water inputs -Farm output (quantities and yields)	Farm-level	
Economic outcomes				
Improved utility cost recovery	-Component 3	-Utility revenue -Utility operating cost -Utility variable operating costs per m^3 supplied -Pump energy consumption kW-hr /$m^3$of water/ m of dynamic pressure -Billing Efficiency -Collection Efficiency -Operating Cost Recovery Ratio (OCCR), %	Utility	In order to fully understand utility cost recovery, data on water imports, exports and own-source production will be required.
Reduced non-network water sales	--Component 1; Element B	-Shop water revenues -Tanker water revenues	Shop / tanker water vendors	
Consumer cost savings	-Component 1; Element A	-Expenditure: Shop, tanker, network, other water -Expenditure for septic pumping -Expenditure for sewage connection -Expenditure for new septic system	Household & enterprise	
Consumer time savings	-Component 1; Element A	-Time spent collecting water -Time spent maintaining sanitation system -Other time expenses related to poor water supply	Household & enterprise	
Increased productivity / capital accumulation	-Component 1; Element A	-Household/enterprise income -Household/enterprise expenditure -Asset ownership -Improved educational status	Household & enterprise	
Aesthetic (quality of life) benefits	-Component 1; Element A	-Satisfaction with water supply -Improved household hygiene behaviors -Non-productive water use	Household	
Reduced economic burden of disease	-Component 1; Element A	-Diarrheal disease: 7-day and 2-wk prevalence; work/school days lost; expenditures on treatment	Household	
Increased land value in Zarqa	-Component 1; Element A	-Self-reported rental/sale value	Household & enterprise;	
Change in net profits from irrigated agriculture	-Component 2; Element B	-Farm input costs -Farm revenues -Farm profits	Farm-level	

Outcome	Component / Element [a]	Indicators	Level of measurement	Comments
Change in farm value	-Component 2; Element B	-Farm land value	Farm-level	

[a] Component 1: Water and wastewater network project impacts on households and enterprises in Zarqa; Component 2: Water network, wastewater network, and As-Samra wastewater treatment plant (WWTP) expansion impacts on wastewater reaching the Jordan Valley farmers; and Component 3: Compact impacts on water utility performance.
[b] Household data collection includes a refugee survey (Element C) to account for changes in population composition and related water consumption over time.

Figure E.1. Map of water and wastewater network improvements (Each type of area identified in Table E.2 is identified; thick blue lines denote water network rehabilitation; thin lines of various colors denote wastewater network expansion).



In practice, this simple picture of exposure to treatment is complicated by a variety of factors, most importantly:

a. The zones in Zarqa that fall into categories A through D are likely to be systematically different and therefore not directly comparable;
b. The timing of exposure of different areas varies over the duration of the investment program, such that some areas are affected before others (e.g., areas move from D to other categories over time);
c. There are likely to be important spillovers from treated areas to untreated areas, given that the network is interconnected, such that the measurable differences between affected

and unaffected areas may be suppressed (for example, water savings in one area may augment water supply throughout the system, even perhaps outside of Zarqa); and

d. The impacts of these investments may only be realized once a threshold of investment is achieved, or following a significant period of time during which other confounding factors will become increasing threats (e.g., as in the case of broader utility performance and reform, population changes from refugee resettlement, or increased system-wide water supply).

E.2.2. Specific evaluation questions associated with Component 1

The major evaluation questions related to determining impacts of the water and wastewater network projects on households and enterprises in Zarqa 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 IE will aim to study the mechanisms by which they are produced, by 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 will 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 (see discussion further below).

We present two quasi-experimental design options for evaluating these questions below. Option 1 is preferred as it combines the strengths of *ex-ante* control for observable baseline differences between treatment and comparison units, while also accounting for time-invariant unobservable differences between these units. Due to the fact that impacts will affect multiple populations and economic activities in Zarqa, we also define three elements within each option, which together would allow consideration of the full set of IE questions presented above. If funds do not allow inclusion of both elements, the IE will not be able to answer question 6.

E.2.3. Evaluation Design Options

Below we describe the IE option that was retained following discussion of two alternatives. We note that Elements B and C of Component 1, and end line surveys for Element A, were pending at the time of this revision to the EDR.

Element A: Analysis of household and enterprise surveys

To measure effects on households and enterprises, the primary approach will be to implement ordinary or generalized propensity score matching (PSM) in combination with difference-in-differences (DiD) and regression analysis. PSM (or GPSM) will be used to predict selection into the various treatment groups A-C (or, if more appropriate, some measure of continuous treatment intensity), using pre-intervention characteristics of those areas. We will then match areas that have similar propensity scores (i.e., that appear equally likely to have received specific exposures to the intervention, based on observable characteristics) to ensure comparability across controls and differentially treated areas, and will conduct balance tests. The DiD design will, in turn, allow us to reduce the threats posed by unobservable differences between affected units that do not vary over time. Ideally, households will be surveyed at baseline, during different seasons (summer and winter) during implementation, and finally during summer and winter after the Compact closes. Next, regression analysis will further allow us to control for factors other than treatment status that may be related to outcomes, thereby increasing 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). Finally, in an effort to address the issue of spillovers, , and in order to test the extent to which households in Zarqa are affected by the awareness-raising activities of the WSH campaign, we will aim to include control areas outside of the Zarqa water and wastewater network, for example areas in Amman that are nearest to Zarqa,

Propensity Score Matching (PSM): The PSM approach assumes that expected outcomes of treatment and comparison groups are independent of the treatment assignment, conditional on a vector of observed baseline characteristics *X* (Rosenbaum and Rubin 1983). The first step in PSM is thus to estimate a logistic regression that explains assignment of unit *i* into treatment controlling for *X*:

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

where $T_i = 1$ if unit *i* has been assigned to treatment and 0 otherwise, and ε_i is an error term. We will conduct PSM at the block level (ideally, at the block level used by the Department of Statistics for Census sampling), and will include as many relevant predictors of participation as possible in the equation 1. This exercise will be facilitated by the fact that very specific criteria were used by the engineering firms that conducted the *ex ante* prioritization of target District Metering Areas (DMA). We will augment these variables with a range of other socio-economic characteristics that may be related to the selection of project zones.

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

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

We will use nearest neighbor (1-1 matching with replacement) and radius-based matching methods to obtain a sample of matched treatment-control pairs in which to conduct our surveys. The results from various model specifications and matching strategies will be compared in order to best obtain balance on the key characteristics that determine treatment assignment at baseline (Lee 2006). Included in these comparisons will be assessment of the overlap between treatment and control units, as well as the quality of the matches, using standard approaches available in STATA (e.g., considering covariate balance, the size and nature of the common support region). These processes will ensure that treatment and control units are as comparable as possible at baseline, and will indicate the extent to which the comparisons are likely to be relevant to the whole of the areas in Zarqa affected by the intervention.

There are three principle threats to the validity of estimates obtained using PSM in this way. 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. We hope to minimize this threat using our understanding of the parameters in the prioritization algorithm that was used to select treatment areas for the water network improvements, and the criteria for selection of expansion of areas treated by the wastewater network expansion, respectively. We will also complement these known parameters with additional data on socio-economic characteristics of the affected zones, since modeling the propensity score using a richer set of regressors tends to improve the performance of the estimator (Heckman, Ichimura et al. 1997; Heckman, Ichimura et al. 1998).

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). At this time, we cannot fully assess the extent of this threat, given the fact that we do not yet have Census or more recent socio-economic data at the block level. 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). We discuss this issue in further details below in section B.6, under risks and mitigation strategy.

Difference-in-Differences (DiD): Once we have generated a sample of treated and control areas that are better balanced on observable characteristics at baseline, we will model survey outcomes in a fixed-effects panel regression framework where:

$$Y_{ijt} = \alpha + \gamma T_{jt} + \delta d_{jt} + \kappa T_{jt} \cdot d_{jt} + \beta X_{ijt} + \delta_{ijt}, \quad (3)$$

where Y_{ijt} is the outcome of interest for household/enterprise or other unit *i* in zone *j* at time *t*, *d* is equal to 1 if household *i* is in a treatment area *j*, and 0 otherwise, T_{jt} is a dummy variable that is equal to 1 once the intervention has occurred in community *j*, X_{ijt} is a vector of time-varying variables that affect the outcome for unit *i* in zone *j* at time *t* (including month of year dummy variables), and δ_{ijt} is a time-varying error term. To be clear, the model we will develop is a multi-level model: we will measure household-and enterprise-specific outcomes (e.g., diarrheal disease prevalence among children under the age of 5) even as treatment is assigned at the community rather than household level. The coefficient κ will measure the "treatment effect," or the change in outcome *Y* for treatment households or enterprises relative to that for controls. This estimate is unbiased so long as the error term δ_{ijt} is not correlated with treatment. Equation 3 readily accommodates data from the multiple waves of data collection that are planned (baseline, during and post-intervention) and that differ in the precise timing of treatment.

Particularly for the wastewater network expansion, κ represents an intention-to-treat (ITT) estimate. This is because κ indicates the average effect of the community-level intervention across all households / enterprises in that community, whether or not they choose to adopt the improvements that come with the intervention; in other words whether or not they actually connect to the expanded sewer network (Galasso, Ravallion et al. 2004). Based on conversations with MCC and MCA-J, we expect that the ITT estimate will be very close to the actual treatment effect on the treated, given that a) the vast majority of households / enterprises are currently connected to the water network and so will automatically receive the rehabilitation treatment; and b) connection to the sewer network is supposedly compulsory in zones that have access to it. However, these may be different if households / enterprises do not adjust their water consumption due to continuing concerns over water quality and reliability, if households / enterprises do not connect to the sewer lines, or if the timing of treatment does not perfectly follow the timing of the infrastructure network, perhaps due to lags in the re-optimization of network operations by the utility. We will assess the degree to which this is true using the data collected in the household survey.

Data: To implement the PSM methodology using pre-intervention characteristics in the different areas, we will rely on several data sources. Some of these data are already available at the DMA and other levels and have been obtained from the PMC. We have also worked with DoS to obtain other critical data (Census and detailed block-level socio-economic data from the income and expenditure surveys conducted by the Department of Statistics (DoS)) needed for the matching procedure.

Once the sampling frame for the IE is specified and pre- and post-intervention data are collected, the major part of this analysis will be conducted using data on intermediate outcomes and social welfare measures obtained from panel surveys – required to carry out the DiD analysis – of households and enterprises within the sample zones identified by the PSM procedure (see prior Table E.2 for major groups of such outcome variables). 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 presented previously. Collecting and analyzing these data will produce a more fine-grained understanding of the mechanisms that have or have not led to real changes in well-being.

Household-level surveys will 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. Baseline and end line surveys will provide pre- and post-completion data collected in the same season, respectively, and the data from these will be augmented with seasonal winter and summer surveys conducted regularly over the course of the evaluation period. When integrating the seasonal survey data into the panel, we will code the variable denoting that treatment has occurred (T_{jt}) based on the status of the block at the time of the interview, exploiting any differences in the *timing* of implementation of the infrastructure improvements.

Enterprise surveys will focus 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 (for which no sampling frame currently exists), we will include informal production activities carried out by households in our sample in this assessment, and supplement this with interviews with informal businesses (using DoS records of informal enterprises. The enterprise surveys will be conducted at baseline and end line.

The household and enterprise survey instruments will be developed based on well-tested existing instruments previously applied by members of the SI team in studies in other countries. These instruments will undergo forward and backward translation to ensure the accuracy and precision of survey language. Challenging and additional questions will also be thoroughly piloted in focus groups with men and women, and through training activities with enumerators. Finally, an extensive pre-test will be conducted in non-sample areas of Zarqa prior to launch of the survey. The household survey instrument is expected to cover up to 10 modules and have a length of approximately 30 pages, and will take 45 minutes to 1 hour to complete. The instrument for the enterprise survey will likely be shorter, between 15 – 20 pages (and take roughly 30-40 minutes to complete).

The power calculations presented in Annex C show that a sample size of 3,440 randomly-selected households would provide sufficient power to detect statistically significant changes of 10% magnitude on six important outcome variables, including water supply, water consumption, water bills, spending on treatment shop water, quantity of water purchased in treatment shops, and monthly expenditures on water. To be clear, the calculations assume that:

- A sample size of 2,500 would be sufficient for detecting 10% differences across treatment and control groups if the sample is comprised of 4 groups (roughly 625 households per group)

- We add 1 additional group from peri-urban zones in Amman to test for Zarqa-wide spillovers (625 households, 3125 total)
- There will be 10% attrition, such that 313 additional households are required (total of about 3,440).

Differences in treating water from the public network as well as self-reported health expenditures because of water consumption would be harder to detect with this sample size, especially for the latter. Reasonable power to detect such differences would require sample sizes beyond the capability of this evaluation. For the enterprise survey, we will target a sample of 345 enterprises, based on the sample-size considerations presented in Annex C. However, we note that the sample size for the enterprise survey will be reassessed following baseline which will allow more detailed power calculations, and consideration of whether more careful stratification by formal/informal status would be warranted.

Both the household and the enterprise survey use panel survey designs to measure outcomes before and after the MCC interventions. Over the course of the evaluation, we also plan to periodically engage with focus groups in affected and unaffected areas to collect qualitative and semi-quantitative information on the water situation in different parts of Zarqa. Such focus groups, though perhaps not strictly necessary for answering the IE questions, will allow for a more nuanced perspective on the way different groups of households and individuals perceive and are affected by the investment program, and are considered to be best practice in IE work since they enable the evaluation team to make adjustments to survey instruments in response to changing conditions. In addition, we propose to collect a variety of utility-level indicators (from WAJ-Zarqa), to implement a small number of additional data collection activities, namely meter tests, and to carry out some independent water quality testing as detailed further below in component 3. Finally, in order to answer the final two specific evaluation questions for this component that are presented above, we would like to conduct systematic, small-scale data collection with shop water vendors and tankers, as described in element B below.

<u>Element B: Analysis of impacts on the non-network water vending industry in Zarqa.</u>

The activities included in this element of component 1 are required for obtaining answers to evaluation question 6, since it represents a potentially important impact on economic welfare in Zarqa that is unlikely to be captured accurately in the household survey. To fully address this question, the costs of the investment on alternative water suppliers (whose profits may or may not be displaced) need to be assessed, to determine whether the cost savings to households simply represent a transfer of income from vendors to households, or whether they result in net welfare changes due to greater consumption and an ability of these industries to adapt to changing demand. Since relatively few households and businesses sell water themselves, the IE would be unlikely to detect such costs through the household and enterprise surveys.

Element B will also enhance learning that is obtained from the evaluation. In many countries, urban utilities fail to provide consistently reliable and high quality piped water supply to households, due to lack of sufficient supply, the threat of infiltration with contaminated water from regular depressurization of pipes, and/or lack of universal connections to the piped network. Although most households are connected to the water network in Zarqa, households have been found to rely on non-network alternatives, perhaps due to concerns over intermittent supply. Many of these households rely on purchased water (from shops, tankers, or other vendors) to supplement unreliable utility supply. This situation is not unusual, global evidence suggests that the reliance on private sector vendors has increased by large amounts in many cities in recent years in different regions (Thompson et al. 2000; Stoler et al. 2012), despite its higher cost (World Bank 2004; Nauges & Strand 2007; Whittington et al. 2008). An important question is whether this reliance is due to unreliable network water delivery, as is assumed in Jordan, or due to perceptions that non-network sources are higher in quality, which has also been observed in numerous settings (Whittington et al. 1991; Davis et al. 2001; Korfali & Jurdi 2009; Orgill et al. 2016). Finally, little systematic research exists on the scale and magnitude of water vending activities in metropolitan areas of developing countries. Nearly all existing studies rely on household survey data, and thus present only a partial picture of water vending, because this supplemental source may be underreported by households. There is anecdotal evidence from several cities (e.g., Manila, Karachi) that vendors often simply repackage network water that may not be obtained legally (David & Inocencio 1998, Chaudhury 2013), which implies large inefficiencies. To be sure, documenting such practices is complicated, further increasing the knowledge gap on urban water vending practices.

We hypothesize that some fraction of the Zarqa population relies on bottled “container” or tanker water for drinking and other purposes, based on data in the DoS Water Survey from 2009. Yet it is not fully clear whether households will switch away from such sources, since they may depend on them in order to cope with unreliable water supply, or alternatively because of a lack of trust in the quality of water delivered by the utility. Shop water containers are 19 or 20 L clear plastic jugs, and are generally sold at retail outlets for basic household goods. We will complete a listing and cross-sectional survey of such shop and tanker water vendors at the conclusion of the Compact implementation, and use it to perform an industry analysis that documents if and how the intervention has affected this business (including the amount of competition in the industry). We anticipate that the level of cooperation among small-scale shops will be good but that cooperation from tanker operators may be more difficult to obtain. Given the potential sensitivities, we plan to enlist the help of a private data collection firm to manage this element. With the data from these surveys, we will be able to develop secondary estimates of changes in household purchases of container water from the supplier side.

We propose a two phase approach to systematic data collection with this population. Phase 1 will be a largely qualitative phase to understand how best to engage and ask sensitive questions with these operators. SI had originally proposed engaging a private firm to conduct this phase, but the budget approved for the vendor survey during the 2017 modification process was smaller than we had proposed for this activity. To minimize costs, the IE team’s local field coordinator will pilot the vendor survey with at least 5 water shop and 5 tanker business owners operating in Amman and Zarqa, and a private data collection firm will be enlisted for Phase 2 only. The findings of these interviews will help the team to adapt a draft survey instrument developed for the quantitative survey based on prior work with water vendors in other settings (e.g., Whittington et al.'s study in Onitsha, Nigeria offers a good and relevant urban example[13]).

Phase 2 will then build on this preparatory work to implement a quantitative survey. Detailed information required for power calculations of this data collection activity is not available. However, we anticipate, based on information obtained from the Ministry of Health, that the number of shop operators in Zarqa does not exceed 200, which should allow a survey of the complete set, even if we assume that there is an equal number of tankers (and a comparison group of shops and tankers from control areas in Amman). We will first confirm the accuracy of the listing of water shops, and will work in parallel with the Division of Motor Vehicles to conduct a similar exercise on tankers, or use other acceptable methods proposed by the firm selected for the work. If the final compiled listing reveals that there is a total of more than 500 water shops and tankers operating in our survey zones, we will draw a sample of 500, spread evenly across our survey arms (100 per arm, so 400 total in Zarqa, and 100 in Amman), and proportional to the relative number of tankers and shops.

The data analysis will then focus on changes in this industry during the Compact period, using data from firm financial records and self-reports. The primary comparisons will look at simple aggregate firm-reported net revenues before and after the Compact (using firms' financial records), which is necessary to obtain an industry-wide perspective since households may purchase from shops or tankers located outside of their immediate neighborhoods. With shops, for whom most purchases are likely to be local, we will supplement this industry-wide analysis with multivariate regression analysis that exploits proximity to infrastructure improvements and controls for observable differences between shops in different groups:

$$Y_i = \alpha + \gamma D_i + \beta X_i + \delta_i, \qquad (4)$$

where Y_i is the outcome of interest for vendor *i* (e.g., change in net revenues over the same time period before and after the Compact), D_i is a measure of distance (or alternatively, a vector of dummy variables that pertain to the different control and treatment arms) from the closest water network project infrastructure, X_i is a vector of shop-specific control variables, and δ_i is an error term. The coefficient γ will measure the "treatment effect," or a measure of the effect of distance from the WNP improvements, on vendors' net revenues

The survey instrument for this activity will contain questions related to basic firm characteristics; water sources and use of inputs; business constraints and other factors influencing profitability (including competition in the sector); volume of water sales; cost and revenue structures, and investment; and adjustment in geographic coverage and/or estimates of changes in the spatial distribution of demand and number of customers. Such data collection instruments have previously been developed and utilized by one of the members of the evaluation team in other urban settings.

[13] Whittington, D.; D. Lauria; X. Mu (1991). "A Study of Water Vending and Willingness to Pay for Water in Onitsha, Nigeria." World Development 19(2/3): 179-198.

Element C: Analysis of refugee survey data.

In Fall 2017, shortly after the end of the Compact, we propose to carry out mixed methods surveys with new populations that settled in Zarqa during Compact implementation. The motivation for this element is that it will allow the evaluation to more fully understand and adjust impact estimates that may be affected by the influx of people into Jordan during this period, in particular as they relate to water demand and consumption. Estimates of the population increase in Zarqa Governorate vary across sources, but conservative sources put the number at 94,000, of whom about 48,000 live in urban or peri-urban areas (UNHCR 2016). These population increases have led to reported reductions in water provision, and the government has issued a National Resilience Plan in response (GoJ 2004).

Ordinarily, one would expect that the treatment-control design of the evaluation would not be vulnerable to confounding effects from population growth (since these increases would occur throughout the sample. Yet differential settling of these new populations across sample areas could create bias and threaten the IEs internal validity. For example, one might expect that refugees would be more likely to settle in lower density areas, where rents are cheaper; these areas are more likely to be in the zones targeted by the WWNP. Alternatively, these populations may prefer to locate in zones that are closer to other services, which may be concentrated in older parts of Zarqa that are more likely targets for the WNP rehabilitation. In addition, simply making adjustments on the basis of consumption levels measured in the general population would not be appropriate, given the particularities of this population. There is ample evidence that refugees who have recently settled in Jordan have a different socio-economic profile than the general population, are more likely to be renters (CARE 2013), may thus be more likely to settle water bills in non-conventional ways (through landlords or other meter sharing arrangements) and less likely to invest in water conservation, and may therefore also exhibit very different water consumption behaviors (UNHCR 2014).

Methodology. The evaluation team will conduct the work in this element in two phases. In phase 1, semi-structured interviews and focus groups will be used to understand the perceptions of this population vis-à-vis surveyors asking water-related questions, and to learn how to ask critical questions about water expenses and behaviors. The results of this exercise will be used to inform revisions to the household questionnaire used in Element A. Phase 2 will then consist of a representative survey of refugee households in Zarqa and in survey areas in Amman in order to document differences in population and consumption behaviors. Specifically, we will work with a data collection firm to create a listing of households who recently settled in our sample zones, aided by UNHCR lists (if these can be obtained). We will then sample 1,500 households randomly from these lists in order to maintain at least 250 households per sample arm (detailed power calculations are again not possible due to lack of data).

The refugee surveys will collect similar data as (and be adapted from) the household surveys utilized in Element A, that is, information on: household demographics; water sourcing, pumping, storage, and use behaviors; preferences and satisfaction with water supply and sewer service; coping and health costs related to intermittent water supply and poor water quality; and expenditures, income, and other socio-economic characteristics. Challenging and additional questions will be thoroughly piloted through training activities with enumerators, and will be pre-tested prior to launch of the survey.

Analytically, the refugee survey will allow us to approximate the *additional* water consumption and wastewater production (and other effects) that arise from this population inflow, and to determine how it adds to water demand in our sample areas and in Zarqa as a whole. This additional amount can then be contextualized against the increases (or lack thereof) in water consumption among different sample sub-groups, so as to obtain a better representation of the effects of the Compact. In short, we will determine refugees' consumption levels (and other outcomes) across treatment arms, in determining how water supply has increased in each zone. We will estimate the following regression:

$$Y_{ij} = \alpha + \delta T_j^1 + \kappa T_j^2 + \beta T_j^3 + \gamma X_{ij} + \delta_{ij}, \qquad (3)$$

where Y_{ij} is the outcome of interest for household *i* in zone *j*, T_j^k is a dummy variable that is equal to 1 if community *j* is in treatment arm *k* (where those arms pertain to WNP, WWNP, and both interventions), and δ_{ij} is a time-varying error term. The coefficients δ*,* κ*,* and β will measure the differences in outcome *Y* for refugees in treatment areas relative to that for refugees living in control zones in Zarqa. We will also control for a vector of household characteristics X_{ij}. This estimate is unbiased so long as the error term δ_{ijt} is not correlated with treatment.

On the basis of these estimates and on the differences in the numbers of new households settling in our sample areas (obtained through the listing exercise), we will be able to estimate how overall water consumption has changed across zones, to complement the information on per-household water consumption collected through Component 1. This will help us determine whether the impact estimates are affected by differential changes in population or household composition / consumption profiles.

E.2.4. Risks and mitigation strategy

This section describes a variety of risks associated with the evaluation activities include in Component 1. Specifically, we consider the most important risks as summarized in Table E.4, and discuss them in greater detail below.

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, it may be difficult to obtain a precise function that maps characteristics observable by our team to selection into treatment. We will however, include as many factors as possible in the estimation of propensity scores, to limit the nature of such threats. We will also try to time our surveys to take advantage of differences in the *timing* of exposure to treatment, perhaps using targeted areas as controls for areas treated early in the investment program.

The second concern then relates to the overlap between treated and potential control units, which may be very limited given the priority of investments in unserved (in the case of sewerage) or problem (for water network rehabilitation) areas. Related to the latter is the fact that our matching algorithm will likely rely on some data sources that are somewhat dated, which will at best contribute random noise to the comparisons (if differences between treated and matched control units are uncorrelated with other factors that help determine outcomes), or at worst contribute bias (if they are). To best preserve flexibility to manage such threats, our power calculations are deliberately conservative, which will 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.

Lack of statistical power. Another distinct concern has to do with the potential lack of statistical power to reliably measure impacts of the MCC investments. 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. For example, diarrheal disease prevalence, even in children under the age of five, is not likely to exceed 5%, such that very large samples would be required to detect modest changes in prevalence, which the literature reviewed previously suggests we should expect. Developing a power analysis for an IE with such a wide range of impacts as this one is a significant challenge, and we have limited ourselves to conservative assumptions related to the key outcomes for which we have data (i.e. household-level outcomes such as expenditures on non-network water, and 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. We also note that oversampling among (or selecting for) households with children would not be a viable strategy for addressing some of these power limitations because it would place undue focus on particular impacts (e.g. diarrhea incidence), to the detriment of measuring general Zarqa-wide impacts.

Outcomes from Table E.2 for which power calculations were not possible given the non-existence of data are: 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, some of the outcomes for which much greater sample sizes would probably be needed are: Customer complaints or sewer overflow hazards; diarrheal disease prevalence; and water quality violations at the tap (bacterial or other).

Table E.4. 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 -Test for systematic differences following baseline -Oversample at baseline; make *ex post* adjustments -Obtain better controls by leveraging differences in timing of exposure to treatment
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 outcomes that are unlikely to be detectable (e.g., 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 -Include property valuation activity -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, differential population changes)	-Statistical control for confounders using DiD -Integration with other components (refer to discussion following Component 3) -Measure of intermediate outcomes to obtain understanding of change mechanisms -Inclusion of a refugee survey element
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
Other important considerations (not discussed in detail in following text)	1) Attrition in sample 2) Limited sample size restricts ability to detect treatment effect heterogeneity 3) Confounding of effect of infrastructure improvement by O&M behavior of WAJ-Zarqa 4) Non-cooperation of vendors/tankers/enterprises	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 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 4) Utilize previously-developed data collection protocols that ensure confidentiality and increase respondents' trust

Short (planned) time horizon of IE. As originally 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, Hanemann 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). The impact of the latter on property values argues for measuring those changes, even though that may require additional data collection (Element B of this evaluation component). 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)). As of September 2017, MCC approved funding to continue supporting monitoring and evaluation activities through September 2019, expanding the timeframe to measure impacts by 2 years.

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 could be Disi water, to the extent that such additional water reaches treated and untreated areas in different amounts. The details of Disi supply to Zarqa are murky at this time; however, it became clear during SI's scoping trip that there will eventually be direct connections between this new supply and the Zarqa system. It is also true that the supply of Disi water to Amman will indirectly affect Zarqa via its indirect effect of reducing Amman's demand for other sources that currently serve both Zarqa and Amman. Whether these changes will increase 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. In addition, differential changes in population in treatment and control areas, and generally increased demand in Zarqa, could mask the gains from the intervention, if they exceed expected amounts (for example due to large population inflows into Zarqa from Syria). These collective changes will provide a large and new influx of water to the Zarqa governorate, on the one hand, and will increase demand, on the other; both types of changes co-occur with implementation of the Compact. Finally, institutional changes to water utilities and water sector governance, as promoted through USAID's Institutional Support and Strengthening Program (ISSP), may have similar differential effects that challenge the ability of the IE to isolate the effects of the Compact investments.

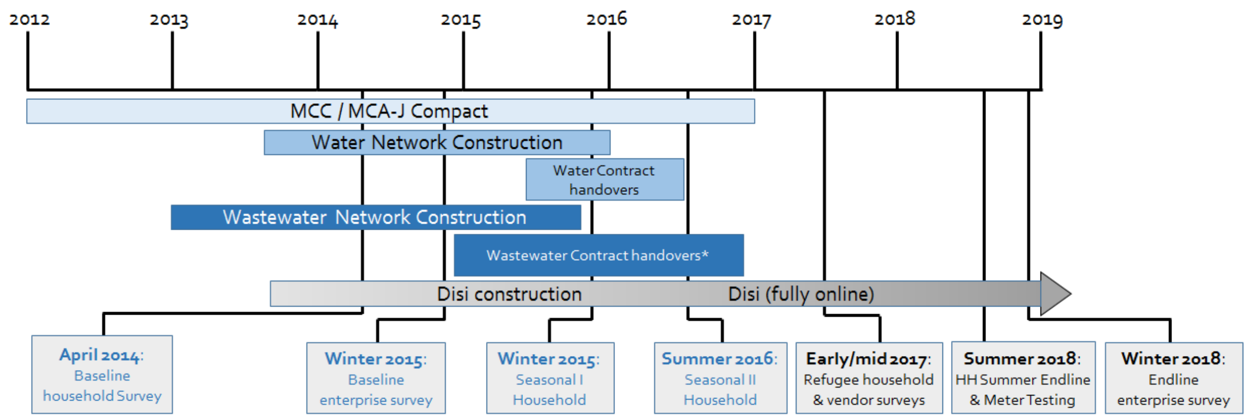
Our approach for dealing with these kinds of confounders will utilize a number of strategies: 1) statistical control (including such time-varying factors explicitly in the DiD estimation through the term X_{ijt} in equation 3 above); 2) integration of the results obtained through the Zarqa-based surveys 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; 3) careful measurement and attribution of changes in intermediary variables hypothesized to lead to the outcomes of compact investment; and 4) measurement of the implications of population growth from refugee resettlement patterns in particular. 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 will hinge 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* GPSM to consider intensity of the water network treatment based on the additional hours of water supply to various DMAs. 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 more convincingly to benefits such that a more complete picture of impacts can be obtained.

E.2.5. Timeframe and implementation (and prioritization) of Component 1

At the time of writing this design report, in early 2013, the wastewater network project construction had already started (although no new households have yet been connected) (Figure E.2). The water network rehabilitation contracts had been issued and construction is scheduled to begin during the fall of 2013. And though new connections and the physical effects will not be felt for some time still, it is critical that the baseline surveys be conducted very soon. This is because behavioral responses (in particular those related to asset accumulation and investment decisions of firms, households, and the water vending industry) change in anticipation of new infrastructure investments. In fact, given that awareness-raising activities are ongoing and actual works projects are underway in some areas, it is likely that expectations have already begun to affect behavior. We think there is a non-negligible risk that this will lead to underestimation of the economic impacts of the Compact, and would therefore push for baseline surveys to take place as soon as possible.

Figure E.2. Water / Wastewater Network Projects and Evaluation Component 1 Timeline (Surveys shown in blue font are complete; black font are planned)



Note: A small portion of the wastewater network project (10 km) was handed over in March 2014.

However, it is also important to acknowledge that the timing of the arrival of Disi water in Zarqa (even if indirect, via effects on water allocation across urban areas in central Jordan) complicates this picture somewhat. To the extent that the baseline occurs prior to the arrival of any or most of the water volumes added by Disi, a naïve evaluation strategy that failed to account for Disi would misattribute this additional water (which may swamp the savings obtained from the reduction of physical losses and the primary substitution effect) to the Compact. Still, we plan to explicitly account for these added volumes by requesting data on water flows from Disi to Zarqa from the WAJ, and integrating this information into our overall integrating water balance analysis. Therefore, it is critical to determine in the near future whether the WAJ can actually provide information on water inputs from Disi and other supply augmentation to Zarqa.

We further suggest inclusion of seasonal surveys in this component (rather than simple baseline and endline) for the following reasons: a) they will allow us to exploit variation induced by the differential timing of treatment across areas affected by the water and wastewater network investments; b) the inclusion of additional surveys will allow us to better understand and track relative changes in control vs. treatment areas that will provide clues regarding the extent of spillovers (which should be much reduced at midline relative to endline); and c) inclusion of additional surveys will also allow for better protection against time-dependent confounders such as Disi or seasonal effects, since these are unlikely to occur all at once.

Finally, the Compact ends in December 2016. Following the close of the Compact, we suggest the following timing for the remaining data collection activities, also as indicated above in Figure E.2:

- A single end line household survey in summer 2018.
- A single end line for the enterprise survey in late fall 2018.
- Water vending and refugee surveys in late 2017 and early 2018

We prioritize these data collection activities as follows: 1) Summer end line for household survey; 2) Refugee survey; 3) Enterprise end line survey; and 4) Water vending survey. The end line surveys for households and enterprises are critical to the longitudinal design of the IE. Meanwhile, the refugee survey is key to assessing and, if warranted, correcting for internal validity threats in the household survey, which is the most significant survey for measuring impacts in Zarqa. The water vendor survey is least critical, but would provide important information concerning water substitution by urban consumers in Zarqa.

E.3. IE Component 2: MCCMCC Compact impacts on agriculture downstream of As-Samra and in the Jordan Valley

E.3.1. Preliminary considerations

As human populations push against the constraints posed by limited conventional freshwater resources, there is hope that wastewater reclamation will become an increasingly valuable means of maintaining human welfare and enabling future growth. In few places is the necessity for viewing wastewater as a resource rather than a nuisance more pronounced than in the water-scarce countries of the Middle East, of which Jordan is a prime example. Indeed, much of the economic rationale for the MCC investment program in Zarqa does not rest in the benefits of these activities to households residing specifically within the zones of Zarqa receiving infrastructure improvements, but rather in its indirect effects on increased water availability in Jordan. These effects would occur via the substitution of reclaimed wastewater (i.e., the product of wastewater treatment that meets water quality requirements for a specific end use) for more expensive water sources (specifically, conventionally sourced freshwater from the Jordan-Yarmuk surface water system), as well as the preservation of high value agricultural activities in the Jordan Valley.

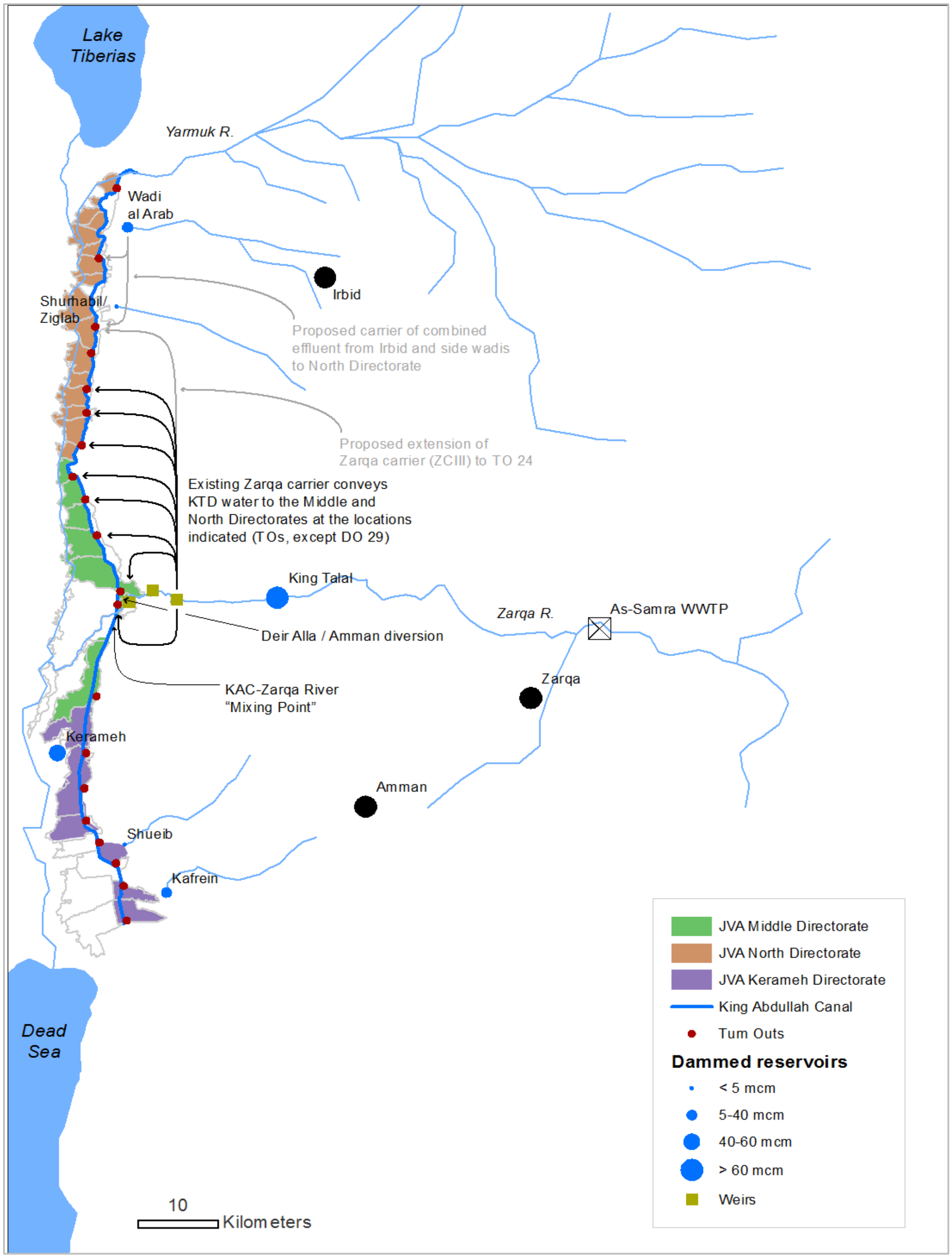
To put into context the logic of this substitution effect, a geographic view of the irrigated water supply system is helpful. Figure E.3 below provides an overview of the locations of major reservoirs, including the King Talal Dam (KTD), as well as water conveyance infrastructure to and within the Jordan Valley. JVA-delineated agricultural "development areas" (DAs) are also shown, as are the route of the King Abdullah Canal (KAC) and the locations of "turnouts" – points at which local networks branch off of the KAC. The existing and planned wastewater delivery systems from reservoirs into the valley distribution system are illustrative diagrams, but the other elements of the map are nominally accurate in geographic terms.

The logic behind the benefits from water substitution stems from the idea that increased wastewater flow to As-Samra, enabled by reduced losses in the water network and increased wastewater collection and conveyance in Zarqa, followed by treatment, will increase the flow of reclaimed water to the KTD, which serves as a crucial source of irrigation water for the Jordan Valley. Yet analyzing the impacts of the water, wastewater, and As-Samra WWTP expansion activities on the agricultural sector is difficult, due to four principal challenges:

1) Overlapping causal relationships;
2) Major non-Compact confounders that limit the potential for attribution;
3) Low reliability of existing hydrological and agricultural water use data; and
4) An inability to observe the counterfactual of no Compact investments (i.e. the lack of a control group that allows definitive attribution to the MCC investment program).

With respect to the first of these challenges, the causal relationships between the water and wastewater activities, respectively, and water made available for agriculture, clearly overlap (see Figure D.1). For example, as noted in Section A of this document, there is a possibility that both water and wastewater network improvements will result in increased effluent production at As-Samra and inflows into the King Talal Reservoir (KTR), the water body that forms behind the King Talal Dam (KTD). The wastewater infrastructure activity will result in more wastewater capture (in exchange for less wastewater being disposed in septic tanks and cesspits). Meanwhile, the water infrastructure activity has a principal objective of reducing NRW, an important element of which is the reduction of physical losses.

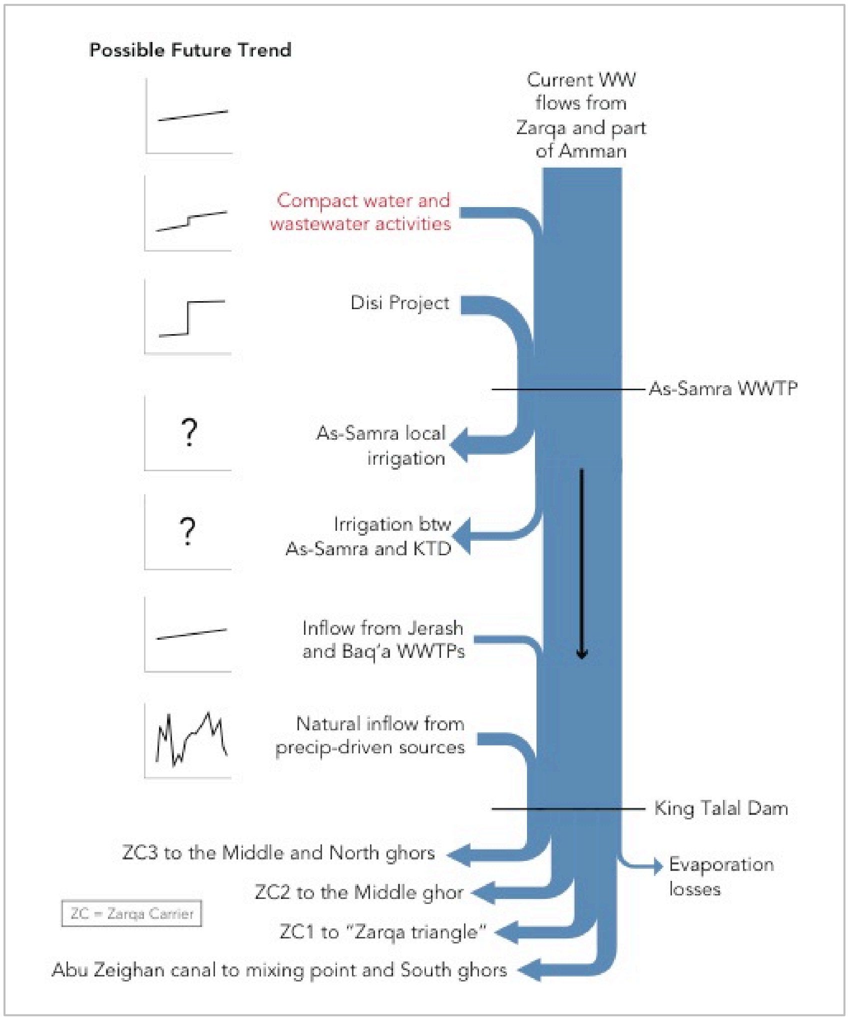
Figure E.3. General schematic of wastewater irrigation conveyance to the Jordan Valley from Zarqa/Amman. Map produced by the SI Team using publicly available sources as well as schematics developed by RTI/ISSP in collaboration with the Jordan Valley Authority.



Physical losses from the water supply system currently infiltrate into groundwater, flow into leaky sewer lines, and/or runoff into wadis. Reduction and control of these physical water losses as well as the rehabilitation/upgrade of sewer lines should lead to some increase in wastewater capture, as long as there is an increase in utility water usage that is larger than the reduction in displaced shop and tanker water. In areas where physical losses are prevented but there is no sewer coverage, additional use of water will instead end up in infiltration and septic systems and not reach As-Samra. We reiterate here that it makes little sense to attribute enhanced water availability for agriculture to the As-Samra treatment plant expansion by itself, since it is the complementary investments in water and wastewater network infrastructure that provide this additional water, *not* the AEP.

The second challenge is more significant. Increases in the volumes that result from the water and wastewater activities must be analyzed in the context of multiple inputs and outputs into a complicated flow regime that connects water users in Zarqa to the production and transport of wastewater from sewer network subscribers (as well as the portions of Amman served by the As-Samra WWTP), and then to agricultural end-users in the Jordan Valley. Figure E.4 below partially illustrates the nature and magnitude of this challenge.

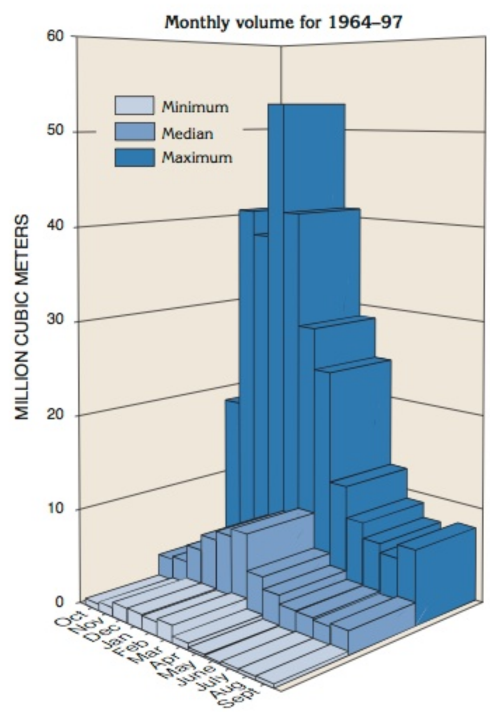
Figure E.4. Inputs and outputs in the flow path between municipal wastewater production in Zarqa and Amman, respectively, and irrigation allocations in the Jordan Valley. The widths of the flow lines are approximate and meant for illustrative purposes only. Trendlines at left correspond to hypotheticals for each inflow/outflow element.



The Compact's water and wastewater activities are expected to increase flow into the As-Samra WWTP. At the same time, natural population growth, imports of new water from the Disi aquifer in southern Jordan will also increase these wastewater inflows into As-Samra, and the timing and magnitude of each of these increases will differ. Other important concurrent changes will affect flows eventually reaching the KTR, including the amount of irrigation taking place in the immediate vicinity of the As-Samra WWTP, other irrigation further downstream (but above the KTD), changes in mixed effluent flows from Jerash and Baq'a, and perhaps most importantly, interannual variations in naturally-occurring streamflow (a function of precipitation). The small trendline plots in Figure E.5 are intended to demonstrate that the nature, timing, and magnitude of these changes differs from element to element. For natural background contributions to the lower Zarqa, the year-to-year differences are likely to be dramatic but will be difficult to establish without extensive precipitation and river gauge measurements as well as modeling (with the unpredictability likely to be compounded by global warming in coming years).

Figure E.5, drawn from the US Geological Survey's EXACT program, indicates the extreme differences in Zarqa River streamflow resulting from interannual precipitation variation. Winter flows can range from the baseflow (consisting nearly entirely of WWTP effluent, some 5 million cubic meters (MCM) per month) to well above 80 MCM/month). In practice, this means that during very wet years, the effective storage capacity of the KTR (75 MCM) is reached or even exceeded, and the Jordan Valley farmers enter the dry season with the KTR at full capacity. By contrast, following dry winters of little to no precipitation, reservoir levels can fall below 30% of capacity (or worse, as in the droughts of the early 1990s). Depending upon winter rains, the KTR can enter the dry irrigation season with volumes varying by more than 40-50 MCM, differences that dwarf the contributions to dry season flows that can be attributed to the Compact's water and wastewater activities. Without precise measurement of these components, there is little hope of accurately quantifying the extent of any primary substitution effect that may stem directly from the Compact investments.

Figure E.5. Minimum, median, and maximum monthly discharge of the Zarqa River measured at the New Jerash stream gage, 1964-1997. Source: EXACT Program of the US Geological Survey.



Third, and building from the comments above, the hydrological data that would aid in modeling the hydrology of the KTD system and the agricultural zones depending on it appear to be quite limited and are of questionable reliability. The widths of inflows and outflows in Figure E.5 were intended to be roughly proportional to their magnitude, but substantial improvements in data availability and quality are required to establish these inflows with any degree of certainty. Published estimates of many variables simply do not exist, and other data that we assume are collected or maintained by JVA, Ministry of Water and Irrigation (MWI), and other agencies do not appear to figure in the M&E plan of the MCA-J. For example, we have not been able thus far to obtain Zarqa River streamflow gage data for the past 15 years from public sources. With respect to data on water allocations among Jordan Valley end users (farmers), it is our understanding that many flow monitoring instruments are non-functional due to elevated contaminants in the distributed water (salinity and/or algae), sabotage, or inadequate maintenance. The data that we have been able to examine up until now have been exceedingly vague and aggregated at levels that would make it impossible to examine fine-resolution effects on Water User Associations, Turnout Zones, or farm DAs, let alone at the level of individual farmers. Thus, a substantial data collection effort is warranted, and proposed, below.

Fourth, and finally, it should be emphasized that the causal linkages shown in the IE logic above (and underlying much of the ERR calculation) assume that the As-Samra expansion (and indeed, the Compact investments as a whole) is an essential input to increased agricultural productivity, in comparison with the counterfactual of no expansion. Just as with household and enterprise impacts, the direct counterfactual in this case cannot be observed. However, in this case, the evaluation problem is even more complicated, because there is clearly no untreated comparison group that can be used to approximate this counterfactual. The best that can be done then is to consider carefully the productivity of agriculture that is naturally exposed to differences in water quality and quantity, in an effort to isolate the effect of these from other factors (e.g. soil fertility, weather, technology, farmer ability) that affect this productivity, and then to combine such micro-level information with a macro- and micro-level picture of the water balance. Achieving a basic understanding of the natural variation in water quality and quantity over space and time provides the only viable option for evaluating whether irrigation water augmentation with blended KTR water (and therefore, recycled wastewater) is in fact occurring in the Valley, and whether this additional water is of sufficient quality to maintain high value agricultural output in areas that newly receive it. Data collected at the farm level can provide such information, and detailed water balance assessments are needed to determine what portion (if any) of increased flows can be attributed to the Compact investments (rather than to natural flow variability, the Disi water supply augmentation, or increased water consumption in zones unaffected by the Compact that eventually make their way into the KTR).

E.3.2. Primary evaluation questions

With these issues in mind, the main evaluation questions related to component 2 of the IE, that is, to the changes faced by irrigators located downstream of Zarqa and the As-Samra treatment plant, are the following:

1. **Impacts on water sourcing:** Does the combined WNP/WWNP/AEP result in increased irrigation with addition blended KTR water? Does the volume of irrigation using KAC freshwater correspondingly decrease?
2. **Impacts on farming costs:** Does the combined WNP / WWNP / AEP lead to changes in farm input costs?
3. **Impacts on farm output:** Does the combined WNP / WWNP / AEP lead to changes in the value of farm output in affected areas?
4. **Impacts on asset value:** Are farm values affected by the WNP / WWNP / AEP investments?
5. **Overall impacts on farm welfare:** What is the net economic value of changes in irrigation?
6. **Impacts on compliance:** Does the AEP result in increases in the quantity of wastewater that meets effluent standards prior to discharge into the environment? (Note: This is not really an impact evaluation question, but it does seem to be the only specific impact that can be attributed to the AEP alone).

As with component 1 of the IE, besides considering each of these evaluation questions, we will aim to study the mechanisms by which they are produced, by 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. A more elaborate list of the measures to be tracked is presented in Table E.2 shown previously.

We present a single quasi-experimental design option for evaluating these questions below. We propose two critical and complementary elements within this option, which together should allow consideration of the full set of IE questions presented above. We would not advise elimination of either of these elements, since all questions depend on careful understanding of both the water balance and changes in agricultural production in affected areas downstream of Zarqa relative to other parts of the Jordan Valley.

E.3.3. Evaluation Design Options

Element A: Enhanced data collection to support future water balance calculations.

The SI team recognizes that potential substitution of reclaimed wastewater for conventional freshwater in agriculture constitutes an important part of the ERR calculation justifying MCC's investment. Attributing changes in agricultural water use to Compact activities is impossible without producing a detailed water balance calculation supported by a detailed and comprehensive dataset. This dataset would include careful measurement and/or defensible estimation of flows into the As-Samra-bound sewer network as well as inflows to and outflows from the Zarqa River upstream from the dam. Improved monitoring of both mixed effluent and conventional freshwater flows within the Jordan Valley irrigation network may also be also required. Producing such a dataset will require a sustained working partnership with a number of relevant institutions. In some cases, the data are properly collected and maintained, and may require only official agreements for data sharing. (We are aware, for example, of the recent development of the Water Management Information System (WMIS) for the JVA.) In other cases, new data collection schemes may be required. Table E.5 presents a list of variables that represent a minimum of information that will be necessary for attributing agricultural changes in the Jordan Valley to Compact activities.

As discussed in the note to Table E.5, proxies for some of the variables required for understanding the agricultural benefits of the investments are being tracked through the IT Table elaborated in the M&E plan. Unfortunately, those indicators do not allow anything close to attribution (as initially discussed in the IT Table memo included in Annex A). Several additional comments on the M&E tracking are warranted here to better understand our concerns and the rationale for the additional data collection efforts proposed in Table E.5.

Table E.5. Variables required for analysis of the role of Compact investments in allowing primary substitution of water away from Jordan Valley irrigation (water balance analysis)

Variable	Source Institution(s)	New data collection required?
New contributions to the As-Samra-bound wastewater network		
Volume of Disi or other new imports (e.g., due to national substitution policy, or new water supply development)	MWI, WAJ, PMU	Unknown
Δ in wastewater production due to population/demand growth	MWI, WAJ, PMU	Unknown
Δ in wastewater capture from Compact in already sewered areas (from reduced physical network losses)	WAJ, PMC	Yes; in Components 1 & 3
Δ in wastewater capture from Compact in newly-sewered areas	WAJ, PMC	Yes; in Components 1 & 3
Water quality of wastewater effluents		
TSS, BOD, and bacteria levels in As-Samra effluent	MWI, MoH, PMC	Unlikely
TDS, TSS, bacteria, and chlorophyll-A in KTD water	JVA, MoH	Unlikely
Water balance between As-Samra and the KTD		
Δ in withdrawals of treated effluent from the Zarqa River in the vicinity of As-Samra	JVA	Yes, in this Component 2 (remotely sensed data)
Δ in withdrawals of treated effluent from the Zarqa River downstream of As-Samra	JVA	Yes, in this Component 2 (remotely sensed data)
Inflows to the Zarqa River from the Jerash and Ba'qa WWTPs	WAJ	Unknown
Precipitation-based inflows into the Zarqa River	MWI	Unknown
Releases from the KTD Releases	JVA	Unknown
Water allocations to JV irrigation systems		
KTD water allocations to irrigated areas, by turnout	JVA	Unknown
KAC water allocations to irrigated areas, by turnout	JVA	Unknown
Weekly water allocations to individual farm units	JVA	Yes; Component 2 surveys

Note: The SI team recognizes that proxies for several of these variables are included in the MCA-J's ITT.

First, knowing the *volume of wastewater collected* from Zarqa (an IT Table indicator) over time is critical, but is only helpful for the IE if it can be placed in the context of a historical time series of data to ensure that changes are a result of Compact activities and not due to seasonal (or interannual) variation in the production of wastewater. Moreover, this attributable quantity will have to be compared with the changes in volumes of water arriving from the Amman sewer system, which may produce more wastewater as Disi comes online.

Second*, w*ithin the category of *non-revenue water* (an IT Table indicator), it is important to understand the change in physical losses due to Compact activities, which will lead to increased flows of wastewater from sewered areas of Zarqa, in contrast to administrative losses, which will not. These components are not tracked separately, and this is a serious limitation of the M&E activities, as will be discussed further in Section E.5.

Third, *treated wastewater used in agriculture* and *agricultural use of treated wastewater (*both IT Table indicators) are listed as specific to the As-Samra expansion component; yet they are really related to <u>all</u> three project activities. This is because the As-Samra expansion itself will have no impact on the volume of water used in agriculture in the Jordan Valley; it only affects effluent quality (see next paragraph). The changes in quantity come from increased water use (due to water network rehabilitation) and increased collection of wastewater (due to sewerage expansion). In addition to this, the first parameter is reported as the percentage of treated wastewater allocated in the JV relative to total irrigation water. Since absolute volumes are not tracked, there is no way of knowing whether the change in this percentage is a function of increased KTD contributions, reduced KAC contributions, or some combination of the two. Finally, as the evaluation team has pointed out, the agricultural area measure does not account for changes in volumetric usage. (For example, if the expansion of the area that receives any KTD water at all is coupled with a reduction in overall water allocations to farmers, the results on welfare will be net-negative.

As noted in Section D.2 of this document, it is critical to understand that the As-Samra expansion will likely have no effect on the quantity of water flowing to downstream farmers. This is a simple conservation of mass; the wastewater increases come from additional water consumption and collection of wastewater from Zarqa and any other areas served by As-Samra. The treatment plant expansion only affects the degree to which these additional volumes can be treated to meet effluent standards. It is therefore logical to ask whether *meeting effluent standards* (another IT Table parameter) is a relevant indicator of project impact, given that it may not indicate anything about the viability of water recycling for use in agriculture. Based on discussions with MCA-J, we have assumed that the parameters that will be tracked are total suspended solids (TSS) and biochemical oxygen demand (BOD), as is conventional for municipal wastewater management. Yet TSS and BOD, which are designed to protect aquatic ecosystems, are not relevant to suitability for agricultural reuse, which is rather related to microbial contamination (for protection of public health), chlorophyll-A (an indicator of algae, for protection of irrigation equipment), and salinity (for protection of crops and soils). In addition, TSS and BOD concentrations change by roughly an order of magnitude (from 150 to 12 parts per million) between the As-Samra WWTP and KTD due to dilution, aeration, settling, and UV exposure (Carr, 2009), and comparable reductions in TSS likely occur prior to discharge into the Jordan Valley irrigation network. The JVA should be contacted to more fully understand the nature of these changes in water quality.

In making these comments, we are not seeking to minimize the importance of the WWTP expansion as a means of protecting aquatic ecosystems to the satisfaction of international standards. We also appreciate that there could be reputational risk to Jordanian agriculture if evidence of fecal contamination is detected in JV crops and is subsequently associated (logically or otherwise) to elevated bacteria concentrations in As-Samra effluent. We do therefore recommend high-frequency measurement of fecal indicators in As-Samra and KTD water to enable evaluation of suitability of this water for irrigation at different locations downstream of As-

Samra (and commit to working with the Ministry of Health to learn more about its pathogen sampling regime and consistency with WHO guidelines for water reuse).

Element B: Longitudinal farm end-user surveys in the Jordan Valley, As-Samra vicinity, and along the downstream reaches of the Zarqa River between the WWTP and KTD.

Developing a rigorous quasi-experimental design for evaluation of the impacts of expanded wastewater availability on farmers is not realistic in this case, given that increased flows of mixed KTD water will affect all farms within areas receiving such supplies, with areas outside of the receiving zones almost certainly being systematically different with regards to soil fertility, climate, and even access to markets. However, the change in allocations of mixed water over time across large swathes of the Jordan Valley does provide a natural experiment (observational study) through which differences in productivity and net agricultural returns over time can be compared for areas that already use KTD water but may receive more, areas that do not yet rely on KTD water but will now receive some, and areas that do not and will not receive KTD water. At present, reclaimed effluent is used in varying proportions by farmers near the As-Samra WWTP, along the Zarqa River above the KTD, and by farmers in the Middle and Southern portions of the North Directorate of the Jordan Valley. There are currently many areas in the North Directorate that are unconnected to the distribution network for mixed effluent distribution, and it is unlikely these areas will all be connected during the period of the IE (see Figure E.6 below).

We would therefore propose to use a DiD methodology to isolate the effects of changes in the quantities of KAC and KTD water supplied to a representative sample of farms extending over these various regions, using a framework similar to that shown in equation 3 above. In this case, we modify the model to be:

$$Y_{ijt} = \alpha + \gamma Q^{KTD}{}_{ijt} + \delta Q^{KAC}{}_{ijt} + \beta X_{ijt} + \upsilon_i + \delta_{ijt}, \qquad (4)$$

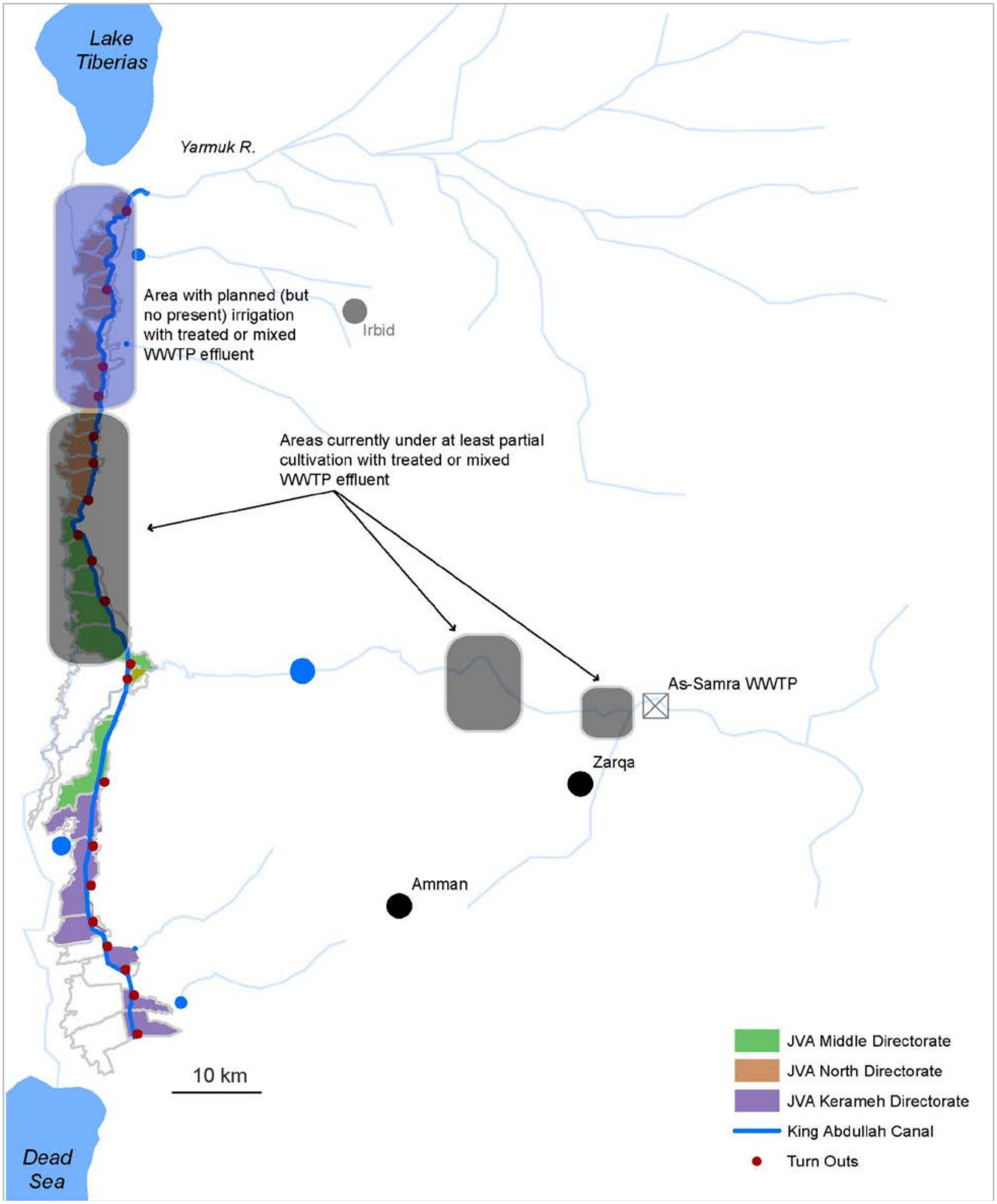
where Y_{ijt} is the outcome of interest for farm *i* in zone *j* at time *t*, $Q^{KTD}{}_{ijt}$ and $Q^{KAC}{}_{ijt}$ indicate the quantity of KTD and KAC water delivered to the farm *i* in zone *j* at time *t*, X_{ijt} is a vector of time-varying variables (farm inputs, weather, crop prices, etc.) that affect the outcome for unit *i* in zone *j* at time *t*, υ_i is a farm fixed effect, and δ_{ijt} is a time-varying error term. Within this framework, the coefficient γ measures the "treatment effect" of an additional unit of KTD water, and δ measures the treatment effect of an additional unit of KAC water. This estimate will be unbiased so long as the error term δ_{ijt} is not correlated with assignment to different water sources, a risk which is considerably reduced through the use of fixed effects panel estimation. The outcomes of interest will be yields for different crops grown in farm *i*, as well as overall net agricultural returns.

For the purposes of baseline data collection, we plan to survey roughly 550 farmers (110 farmers in each of five differentially affected areas[14]) to determine crop production and returns for the

[14] One hundred per group +10% to allow for attrition. The proposed areas are the following: 1) immediate vicinity of As-Samra; 2) between As-Samra and the KTD; 3) in Middle Valley areas currently receiving both KAC and blended KTR water; 4) in Upper Valley areas that may newly receive a mixture of KAC and blended KTR water; and 5) in Upper Valley/Yarmuk areas that are unlikely to receive blended KTR water.

previous year, along with measures of water supply from different sources. Using the data from these 550 farms, we will conduct more detailed power calculations to determine the appropriate sample size for annual tracking of the balance of water sources, production, and net profits. It is our understanding that metering is very limited in the Jordan Valley, so we will rely on self-reports of water consumption and third-party ground-truthing from the JVA and other sources. The annual or biannual surveys – conducted in summer 2015, 2016, and finally 2018 – will include questions on farmer characteristics (education, training, knowledge, relative influence, risk preferences, etc.), farm attributes (soils, canal location, etc.), farm equipment and use of advanced technology, inputs and production, animal husbandry, prices of agricultural products, and farm and non-farm sources of income. Unfortunately, we are unable to provide detailed power calculations to select a final sample size at this time since we do not have access to data on the variability of crop productivity and net returns in different parts of Jordan. It is possible that such data could be obtained from other sources – DoS farm surveys, a recent USAID-funded agriculture survey, or the JVA – and we will investigate this even as we begin planning for the baseline described above.

Figure E.6. Target locations for agricultural surveys in Jordan based on differential sourcing of irrigation water.



E.3.4 Risks and mitigation strategy

The primary risks to attribution in Component 2, and our mitigation strategy for addressing them, are summarized in Table E.6. As the above section emphasizes, attributing agricultural and farm welfare changes to Compact activities will be challenging even with an abundance of information on influent flows into and effluent flows out of the As-Samra facility, hydrological and water quality testing at different points along the Zarqa River, the King Talal Reservoir, and through the Jordan Valley irrigation networks of the North, Middle, and Kerama directorates.

Our central concern is that relevant present and historical data either do not exist or will not be readily shared with our evaluation team. Our preliminary investigations suggest that the non-functioning status of some flow instrumentation is not entirely due to wear-and-tear, as the stakes associated with the dissemination of water allocation information are quite high. We also caution that the incentives facing the institutions and people responsible for water allocation are not always consistent with open data sharing. Thus, the surest way to produce accurate information will be conduct independent measurements of flows and water quality parameters at selected locations, particularly those downstream of the As-Samra facility. This would entail dedicated personnel taking visual readings at notched weirs at selected locations as well as collecting and testing water quality samples both in situ and in the laboratory at multiple locations. The SI team realizes that there are likely to be sensitivities associated with such measures.

In the event that there is not a satisfactory transfer of information required to estimate causal relationships between Compact activities and agriculture downstream of As-Samra, a possible alternative may be to position a qualified member of the MCA-J in the offices of the JVA or the MWI (as has been mentioned with regard to data transfer from DoS in the past).In any case, these issues should be investigated during the team's next trip to Jordan.

Table E.6. Categorization of threats to identification of impacts, and mitigation strategy

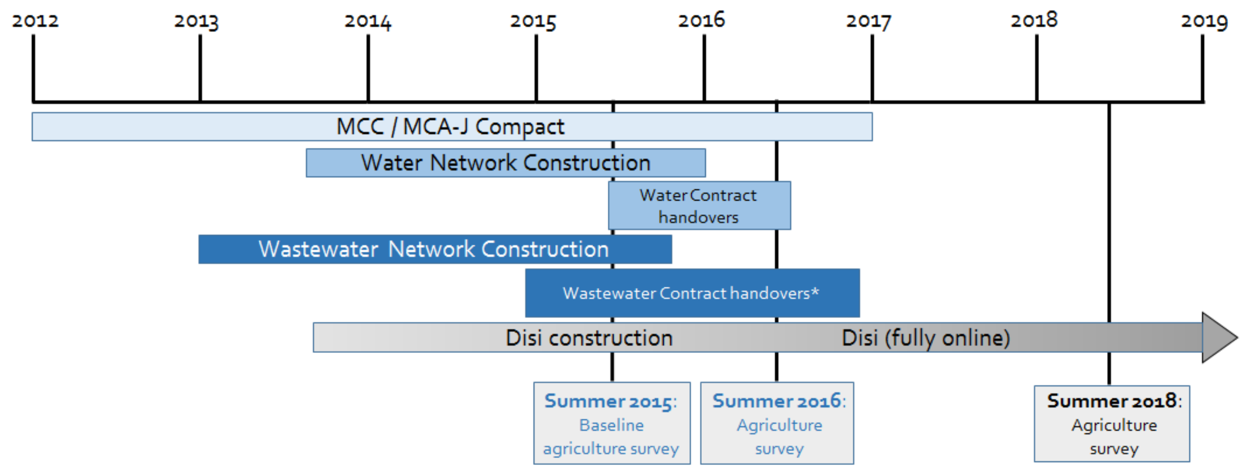
Description	Type of Risk	Mitigation strategy
Inflows of Disi water into Amman and Zarqa drinking water distribution networks cannot be determined directly or with sufficient precision	Confounder	Produce estimate based on influent volume trend at As-Samra and estimates from MWI, WAJ-Amman, and WAJ-Zarqa
Drinking water metering data at DMA level cannot be determined directly or with sufficient precision	Confounder	Produce estimate based on influent volume trend at As-Samra and estimates from MWI, WAJ-Amman, and WAJ-Zarqa
Wastewater production resulting from WWNP cannot be determined directly or with sufficient precision	Confounder	Produce estimate based on influent volume trend at As-Samra and estimates from MWI, WAJ-Amman, and WAJ-Zarqa
Zarqa River withdrawals in the vicinity of As-Samra (downstream of the treatment plant) cannot be determined directly or with sufficient precision	Confounder	Estimate using combination of: publicly available satellite image time series, interviews with JVA personnel, and field visits

Natural inflows of Zarqa River water into the KTR cannot be determined directly or with sufficient precision	Confounder	Locate streamflow proxy data from precipitation data from the Jordan Meteorological Department and/or other publicly available weather or climate models for the Jordan basin
Releases from the KTR into the Jordan Valley through the Zarqa Connectors and the Abu Zeighan Canal cannot be determined directly or with sufficient precision	Confounder	Direct measurement of inflows, outflows, and elevation of the reservoir surface on a bi-weekly basis
Inflows of Jordan-Yarmuk water into the KAC cannot be determined directly or with sufficient precision	Confounder	Unknown at this point
Specific volumetric allocations of the KAC and KTR water to Jordan Valley farmers cannot be determined directly or with sufficient precision	Confounder	Produce estimates based on contact with JVA personnel and farmer surveys

E.3.5. Timeframe and implementation of Component 2

At the time of writing this design report the following period of performance is envisioned for the As-Samra extension project:

Figure E.7. Evaluation Component 2 Timeline ((Surveys shown in blue font are complete; black font are planned)



Note: A small portion of the wastewater network project (10 km) was handed over in March 2014.

E.4 Impacts on WAJ Zarqa – Impact/Performance Evaluation

E.4.1. Preliminary considerations

Components 1 and 2 of our IE design aim to measure the welfare changes among direct beneficiaries of the water and wastewater sector interventions included in the Jordan Compact. Yet one of the very important challenges facing the IE of this program stems from the very real possibility that most of the benefits may not be directly reflected in welfare changes among households and enterprises in Zarqa, nor among the farmers who may receive additional flows of treated wastewater for their irrigation activities. Indeed, many of the benefits of the investments may be captured by the local water utility, the Water Authority of Jordan in Zarqa (WAJ-Zarqa), or by other larger government institutions responsible for water delivery in Jordan, including the central WAJ, the JVA, or the MWI. Benefits captured by these institutions could in turn lead to reductions in public debt in Jordan and free up capital for other productive economic activities.

For this reason, though we cannot estimate these types of benefits using traditional IE methods (there is no appropriate design to implement appropriate experimental or quasi-experimental control for such changes within the scope of this IE), we believe that the IE should at least track improvements in utility performance indicators, so as to provide some approximate sense of their importance. In addition, given that utility performance is subject to change due to a variety of dynamic influences, and that service improvements may benefit households in ways that are not measured through the household survey (e.g., general water supply and demand forces in Jordan; institutional reform and corporatization of utilities), there is a need for a comparative study with other WAJ units in places that are less likely to be affected by the infrastructure and other MCC investments in Zarqa.

E.4.2. Primary evaluation questions

There are thus two main questions related to utility performance, which can be summarized as follows:

1. **Impacts on utility cost recovery**: Does the net cost recovery of the utility improve due to the Compact, and is this related to service improvements?
2. **Operations and maintenance**: What is the impact of the Compact on the budget and execution of O&M?
3. **Service improvements**: At the utility level, are there measurable changes in service delivery quality trends in Zarqa relative to those of other municipal utilities in Jordan?

Of course, beneath these questions lie a series of sub-questions related to specific indicators of utility performance, including those dealing with the reliability of water delivery, magnitude of lost revenues due to NRW (physical and administrative components), the cost savings from reduced pumping requirements and more efficient operations, the reduced financial and aesthetic losses from avoided repairs to the distribution network or for management of sewer overflows. Our team has concerns about the lack of comprehensive utility monitoring in the M&E plan, which partly stems from removal of funding for important monitoring components (e.g., SCADA) during the compact development process. In light of these realities, we suggest implementing an approach that augments this regular and planned reporting of indicators, and that incorporates a manageable number of engineering tests, as described below.

E.4.3. Evaluation design

Element A: Enhanced analysis of standard indicators of system-wide and local-scale utility performance – Performance Evaluation

Comparative utility-scale performance. The Monitoring & Evaluation Unit of MCA-Jordan is already collecting a variety of utility performance indicators at the level of the WAJ-Zarqa utility, and these are useful for the evaluation. Nonetheless, we believe it important to augment these measures with additional indicators that will enable generation of a more complete picture of the performance of the water and sewer networks under utility management, as well as operational efficiency, the degree of utility cost recovery, and overall financial sustainability. These proposed measurements are consistent with typical norms for utility management / monitoring best practice, as well as with the current reporting and analysis conducted by the Jordanian water utilities (Aqaba, Yarmouk and Miyahuna) currently reporting to the Project Management Unit (PMU) of the WAJ, responsible for privatization of water utilities. Annex B provides a complete listing of the indicators we propose to monitor, which are also summarized below (Table E.7). Annex B also presents an abbreviated set of indicators which mostly entail compiling and reviewing the quality of existing data (many of these indicators are already reported to MCA-J though they may not appear in the IT Table), so as to provide synthesized information on the benefits of the project at the utility level.

Table E.7. Detailed summary of system utility-wide and local-scale performance indicators

Category	Example	Purpose
Basic Information	Connections, Customers, Meters, Line lengths, Bills Issued	Base parameters for indicator computation
Summary Water Balance	Own source production, Imports (by source), Exports (by destination), Main categories of NRW *Local-scale only*: Inflows and outflows to DMA	Base parameters for indicator computation
Operational Indicators	Hours of service, Water quality, Pressure, Pump breakdown, Leaks, Overflows, Complaints, Response times, Specific energy consumption	Provide objective indicators of service quality and efficiency and effectiveness of utility O&M
Financial Statement	Income statement including details of revenues and expenditures for water, Operating income *Local-scale*: Customer billing records, repair costs, and other O&M cost accounting in survey DMAs	Base information for key financial indicators and understanding of costs of NRW Corroboration of survey data
Key Indicators	Consumption per subscriber, Water variable cost, Real losses[a], Apparent losses[a], Hours of service, Burst rate, Sewer overflow rate, Collection efficiency, Cost recovery ratio.	Comparison to key project objectives and assumptions, especially indicators of long-term utility sustainability

[a] Note that the method for determining the real losses and apparent losses is discussed in the next section of this report.

The effort we envision here will not require large amounts of additional data collection, as many of the parameters are already reported to the MCA-J or the PMU, or are otherwise available at WAJ Headquarters. More specifically, we intend to carry out the following:

1. Improve the understanding of baseline performance through computation of indicators for 2009, 2010 and 2011, to the extent that required data are available. These tabulations would be very similar to the various analyses for 2008 that have been compiled in the various project preparation documents, but would be extended to cover additional water utilities in Jordan. This information gathering should also document the methods used (if possible) to determine or estimate the parameters.

2. Establish a protocol for utility data collection that allows base data to be gathered in a way that does not duplicate other reporting methods. For example the quarterly data collection / transmission of information from Zarqa utility to MCA-J would be the obvious starting point for an expanded data collection and analysis protocol, that could be extended to include monitoring from WAJ Headquarters in Amman (see item number 6 below). The protocol would also record the method of measurement / estimation of the parameters.
3. Optional: Develop a data accuracy rating scale for different parameters, based on rating scales developed by the American Water Works Association (AWWA), and merge it into the data collection protocol. It would be highly desirable to arrange for annual calibration of key metering infrastructure, such as production wells, import and export lines and zonal boundaries. A small local contract could achieve such an objective.
4. Begin collection of data and data accuracy ratings (preferably on a quarterly basis). Assess coherency and reliability of the information provided, and potential impact of parameter uncertainty, and make adjustments to the data collection protocol as needed.
5. Conduct basic trend analysis of key indicators alongside analyses of progress on network restructuring and other Compact activities, using the several years of baseline information.
6. Carry out comparative analyses of Zarqa and other Jordanian utilities' performance indicators, controlling for underlying differences in starting points and trends (due to geographic, population/scale, climatic, water availability, or institutional factors) across utilities. While it is not clear if sufficient information is available to conduct these comparative analyses on all parameters, SI Team members have performed such comparative analyses on a subset of indicators in other contexts, without large extra data collection efforts.

In addition to this, if funds allow, the IE team could also address the topic of "economic level of losses", which is suggested as a focus in the Investment Master Plan. This topic, which actually relates to a financial criterion for a water supplier, has been widely discussed by water analysts and researchers interested in NRW. The basic idea is that a utility can increase spending on O&M, thereby reducing losses over time, but that eventually, the marginal cost of saving NRW will exceed the marginal returns of savings. For the case of Zarqa, given the high cost of alternative water sources (e.g., imported Disi water), such a modeling exercise could be highly informative for determining how to allocate resources for O&M. It would also allow re-examination of the IT Table End-of Compact Target of 35% NRW), which appears to be far above what would typically be the optimal level (and is different from the reduction assumed in the ERR analysis). The SI team notes that a collaborative pilot project run by RTI has applied such a model to the case of Aqaba, and a paper based on the analysis has been published in a peer-reviewed IWA journal (Wyatt and Alshafey 2012). With little difficulty, the model could be applied in Zarqa, and even to WSAs or DMA sub-units within Zarqa.

Local scale intensive monitoring. In keeping with the basic scientific premises of an IE, assessment must be conducted at specific project intervention sites and control locations that are deemed comparable and are subject to different levels of treatment with the Compact interventions. In order to complement and enrich the household survey work we propose to collect certain "utility" indicators at the level of the surveyed DMAs – whether treatment or control areas.

Those indicators include a subset of the same indicators listed above for the Utility level, as noted for “local-scale monitoring” in Table E.7.

The locations for this intensive monitoring will be driven by the selection of DMAs for the household survey, to increase confidence in the results in the target areas, but also increase the richness of the operational monitoring knowledge. The former will provide a particularly valuable source of corroboration of data from the surveys given persistent concerns in the literature over the reliability or stability of self-reported survey measures (Zwane, Zinman et al. 2011). The analyses that will be performed (most likely trend and regression analyses, with an eye to extrapolating to other zones in the city) using these local data will be developed in concert with the development and pre-testing of household survey forms. Once refined, such local monitoring should be carried out at the time of surveys, allowing for its incorporation into the impact evaluation of Component 1. Ideally, at the same time, the Zarqa utility would be trained to carry out additional analyses of its own, to help pinpoint areas of operational concerns such as poor service quality, high apparent losses, etc.

<u>Element B: Engineering tests to better understand the components of Non-Revenue Water (NRW) – Impact Evaluation.</u>

One of the most important indicators associated with the Water Network Project is NRW. Yet the indicators of NRW in the M&E plan are misleading. If NRW is reported as a percentage of the water supplied, the indicator is in fact dependent on the amount of NRW *and* the water consumption, since the water supplied is the sum of the NRW and the consumption. Since the objectives of the project include both increasing network water consumption and reducing NRW, the percentage measure becomes misleading. Table E.8 illustrates this problem, using figures from the Compact IT Table. To reach a target of 35% NRW, given the targeted increase in consumption, the required reduction in losses would only have to modest, but water supply would have to be increased considerably. Also, the drop from 50% to 35% NRW (a reduction of 30%) does not correspond to the required reduction of losses from 65 lpcd to 52 lpcd (a reduction of only 20%), because the consumption has changed at the same time. The percentage of NRW is not really measuring NRW; rather it is measuring concurrent changes in both NRW and consumption.

Table E.8. Examination of Compact targets for NRW

Zarqa Water Network Project - ITT Non-Revenue Water Indicators

	Baseline	End of Compact	
Network Water Use, Lpcd	65	96	Water consumption is increased 47%
Non-Revenue Water, %	50%	35%	% NRW is decreasing 30%
Required Water Input,lpcd	130	148	Water requirements are increased 23% !!
Calculated NRW, lpcd	65	52	NRW need only be reduced by 13 lpcd or 20%

Thus, instead of % NRW, we recommend using the IWA recommended NRW indicators - Liters per Connection per day and m^3 / kilometer of mains / day. In addition, the IE should measure the 2 main components of NRW – Apparent (Administrative) Losses and Real (Physical) Losses (Table E.9). These different types of losses have different implications for project finances, the behavioral responses of and benefits flowing to potential beneficiaries, and the ultimate changes and distribution of project economic outcomes. A good baseline and accurate ongoing measurement of the separate components of NRW – unbilled, authorized consumption, real losses, apparent losses – is therefore necessary to track the impacts of the Compact investments, and especially for teasing out their effects on utility cost recovery and consumer well-being.

The conventional method of determining the real losses and apparent losses in a specific zone of interest involves the pairing of records of the frequency, flow and duration of leaks and bursts in a system with a network operational test – called *minimum night flow analysis*. During night flow analysis, utility personnel measure water consumption on several successive nights and determine the minimum flow (often around 2-3 am, when both consumption and apparent losses are likely to be low; surveys must be conducted to identify and characterize commercial or other users consuming water late at night). An accurate assessment of night time real losses can then be made from minimum night flow less known night time consumption. In addition, pressure transducers are used for water pressure measurements throughout the test period, so that pressure corrections can be applied to compute daytime real losses. Such tests require trained operational personnel and public involvement. In the case of water systems with intermittent supply, the water supply schedules may have to be modified, causing disruption to users inside and outside the test area. In fact, night flow measurements are challenging in intermittent flow situations, for a variety of reasons, and the tests can be both laborious and expensive. They are particularly difficult in baseline conditions with networks in poor condition.

Table E.9. Sub-components of Non-Revenue Water

NRW Category	Examples / Components	Impact and Potential Significance for IE
Unbilled authorized consumption	Some public buildings, line flushing, firefighting, municipal uses such as park irrigation, etc.	Decreased revenue for the utility. Billing policy changes could allow increased utility cost recovery
Apparent Loss	Meter under-registration, illegal use, database / data handling errors	Decreased revenue for the utility, lower cost of network water supplied to consumers. If apparent losses (meter error and illegal use) are significant, 1) actual water use would be under-estimated, 2) the estimates for effluent arriving at the wastewater treatment plant could be underestimated. Furthermore, reductions in apparent losses due to the Project would result in consumers paying more per unit of network water. This could reduce consumption of network water (substitution effect), and/or lead to increased expenditure (income effect) by households, both of which would decrease household welfare.

Real Loss	Leakage on main lines, service connections, service lines, reservoir overflows	Decreased revenue for the utility, lower cost of network water supplied to consumers Water not available for users, which can increase the need for rationing, curtail usage, or increase the use of more expensive imported water. If real losses are overestimated, the potential water savings from reducing them by installing new infrastructure is likely to be over-estimated (absent methods to curb apparent losses). Potential reduction of the cost of expensive imported water is then overestimated. Accurate measurement of the amount of real loss savings is critical to the economics of the project, as discussed in the ERR commentary.

In Annex D, we present a new methodology, developed by the authors of this IE design, which could make the measurement of real and apparent losses simpler and less expensive. It can reduce the number and frequency of night time engineering tests. By using limit analysis on inflows and outflows of zones, distribution areas, and DMAs, a reasonably accurate water balance (all components of NRW) may be developed, though validation of results and assumptions using conventional night flow tests would be ideal.

The first section of Annex D describes the “baseline” (2008) water balance developed in the Project Preparation documents (Investment Master Plan). As also noted in the section above on comments on the ERR, the IE team has concerns about the accuracy of the baseline water balance which estimated that nearly all the NRW consisted of real losses. Annex D provides empirical information from engineering tests in various cities (including night flow tests in Zarqa in 2006-2007), which shows much lower real losses and much higher apparent losses. In fact, the application of the new method we propose, using data for 2008, concludes that real losses are much lower and apparent losses much higher than the assumptions documented in the Investment Master Plan. The apparent losses rise from 7% of system input to 18% - 24% of system input (depending on the specific assumptions of the methodology). The corresponding fall in real losses is from 50% to 34% - 39% of system input. These calculations confirm that the baseline water balance should be re-examined lest the estimates of changes following the Compact investments be misreported.

Based on these observations, the IE Team proposes that the following procedure be carried out:

1) The general method be applied to 2009, 2010 and 2011 data and results compared with 2008.
2) At baseline and follow-up, some meter tests be conducted and compared to the results obtained using the proposed methodology in areas that correspond to the household survey, such that any necessary refinements to the method can be made, and such that the impacts of the investments on NRW can be adequately assessed.
3) Repeated similar calculations on a regular and ongoing basis throughout the duration of the IE.

4) Possibly, additional analyses be conducted to coincide with other local monitoring activities detailed above, supplemented by a small number engineering tests, to be determined based on results obtained from the three preceding activities.

Element C. Other data analysis – Impact Evaluation *(funds permitting).*

Secondary geo-coded data on water quality. Though details on existing water quality monitoring protocols and data availability in Zarqa are somewhat unclear at this time, both WAJ-Zarqa and the Ministry of Health regularly conduct tests at pumping stations, wells, and various locations in the distribution network. According to the director of the central labs for WAJ, tests are conducted:

- Three times per week at treatment stations;
- Once per week at pump stations and locations in the distribution system; and
- Once per month at wells and other "secure" sources.

The locations of these tests are apparently geo-coded by WAJ. However, we do not know the specific number of each type of sample, though the WAJ has indicated the data would be shared if requested using official channels. Our hope is to collect and analyze such data to complement the water quality assessments we conduct at the household level as described above. There would be no additional water tests included in this element; rather we would devote some effort to compiling existing data and comparing changes in time in water quality test results in Compact and non-Compact-affected areas of Zarqa Governorate.

Secondary geo-coded customer complaints, and sewer overflows. WAJ-Zarqa maintains a detailed database of consumer complaints and other system problems. This system includes customer addresses associated with each complaint that the WAJ said would be made available to the team (contingent on official requests). We will aim to incorporate such secondary data into our DiD comparisons between treatment and control areas, both within the sampling frame for our household surveys (obtained using PSM), and more generally in Zarqa as a whole (for which only DiD will be used for comparing changes over time in treated and untreated areas). As these data are already stored in a database at WAJ and complaints are tied to WAJ customer numbers, the timing of this geocoding is not critical, but could once again be done on an ongoing basis by allocating some effort to a local SI staff.

E.4.4. Timeline and implementation of Component 3

The timeline for Component 3 should be closely aligned with that for Component 1, since these provide complementary information that is critical for answering general integrating questions about the welfare implications of the Compact (see next section). Given that there are many unknowns regarding the availability of the data needed for implementing the three elements described above, we therefore recommend that the evaluation team's specialist travel to Jordan in the near future to work with the WAJ-Zarqa to develop a detailed work plan for this component. This trip will be followed by submission of a memorandum to MCC with a more concrete proposal and costing estimate related to Component 3 data collection activities.

E.5. Summary of proposed IE Activities and integration

E.5.1. Summary of components 1-3

The MCC Jordan Compact is composed of three inter-linked projects that are complementary in trying to facilitate growth and reduce poverty in Zarqa governorate. The preceding discussion has shown that the task of this IE – measuring the impact of the Jordan Compact in a manner that changes in household, enterprise and utility outcomes can be attributed to specific activities of the Jordan compact – requires careful decisions regarding method, time, and cost, which will have bearing on the nature of the questions it is able to answer. Table E.10 summarizes the IE activities for each component suggested by the evaluation team based on the discussion above. The support of the collection of suggested elements in all three components (and including the preferred option in component 1) will provide the MCC and the GOJ with the most rigorous results possible, given the challenges of conducting an evaluation of an urban water intervention.

E.5.2. Integrating questions

Assuming that the three components can be implemented as suggested, we would be able to carry out analyses at the end of the evaluation period that would attempt to answer the following (perhaps in a final synthesis report) more general questions about the aggregated impact of the Compact investments:

1. **Primary substitution effect**: What reallocation of water is made possible by the Compact investments? (i.e., what is the scale of the 'primary' substitution effect)?
2. **Economic valuation of water**: What is the comparative economic value of water consumption for different uses (domestic, commercial/ industrial, irrigation)?
3. **Distribution of benefits from MCC investment:** What are the overall net benefits from the Compact, and what are its distributional consequences?

The first of these questions requires time-series data collected in all three components, on the following types of variables (note the list is not exhaustive): volume of new water supplies (e.g., Disi) provided to Zarqa water consumption changes in areas exposed to the investments in Zarqa; sewage production from Zarqa and other zones discharging to As-Samra; contribution of other water inputs to the water system upstream of the KTD; evolution of water use from different sources (freshwater vs. blended KTR water) in irrigation. These types of variables will be measured using measures recorded through household/enterprise surveys as well as aggregate water balance indicators. It is critical that these data be collected at several intervals (as described above for each of the components) in order to better understand the evolution of the changes over time.

The second question follows mainly from analysis of data that will be collected through the various surveys conducted under Components 1 and 2: household, enterprise, and farm-level surveys. Given the probable lack of variation in water tariffs paid over time in Zarqa, the first of these will include a set of preference-elicitation questions to derive willingness-to-pay measures that will complement revealed preference data on households' actual water source choices (e.g., the tradeoffs between network, shop and tanker water). The second and third of these will be based on application of a production function approach to explore the marginal contribution of an additional unit of water provided to the enterprise and farm sectors.

Finally, the third question, which could be studied by our evaluation team at the end of the evaluation, would essentially consist of providing inputs needed to conduct a revised ERR analysis, updated with our estimates of actual changes that have occurred. If desired, the SI team could help carry out the analysis, and supplement it with distributional analysis to better understand who really benefited from the Compact investments. We have made every effort to include data collection activities that would provide the required data, although assumptions will have to be made to fill gaps related to the unobserved counterfactual in some cases.

Table E.10. Summary of proposed IE Activities and evaluation questions (*Italicized dates are planned*)

Component	IE Question(s)	Methodology[a]	Estimated Sample Size	Estimated Timing
Component 1: Impacts of infrastructure improvements on urban households and enterprises in Zarqa (WNP and WWNP)	• **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)? • **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? • **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? • **Impacts on income**: Does the WNP change HH / E income? • **Impacts on asset value**: Does the WNP / WWNP affect property/asset values? • **Overall impacts on welfare in Zarqa**: What is the net economic value of changes in quantity and quality of water consumed?[d]	**Impact Evaluation[b]** Element A: HH / E survey, Sample construction & analysis using PSM + DiD[c] Element B: Water vendor industry analyses Element C: Cross-sectional survey of newly-settled households (refugees) in survey zones	Element A: 3,440 hhs; 345 enterprises Element B: 500 vendors Element C: 1500 households	Element A: HH Baseline: Feb. 2014 HH Midline: Winter 2015; Summer 2016 HH Endline: *Summer 2018* Enterprise: 2015, *2018* Element B: *Late 2017 and early 2018* Element C: *2017* TBA
Component 2: Impacts of Compact on irrigators downstream of As-Samra treatment plant (WNP; WWNP; and AEP)	• **Impacts on water sourcing**: Does the combined WNP/WWNP/AEP result in increased irrigation with addition blended KTR water? Does the volume of irrigation using KAC freshwater correspondingly decrease? • **Impacts on farming costs**: Does the combined WNP / WWNP / AEP lead to changes in farm input costs? • **Impacts on farm output**: Does the combined WNP / WWNP / AEP lead to changes in the value of farm output in affected areas? • **Impacts on asset value**: Are farm values affected by the WNP / WWNP / AEP investments? • **Overall impacts on farm welfare**: What is the net economic value of changes in irrigation? • **Impacts on compliance**: Does the AEP result in increases in the quantity of wastewater that meets effluent standards prior to discharge into the environment?	**Impact Evaluation** Element A: Water balance modeling Element B: DiD methods comparing agricultural production at locations in the JV that do and do not receive reclaimed wastewater	Element A: n/a Element B: 550 farmers	Element A: Ongoing Element A: Ongoing Element B: Baseline Survey: Summer 2015 Follow-up surveys: Summer 2016, *2018*

Component 3: Impacts on WAJ-Zarqa	• **Impacts on utility cost recovery**: Does the net cost recovery of the utility improve due to the Compact, and is this related to service improvements? • **Operations and maintenance**: What is the impact of the Compact on the budget and execution of O&M? • **Service improvements**: At the utility level, are there measurable changes in service delivery quality trends in Zarqa relative to those of other municipal utilities in Jordan?	**Performance/Impact Evaluation** Element A: Augmented tracking of utility performance (PE) Element B: Small number of basic engineering tests (IE) Element C: Other data analysis (IE)	n/a	Ongoing data collection Meter testing: 2015, *2018*
Integration of component s	• What reallocation of water is made possible by the Compact investments? (‘**primary’** substitution effect)? • What is the comparative economic value of water consumption for different uses (domestic, commercial/ industrial, irrigation)? • What are the overall net benefits from the Compact, and what are its distributional consequences?[d]	Detailed water balance and distributional analysis	Data from other Components	n/a

[a] MCC distinguishes between two types of evaluations, 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.

[b] Element A is essential; elements B and C will be conducted if funds allow, but note implications for IE questions and internal validity.

[c] PSM = Propensity score matching; DiD = Difference-in-differences.

[d] If the elements in B are not funded, this question cannot be fully answered.

A. Sampling

The SI team conducted an analysis of the minimum sample required to produce reliable statistical estimates of key outcomes detailed in Annex C. For the water / wastewater interventions we assume a clustered, quasi-randomized evaluation design with a discrete treatment whose magnitude varies by cluster and by household within clusters. We also assume that data collection will occur before and after implementation of the Compact activities.

The purpose of the sample size estimates is to determine the minimum impact that can be detected for a given sample size. Sample size calculation includes the number of clusters in the sample and the number of households within each cluster. If the measured impact of the treatment is at least as large this minimum impact, we will be able to detect it 80 percent of the time with a given sample size. If the treatment impact is less than the calculated minimum impact, however, we are less likely to detect it.

The optimal sample size depends, among other things, on the focal outcome variable. In determining the sample size we used several alternatives: water supply in both summer and winter; household water consumption according to quarterly bill; amount of quarterly water bill; household spending on treatment shop water; quantity of water purchased in water treatment shops; monthly expenditures on treating public network water; household reported health expenditures due to water; and monthly expenditures on tanker water. These indicators were chosen because they coincide with the project logic and were available from the 2009 Water Survey conducted by the DoS.

The results of the sample size calculations are detailed in Annex C. For Component 1, the impacts of infrastructure improvements on urban households and enterprises in Zarqa, the results show that a sample size of 3,440 would provide sufficient power to detect statistically significant changes of 10% magnitude in six important outcomes: water supply, water consumption, water bill, spending on treatment shop water, quantity of water purchased in treatment shops, and monthly expenditures on water. Differences in treating water from the public network as well as self-reported health expenditures because of water consumption would be harder to detect with a sample size of 3,440, especially for the latter. Reasonable power to detect such differences would require sample sizes beyond the capability of this evaluation. The same is true for changes of a magnitude of around 5%. While we may be able to detect some the changes, it is likely that we may not achieve the power to detect statistically significant effects.

A sample size of 2,500 would be sufficient for detecting 10% differences across treatment and control groups if the sample is comprised of 4 groups (roughly 625 households per group). However, in consideration of the likelihood of Zarqa-wide spillovers, the optimal final sample includes one additional group from peri-urban zones in Amman (roughly 625 households). Accounting for 10% attrition over the life of the evaluation, 313 households are added to the final sample, such that about **3,440 randomly-selected households** are required.

B. Implementation

G.1. Institutional Review Board (IRB) Requirements

SI has an internal IRB which will be used to review and approve the study before data collection begins. Upon addressing final comments received from MCC and local stakeholders, the evaluation team will submit all surveys to SI's Internal IRB as a terminal step in the survey development process. Internal IRB approval is typically granted within two weeks of submission, whereby approval documentation will be submitted to MCC prior to formal survey implementation.

Participation in a local Jordanian IRB is not required; however, the Terms of Reference for the data collection partner ensures that DoS will assume responsibility for receiving IRB approval from a local institution. Key contacts at the local IRB will also provide guidance on local issues like respondent compensation, survey timing, instrument design, and so forth.

G.2. Analysis Plan

We have developed our evaluation design to follow the logic of the impact of the water investment as expressed in the background documents pertaining to the MCC/MCA-J investment plan (feasibility studies and other studies). The evaluation design was also influenced by discussions held with MCC and local stakeholders, which furthered the team's understanding of the project's intended logic. In our IE logic diagram (Figure D.1), the impact mechanisms have been clearly spelled out, and challenges related to attribution and confounding (as well as other risks) have been identified. Our analytical strategy, based largely on combined matching methods and statistical control with a panel data set, will further allow us to explore the extent to which these threats may call into question the impacts that we measure. To take advantages of the staged implementation of investments, we hope to collect a) three waves of data in Zarqa, at baseline, midline (when roughly half of treated units have been affected), and endline (in the final year of the evaluation period); and b) two waves of farm data (at baseline and in the final year of the evaluation period).

In short, we will examine the findings obtained using several analytical techniques to assess their consistency. With regards to impacts on enterprises and households in Zarqa, four specific measures of impacts will be compared:

1) means comparisons using post-intervention data for untreated and matched, differentially-treated units;
2) means comparisons using post-intervention data for untreated and treated units, adjusting for potential confounding variables;
3) comparisons of DiD estimates for untreated and treated units of the change in intermediate and final outcomes using pre- and post-intervention data; and
4) comparisons of DiD estimates with additional statistical control for time-varying confounders.

Evidence of differences in our estimates using these four approaches will indicate the potential severity of bias in our IE strategy.

For analysis of impacts on the agricultural sector, we will primarily focus on the development of production functions at baseline that explain variation in yields and net farm revenue as a function of inputs, farm and farmer characteristics, local weather conditions, and water supply characteristics, in an effort to better understand whether changes in water supply are likely to affect irrigators positively or negatively, in different locations affected by increased use of mixed water from the KTD.

In addition to this, we will conduct *ex post* balance tests on our matched baseline samples, to assess the degree to which these may be different, and, if sample size permits, to improve the quality of our matches. This phased analysis will allow adjustments to be made prior to the launch of the follow-up surveys, and will help in our identification of potential confounding factors that should be included in the statistical adjustments described in strategies 2 and 4 described above. It will also serve as a valuable quality control strategy, since it will be possible to identify and revise problematic questions prior to the follow-up survey.

Finally, to ensure that the correct mechanisms for impacts are identified and to obtain a richer understanding of the effects of the Compact, we will complement these rigorous quantitative analyses with qualitative information collection from semi-structured interviews and focus group discussions (FGDs) with utility personnel (to obtain insights on impacts with regards to the performance of WAJ-Zarqa), targeted beneficiaries (poor and non-poor households, men and women, enterprises and farmers), and sectors that may have been adversely affected by the investment program (water shops and tankers). These interviews will first take place prior to baseline data collection and will inform the baseline survey design. We will then repeat these activities with the same informants at annual intervals, and just prior to the follow-up survey.

<u>G.2.1. Baseline Analyses</u>

As noted above, we will conduct a variety of analyses following baseline data collection. This will include:

1) Initial review of data quality, identification and correction of data errors, and cleaning of variables.
2) <u>Generation of summary baseline statistics</u> on outcomes and covariates.
    a. Descriptive statistics of outcome variables and covariates
    b. Statistical analysis of outcomes and covariates by treatment group (in Zarqa samples) or key location (in farm samples)
    c. Comparison of results from balance tests on matched samples using pre- and post-baseline data (including tests of statistical significance); evaluate the need for *ex-post* matching to re-balance sample.
    d. Generation of GIS maps for key outcomes and covariates averaged at the sampling unit level, and linking with GIS data obtained from WAJ-Zarqa
3) Analysis of the determinants of outcomes at baseline (water sourcing/choice and water-use behaviors; self-reported health outcomes; agricultural and enterprise-level production and use of water), as a function of covariates such as socio-economic status and other household characteristics, location, etc. Such observational analyses will not be sufficient to make causal arguments, but they do at least provide an initial understanding of the

associations between covariates and outcomes (for example allowing us to explore whether the amount and/or perceived or real quality of water delivered to households is correlated with purchases of non-network water). We also expect to measure the demand or preferences for different features of water supply, using stated preference techniques or revealed preference techniques, exploiting natural variation in the total prices households face for their water (Whittington, Mu et al. 1990; Whittington, Lauria et al. 1991; Nauges and Whittington 2010; Orgill, Jeuland et al. 2012).

4) Baseline Report detailing findings and analyses. SI will submit a draft report to MCC within two months of receiving a clean dataset from the data collection partner. Upon presentation to MCC and MCA-J, the draft report will be revised to incorporate comments and resubmitted in final form shortly thereafter. Summary of baseline findings in a summary report.

G.2.2. Interim Analyses at Follow-up Intervals

Midline data collection is envisioned for the household survey, while annual follow-up data collection is planned for the agricultural survey. After each round of data collection, analyses will be conducted to track progress since baseline analyses. Interim analyses are intended to provide additional opportunities for observing differences in outcome variables over the course of the multi-year evaluation. The precise timing and feasibility of interim reporting is yet to be determined.

The final endline analysis, presented in its own report, will draw from baseline and follow-up survey data to include:

1) Generation of summary statistics at follow-up, including changes in outcomes and statistical significance.
    a. Differences in all outcome variables by treatment group in Zarqa with and without statistical adjustment for potential confounders; tests of significant differences.
    b. DiD analysis of all outcomes, with and without statistical adjustment for confounders; tests of significance tests.
    c. Generation of GIS maps for changes in key outcomes and covariates averaged at the sampling unit level, and linking with GIS data obtained from WAJ-Zarqa
2) Modeling of the mechanisms of changes in ultimate outcomes, as a function of intermediate changes (using instrumental variable of treatment status).
3) Test for the importance of variation in treatment intensity across units in Zarqa; implement GPSM as needed (Dehejia and Wahba 2002). The continuous treatment approach embodies the central idea that there is in reality only one "group" in Zarqa, because all households are treated (there is no control group), and what varies is the intensity of the treatment assigned to each household. Hypothesis testing is complicated by the need to test joint hypotheses (rather than using t statistics, we need F statistics). Software for estimating treatment effects is directly available in Stata (Bia and Mattei 2008).
4) Final evaluation report presenting a detailed discussion of findings will be submitted to MCC following receipt a clean dataset. A draft report will be presented to MCC and MCA-J for feedback, and comments will be incorporated into a final evaluation report to be submitted for approval and publication online.

G.3. Work Plan and Key Deliverables

The basic work plan and timeline for the IE is presented in Table F.1 below. We will aim to take baseline measurements as soon as possible and prior to the investments delivering benefits; though we cannot avoid the effect of anticipation of benefits since some construction work has already been initiated. Thus, the aim is to collect baseline data during February-March 2014.

Pending funding availability, interim data collection under Component 1 is envisioned to be timed according to the sampling strategy and implementation progress in order to take advantage of differences in exposure to treatment (see risk mitigation considerations above). Ideally, we will conduct interim data collection once half of each type of sample (A, B and C) has been treated, which may not coincide perfectly. Funding permitting, we will then conduct additional interim data collection once all infrastructure improvements are complete. More precise details on this timeline will emerge once implementation schedules become clearer and data to construct the sampling frame become available.

As described in the SOW provided to SI by the MCC, the IE will produce several deliverables. These key products will include a baseline report, a final evaluation report, and associated datasets. The baseline data report will include a summary of baseline analyses described above, include preliminary qualitative research findings, and discuss the suitability of the implemented sampling design for addressing the evaluation questions. Once data collection activities are complete, SI will generate a final report, and cleaned dataset, which will be shared with MCC and key stakeholders for review and comment before drafts are finalized. SI will present and share documents with MCC, MCA-J, and other stakeholders as requested. SI will also endeavor to present baseline and final evaluation findings at academic meetings and in peer-reviewed journals, subject to approval from MCC.

Table G.1. Work Plan

Task Area 1: Evaluation Plan			
Deliverable	**Comments**	**Period Active**	**Due Date**
1. Scoping trip SOW	SI staff will draft a Scope-of-Work for the initial scoping trip and submit for approval to MCC.	November - December 2012	December 21, 2012
2. Scoping trip report	MCC will provide a template for the trip report. The PI and SI staff will report on scoping trip, making recommendations for IE design and outlining major data needs, as well as comment on MCC's IT Table and the Economic Rate of Return (ERR) calculations.	January - February 2013	February 15, 2013
3. Preliminary IE design	Using the trip report, SI will develop a preliminary IE design (template or outline to be provided by MCC) which outlines study methodology, data collection plans, and analysis options. SI will submit the report for possible internal peer review by MCC.	February - March 2013	March 29, 2013
4. Final IE design	After receiving comments from MCC, SI will finalize IE design.	March – June 2013	September 2013
5. IE Executive Summary/Website Description	IE design will be summarized and disseminated in executive summary form for non-technical experts to read and understand. This summary will also be translated into Arabic. IE design will also be explained in a one-page summary for posting on the MCC website.	September 2013	September 2013
6. Work and Implementation Plan	The Work Plan will include the specific team members for each task under this task order and specific dates for each deliverable. Work Plan will be in Gantt chart form and outline major tasks, detailed to the month and quarterly levels.	September - October 2013	October 2013
7. Staffing Agreement	SI will describe the arrangements for field staffing, and provide a copy of any agreements with local Jordanian entities.	October 2013	October 2013
Task Area 2: Survey Development			
Deliverable	**Comments**	**Period Active**	**Due Date**
1. Sample Design for each survey identified	SI will develop a sample design for each survey identified including methods to optimize power within budget constraints.	November 2013 – February 2014	February 2014
2. Questionnaire for each survey identified	SI will inform future data collection needs for a household survey. SI will provide comments and advice to MCC on questionnaire design and content, strategies and techniques to minimize the non-response rate, minimizing threats to internal validity and data analysis.	October 2013 – March 2014	March 2014
3. Work Plan and expected timeline for survey administration	The Work Plan and expected timeline for survey administration will include the specific team members for each task and specific dates for each deliverable. Work Plan will be in Gantt chart form and outline major tasks, detailed to the month and quarterly levels and ensuring coordination between surveys.	September - November 2013	November 2013
Task Area 3: Evaluation Implementation and Baseline Data Collection Support			

Deliverable	Comments	Period Active	Due Date
1. Ongoing MCA-J support	Provide support to MCA-J staff as needed	January 2013 – September 2017	
2. Ongoing support to data collection firm on enumerator training	SI will coordinate with data collection partner, revise/develop data collection instrument as necessary and set systems in place for successful baseline data collection. SI will provide a brief report on this.	September 2013 – March 2017	Reports submitted yearly
3. Ongoing support to data collection firm on survey pretest	SI will advise the data collection firm as it designs and administers a pre-test of the draft questionnaires. SI will provide a brief report on this.	September 2013 – March 2017	Reports submitted yearly
4. Assessment of data entry and management process	SI will provide quality assurance oversight for data collection activities, including a review of the quality of the data, identifying potential errors to be resolved and propose methods to address any errors. SI will provide a brief report on this.	September 2013 – March 2017	Reports submitted yearly
5. Assessment of raw data files and feedback to survey firm	SI will provide quality assurance to the data collection firm in producing a cleaned raw dataset to include STATA .do files and final data files for documentation purposes. SI will provide a brief report on this.	September 2013 – March 2017	Reports submitted yearly
6. Final public use data files	Working with the local data collection firm, SI will submit to MCC all final public use data files with corresponding STATA code and metadata.	September 2013 – March 2017	Data submitted yearly
Task Area 4: Baseline Data Analysis, Reports and Dissemination			
Deliverables	**Comments**	**Period Active**	**Due Date**
1. Draft Project Baseline Report for each Activity	Once survey and/or other relevant data are collected and cleaned, SI will conduct a full data analysis of the baseline data to assess baseline conditions of the Projects and Activities. Analysis should include a preliminary assessment of the research questions using baseline data, the extent to which anticipated project results may be attributable to the actual intervention, and other findings about the intervention.	March – May 2014	May 2014
2. Presentation and feedback to Evaluation Review Committee on Draft Project Baseline Report	The SI team will present the Project Baseline Report to the MCC Evaluation Review Committee and MCA-J stakeholders. SI will also document the ERC's comments and provide its own response to those comments for documentation purposes.	May – June 2014	June 2014
3. Final Project Baseline Report for each Activity	After all comments have been received, the SI will revise and finalize the draft report.	May – June 2014	June 2014
4. Executive Summary	An executive summary of the Project Baseline Report will be written for a broad, less technical audience, and will be approximately 10 pages in length in both English and Arabic.		June 2014

5. Outreach Session – Jordan	With support from the local MCA's outreach team, SI will conduct an outreach session in Jordan for local academics, government statisticians, development professionals, and other interested parties to discuss the IE's implementation, baseline findings, lessons learned, and other relevant topics.	June - July 2014	
6. Outreach Session – Washington, DC	SI will conduct an outreach session in Washington, DC for MCC staff, development professionals and other interested parties to discuss the IE's implementation, baseline findings, lessons learned, and other relevant topics.	June - July 2014	
7. Updated Baseline Report	SI will ensure that the Baseline Report is properly updated to reflect any changes in timeline, budget, risk analysis, data access, and dissemination procedures.	July 2014	
8. Support MCC to revise ERR for each Activity	The SI team will work with MCC to inform the revision of ERR estimates for each Activity being evaluated.		
Task Area 5: Support to MCA-J between Data Collection Periods			
Deliverables	**Comments**	**Period Active**	**Due Date**
1. Ongoing support to MCA-J	The SI team will provide guidance to MCA-J on any tasks related to the Water Project Evaluation as needed.	January 2013 – September 2017	
2. Qualitative research on interim project results	SI will conduct qualitative research on interim project results, if necessary and deemed useful, in order to describe interim impact results at Compact closeout.	June 2014 – September 2017	
Task Area 6: Evaluation Implementation and Endline Data Collection Support			
Deliverables	**Comments**	**Period Active**	**Due Date**
1. Provide ongoing support to MCA-J on oversight of data collection efforts	SI will provide support to MCA-J staff as needed.	June 2013 – September 2018	
2. Provide ongoing support to data collection firm on enumerator recruiting, training and mobilization	SI will coordinate with data collection partner, revise/develop data collection instrument as necessary and set systems in place for successful baseline data collection. SI will provide a brief report on this.	June 2013 – September 2018	
3. Provide ongoing support to data collection firm on survey pretest	SI will advise the data collection firm as it designs and administers a pre-test of the draft questionnaires. SI will provide a brief report on this.	June 2013 – November 2018	

4. Assessment of raw data files and feedback to survey firm	SI will provide quality assurance oversight for data collection activities, including a review the quality of the data, identifying potential errors to be resolved and propose methods to address any errors. SI will provide a brief report on this.	June 2013 – February 2019	
5. Final public use data files	SI will provide quality assurance to the data collection firm in producing a cleaned raw dataset to include STATA .do files and final data files for documentation purposes. SI will provide a brief report on this.	June 2013 – September 2019	
Task Area 7: Endline Data Analysis and Final Evaluation Reporting			
Deliverables	**Comments**	**Period Active**	**Due Date**
1. Draft IE Report for each Activity	Once survey and/or other relevant data are cleaned, the SI team will conduct a full data analysis of the baseline and follow-up data to assess the impact of the Activities, and will validate evaluation design, demonstrate baseline difference in means tests between treatment and control, power analysis to confirm appropriate sample size, suitability of the sample for the evaluation (where applicable), estimate interim and final results and update beneficiary analysis with any new data and new information available.		June 2019
2. Presentation and feedback to Evaluation Review Committee on draft IE Report	SI will present the IE Report to the MCC Evaluation Review Committee.		October 2017
3. Presentation and feedback to Jordan stakeholders on draft IE Report	Present IE Report to stakeholders' workshop in Jordan that includes a presentation of the report and results and feedback discussion. Former MCA-Jordan staff, implementers, relevant government agencies, and data collectors should all be involved in this workshop. Document ERC and Jordan stakeholder's comments and Evaluator's response to those comments for documentation purposes.		October 2017
4. Final IE Report for each Activity	After all comments have been received, the SI will revise and finalize the draft report.		November 2019
5. Executive Summary IE Report	An executive summary of the Project Baseline Report will be written for a broad, less technical audience, and will be approximately 10 pages in length in both English and Arabic.		November 2019
6. Support to MCC on revision of ERRs for each Activity	The SI team will work with MCC to inform the revision of ERR estimates for each Activity being evaluated.	July 2014 – November 2017	
Task Area 8: Dissemination of Results			
Deliverables	**Comments**	**Period Active**	**Due Date**
7. Outreach Sessions – Washington, DC	SI will conduct 3 to 4 outreach sessions for each different evaluation in Washington for MCC staff, development professionals, and other		September 2019

	interested parties to discuss the IE's implementation, lessons learned, relevant interim and final results, and other relevant topics.		
8. Outreach Sessions – Jordan	SI will also conduct outreach for a similar set of stakeholders in Jordan and should work to ensure that information is disseminated to the most relevant parties. These presentations will be based on the final reports and would be in addition to the ERC and stakeholders' workshops for feedback on the draft report.		September 2019

C. Evaluation Team

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 H.1. as follows:

Table H.1. Evaluation Team Roles and Responsibilities

Position	Responsibilities
Senior Analyst/ Water Specialist **Marc Jeuland**	Dr. Jeuland will serve as the technical and methodological lead. He will be heavily involved in the evolution of the proposed IE design throughout consultations with MCC DC staff and MCA-Jordan. Dr. Jeuland will lead the IE design and ERR activities, manage any changes to the design required during the implementation process and provide guidance to data analysis, consulting with the Senior Network Engineer, Dr. Albert, as necessary. He will contribute to written sections of evaluation reports, and other project deliverables, including taking the lead on the final IE report.
Program Manager **Mateusz Pucilowski**	Mr. Pucilowski will be the primary responsible party for technical work and deliverables and will manage the evaluation design and implementation process. From the SI HQ in Arlington, VA, Mr. Pucilowski will supervise all staff working on this contract and facilitate communication between all key personnel, non-key personnel, local data collection firms, and MCC and MCA-Jordan. Mr. Pucilowski will oversee both quantitative and qualitative data collection. Mr. Pucilowski also will monitor procurement and expenditures to assure compliance with contract and budgetary terms and rates and will provide overall quality control on the production and dissemination of all project deliverables, including the final IE reports.
Statistician/ Sampling Expert **Danae Roumis**	Ms. Roumis will advise on statistical and sampling issues. She is responsible for designing and implementing the sampling framework being implemented in the study. She will oversee the technical aspects of the propensity score matching and survey sampling design.
Gender Specialist **Jennifer Mudge**	Ms. Jennifer Mudge will implement gender integration components in evaluation design and implementation. She will develop specialized indicators to measure the impact of MCA Jordan Compact programs on women and will use this data to provide a thorough gender analysis.
Junior Analyst **Billy Hoo**	Mr. Hoo will support the Program Manager and Senior Analysts with project coordination, data collection, data analysis, and coordination with MCA during data collection and data analysis.
Junior Analyst **Daniel Hudner**	Mr. Hudner will support the Program Manager and Senior Analysts with project coordination, data collection, data analysis, and coordination with MCA during data collection and data analysis.
Program Assistant **Alison Smith**	Ms. Smith will support the Program Manager and Senior Analysts with project coordination, data collection, coordination with MCA during data collection and data analysis.

REFERENCES

Al-Salaymeh, A. and M. Al-Salaymeh (2008). The Environmental Impact and the Economic Feasibility of Disi-Amman Water Conveyor Project. 3rd International Conference on Water Resources and Arid Environments and the 1st Arab Water Forum.

Bennear, L., A. Tarozzi, et al. (2012). "Impact of a randomized controlled trial in arsenic risk communication on household water-source choices in Bangladesh." Journal of Environmental Economics and Management.

Bia, M. and A. Mattei (2008). "A Stata package for the estimation of the dose–response function through adjustment for the generalized propensity score." The Stata Journal **8**(3): 354-373.

Bucknall, J., A. Kremer, et al. (2007). Making the most of scarcity: accountability for better water management results in the Middle East and North Africa, World Bank Publications.

Cairncross, S. and P. J. Kolsky (1997). "Re: "Water, waste and well-being: A multi-country study"." American Journal of Epidemiology **146**(4): 359-360.

CARE. 2013. Syrian Refugees in urban Jordan: Baseline Assessment of Community-Identified Vulnerabilities Among Syrian Refugees Living in Irbid, Madaba, Mufraq, and Zarqa.

Chaudhury, A.S. (2013). “Can private water vendors help meet the Millennium Development Goals? A study of Karachi city urban water market” SSRN Working Paper.

Clasen, T., I. Roberts, et al. (2006). "Interventions to improve water quality for preventing diarrhoea." Cochrane Database Syst Rev **3**.

Clasen, T., W.-P. Schmidt, et al. (2007). "Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis." Bmj **334**(7597): 782.

Cutler, D. M. and G. Miller (2004). The role of public health improvements in health advances: the 20th century United States, National Bureau of Economic Research.

David, C. & A. Inocencio (1998). “Understanding Household Demand for Water: The Metro Manila Case.” Unpublished manuscript.

Davis, J., A. Kang, et al. (2001). "How important is improved water infrastructure to microenterprises? Evidence from Uganda." World Development **29**(10): 1753-1767.

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

Devoto, F., E. Duflo, et al. (2011). Happiness on tap: Piped water adoption in urban Morocco, National Bureau of Economic Research.

Eisenberg, J. N., J. C. Scott, et al. (2007). "Integrating disease control strategies: balancing water sanitation and hygiene interventions to reduce diarrheal disease burden." Journal Information **97**(5).

Esrey, S. A., J. B. Potash, et al. (1991). "Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma." Bulletin of the World Health Organization **69**(5): 609.

Fewtrell, L., R. B. Kaufmann, et al. (2005). "Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis." The Lancet infectious diseases **5**(1): 42-52.

Galasso, E., M. Ravallion, et al. (2004). "Assisting the transition from workfare to work: A randomized experiment." Indus. & Lab. Rel. Rev. **58**: 128.

Galiani, S., P. Gertler, et al. (2005). "Water for life: The impact of the privatization of water services on child mortality." Journal of Political Economy **113**(1): 83-120.

Galiani, S., M. Gonzalez-Rozada, et al. (2009). "Water expansions in shantytowns: Health and savings." Economica **76**(304): 607-622.

Gamper-Rabindran, S., S. Khan, et al. (2008). The impact of piped water provision on infant mortality in Brazil: A quantile panel data approach, National Bureau of Economic Research.

Government of Jordan (GoJ). 2014. Final Draft: National Resilience Plan 2014-2016: Proposed Priority Responses to Mitigate the Impact of the Syrian Crisis on Jordan and Jordanian Host Communities. Available at http://www.jordanembassyus.org/sites/default/files/NRP_FinalDraft_08.29.2014_MOPIC.pdf.

Hamoudi, A., M. Jeuland, et al. (2012). "The Effect of Water Quality Testing on Household Behavior: Evidence from an Experiment in Rural India." The American journal of tropical medicine and hygiene **87**(1): 18-22.

Heckman, J. J., H. Ichimura, et al. (1998). "Matching as an econometric evaluation estimator." The Review of Economic Studies **65**(2): 261-294.

Heckman, J. J., H. Ichimura, et al. (1997). "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme." The Review of Economic Studies **64**(4): 605-654.

Jalan, J. and M. Ravallion (2003). "Does piped water reduce diarrhea for children in rural India?" Journal of econometrics **112**(1): 153-173.

Jasem, A. H., M. a. Shammout, et al. (2011). "The Fate of Disi Aquifer as Stratigic Groundwater Reserve for Shared Countries (Jordan and Saudi Arabia)." Journal of Water Resources and Protection **3**: 711-714.

Jeuland, M. (2012). Creating incentives for more effective wastewater reuse in the Middle East and North Africa. Environmental Incentives and Regulation. H. Abou-Ali, Edward-Elgar Press.

Kijne, J. W., R. Barker, et al. (2003). Water productivity in agriculture: limits and opportunities for improvement, Cabi.

Korfali, S.I. & Jurdi, M. (2009). "Provision of safe domestic water for the promotion and protection of public health: a case study of the city of Beirut, Lebanon." Environmental Geochemistry and Health 31:283–295.

Lee, W.-S. (2006). "Propensity score matching and variations on the balancing test." Empirical Economics: 1-34.

Lewin, S., C. Stephens, et al. (1997). "Reviewing the health impacts of improvements in water supply." Urbanisation and health newsletter(32): 34.

Lokshin, M. M. and R. Yemtsov (2003). Evaluating the impact of infrastructure rehabilitation projects on household welfare in rural Georgia, World Bank Publications.

Luoto, J., D. Levine, et al. (2011). "Information and persuasion: achieving safe water behaviors in Kenya."

Mosler, H.-J. (2012). "A systematic approach to behavior change interventions for the water and sanitation sector in developing countries: a conceptual model, a review, and a guideline." International Journal of Environmental Health Research **22**(5): 431-449.

Nauges, C. & J. Strand. (2007). "Estimation of non-tap water demand in Central American cities." Resource & Energy Economics 29: 165–182.

Nauges, C. and D. Whittington (2010). "Estimation of water demand in developing countries: An overview." The World Bank Research Observer **25**(2): 263-294.

Nimri, L. F., Z. Elnasser, et al. (2004). "Polymicrobial infections in children with diarrhoea in a rural area of Jordan." FEMS Immunology & Medical Microbiology **42**(2): 255-259.

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.

Orgill, J., M. Jeuland, et al. (2012). "Water quality perceptions and willingness to pay for clean water in peri-urban communities in Cambodia." Under Review.

Patil, S. R. and S. K. Pattanayak (2010). Behaviors exposed: panel data evidence on environmental health externalities in rural India. Proceedings of the 3rd World Congress of Environmental and Resource Economists.

Pattanayak, S. K., C. Poulos, et al. (2010). "How valuable are environmental health interventions? Evaluation of water and sanitation programmes in India." Bulletin of the World Health Organization **88**(7): 535-542.

Pattanayak, S. K., J.-C. Yang, et al. (2009). "Shame or subsidy revisited: social mobilization for sanitation in Orissa, India." Bulletin of the World Health Organization **87**(8): 580-587.

Pickering, A. J. and J. Davis (2012). "Freshwater availability and water fetching distance affect child health in sub-Saharan Africa." Environmental science & technology **46**(4): 2391-2397.

Puri, S., H. Wong, et al. (1999). The Rum-Saq Aquifer Resource–Risk Assessment for Long Term Resource Reliability. Proceedings of UNESCO Conference on Regional Aquifer Systems, Tripoli.

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.

Schwartz, J. B. and R. W. Johnson (1992). Maximizing the Economic Impact of Urban Water Supply and Sanitation Investments, Water and Sanitation for Health Project.

Stoler, J.; J. Weeks & G. Fink (2012). “Sachet drinking water in Ghana’s Accra-Tema metropolitan area: Past, present, and future.” Journal of Water Sanitation & Hygiene for Development 2(4): doi:10.2166/washdev.2012.104.

Strand, J. & I. Walker (2005). “Water markets and demand in Central American cities.” Environment & Development Economics 10: 313–335.

Thompson, J. et al. (2000). “Waiting at the tap: changes in urban water use in East Africa over three decades.” Environment & Urbanization 12(2): 37-52.

UNHCR. 2014. Living in the Shadows: Jordan Home Visits Report 2014.

UNHCR. 2016. “Syria Regional Refugee Response.” Available at: http://data.unhcr.org/syrianrefugees/country.php?id=107.

Waddington, H., B. Snilstveit, et al. (2009). Water, sanitation and hygiene interventions to combat childhood diarrhoea in developing countries, International Initiative for Impact Evaluation New Delhi. **31**.

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

Whittington, D., M. Jeuland, et al. (2012). "Setting Priorities, Targeting Subsidies among Water, Sanitation, and Preventive Health Interventions in Developing Countries." World Development.

Whittington, D., D. T. Lauria, et al. (1991). "A study of water vending and willingness to pay for water in Onitsha, Nigeria." World Development **19**(2): 179-198.

Whittington, D., X. Mu, et al. (1990). "Calculating the value of time spent collecting water: Some estimates for Ukunda, Kenya." World Development **18**(2): 269-280.

Whittington, D. & C. Nauges (2009). “Estimation of Water Demand in Developing Countries: An Overview.” World Bank Research Observer 25:263–294.

World Bank. (2004). “World development report 2004—Making services work for poor people.” New York: Oxford University Press.

Wyatt, A. and M. Alshafey (2012). "Non-revenue water: financial model for optimal management in developing countries: application in Aqaba, Jordan." Water science and technology: water supply **12**(4): 451-462.

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.

Zérah, M.-H. (1998). "How to assess the quality dimension of urban infrastructure:: The case of water supply in Delhi." Cities **15**(4): 285-290.

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

Zwane, A. P., J. Zinman, et al. (2011). "Being surveyed can change later behavior and related parameter estimates." Proceedings of the National Academy of Sciences **108**(5): 1821-1826.

ANNEXES

A. IT Table Memorandum (January 24, 2013)
B. Utility Base Data Indicators
C. Sampling Considerations
D. New Method for Determining Water Loss Components in Water Distribution Systems
E. Component 3: Impacts on Water Authority of Jordan-Zarqa – Details of impact/performance evaluation
F. Component 4: Evaluation design for the Water Smart Homes activity
G. Proposed evaluation items that were not approved for funding

Annex A: IT Table Memorandum

To: Lola Hermosillo, Barry Deren, Amjad Attad, Raed Zahrawi
From: Marc Jeuland, Jeff Albert, Luca Etter, Erika Holzaepfel
Date: January 24 2013
Re: Comments on the IT Table and its Relation to the Jordan Compact Impact Evaluation

This memorandum summarizes comments from the SI Impact Evaluation (IE) team to the MCC and MCA-J, regarding the Indicator Tracking Table for M & E and its relation and usefulness to our IE of the Jordan Compact.

<u>General comments</u>

1. The IT Table contains a number of indicators that will be of great value to the IE, either because they directly measure outputs or outcomes that are relevant to the impacts we will measure, or because they serve as useful proxies for such outputs and outcomes (see attached Project Logic). Those specific indicators are identified in the table below and on the logic diagram:

No.	IT Table indicator	Direct measure in IE logic? (Y/N)	Corresponding measure in IE logic
1	Network water consumption per capita	N	Change in per capita **usage** of utility water
2	Operating cost coverage	Y	Increased utility cost recovery
3	Non-revenue water	Y	Reduced NRW
4	Continuity of supply time	N	Increased # hrs and continuity of **water service**
5	Replacement of customer meters	N	Increased coverage of served population with functioning meters
6	Sewer blockage events	N	Reduced frequency of overflows
7	Volume of wastewater collected	Y	Increased capture of municipal sewage
8	Residential population connected to the sewer system	Y	Reduced reliance on cesspits
9	Quality of As-Samra effluent discharged meets standard	N	Compliance with effluent standards
10	Volume of wastewater effluent discharged from the As-Samra plant per year	Y	Increased treatment of municipal sewage
11	Agriculture use of treated wastewater (hA)	N	Increased usage of treated sewage for irrigated agriculture (Volume)
12	Use of tanker water	Y	Reduced purchase of tanker water by consumers
13	Use of treatment shop water	Y	Reduced purchase of shop water by consumers
14	Incidence of diarrhea	N	Reduced disease
15	Customer dissatisfaction with water quality	N	Improved water quality at point of delivery
16	Households cleaning their water storage facilities	N	Improved household water practices

2. It does appear that there is an assumption that several of the IT Table indicators (specifically 12-16) are actually to be measured by the SI team, through IE surveys and/or other data collection activities. While this is very reasonable, it should be clarified that the statistics derived from our IE surveys will not be directly comparable to the 2009 Water Survey baseline results given that the IE sample will not be a representative sample for Zarqa.

3. Several indicators in the IT Table have a single baseline value that we believe is likely to be misleading. Socio-hydrological systems in this part of the world are highly seasonal, and interannual variability is high, such that a single value gives a very messy understanding of the baseline. It would be preferable to obtain a retrospective time series of at least 5-10 years for the following types of measures: Water consumption and demand indicators; incidence of diarrhea; continuity of supply time; sewer blockages / overflows; volume of wastewater (collected and treated); treated wastewater used in agriculture; agriculture use of treated wastewater.

4. A number of the indicators in the IT Table that are identified above are not very clear or only partially measure things that are relevant to the project logic. See more specific comments on the indicators of use to the IE below.

<u>Specific comments on indicators in the M&E IT Table that are needed for the IE</u>

We find that a number of the indicators in the IT Table are difficult to interpret, and detail our observations and suggestions below, understanding that some of the definitions have evolved beyond what is noted in the M&E plan dated March 2012.

1. **Network water consumption per capita**: We recommend that this indicator be disaggregated into separate indicators for residential, non-residential and public water usage per capita. In addition, this indicator should be re-labeled as it appears to only include "billed" water consumption (and perhaps public uses, though that is not clear). Actual network water consumption is much higher due unless administrative losses are nil. Also, for M&E purposes, it may be better to measure this in per billed customer, rather than per capita, because the latter requires an assumption about household size.

2. **Operating cost coverage**: Since this indicator has to be generated by two other indicators (revenues and costs), why not disaggregate them to report information on both cost and revenue over time?

3. **Incidence of diarrhea**: The source for this indicator is unclear (the 2009 Water Survey did not measure incidence), and is not likely to be comparable to a survey measure we will obtain.

4. **Customer dissatisfaction with supply service / water quality:** The 2009 Water Survey questions that generated data for these two indicators will likely be revised in our baseline survey to be consistent with best practices for measuring subjective perceptions of water service.

5. **Non-revenue water**: NRW should not be reported as a percentage. This is misleading, because it is being compared to consumption, which is changing. IWA recommends against this. The PMU uses Liters / Connection / Day which is much better. This indicator should also have a physical portion and administrative portion, though we acknowledge those may be unknown.

6. **Continuity of supply time**: This indicator should be termed "Continuity of pumping time" since supply time at the household level (the outcome of interest) is highly varied within Zarqa.

7. **Households cleaning their water storage facilities**: This indicator as presented is not very useful, given the complexity of water storage behaviors. Best practice for assessing the adequacy of storage cleaning behaviors relies on frequency of specific cleaning practices (e.g., disinfection).

8. **Water Smart Homes activity indicators**: The IE team finds several of these indicators to be quite confusing and actually not specific to WSH. The number of NAF households with improved water and wastewater network connections (overall and female headed households) are actually Compact-wide indicators that may improve outside of the WSH activity, and it is unclear whether water and wastewater connections should be weighted equally. The output indicator: "Number of NAF households connected to the wastewater network as a result of the WSH activity" is misleading, since this indicator, measured quarterly, will really measure: "Number of NAF households with connections to the wastewater network financed by the WSH activity." It is also interesting to note that there are no within-household indicators of other infrastructure improvements.

Water network project

9. **Replacement of customer meters**: Is this indicator supposed to be installation of new meters as well as replacement of old ones, or only replacement? That should be clarified, since it is our understanding that some of the meters will be new where there are currently no meters.

10. **Construct new pumping station**: What is the parameter (%) being measured here? Is it financing dispersed or some other quantitative measure of construction progress?

11. **Restructure and construct District Meter Areas (DMAs)**: What is the parameter (number) being measured here? Is it when works in a DMA are complete or something else?

12. **Temporary employment indicators** (also in water network project): What is the parameter here? Number of jobs; man-hours, something else?

Wastewater network project

13. **Sewer blockage events**: How are these defined and measured? (Customer complaints or otherwise?). Do they have to result in overflows?

14. **Volume of wastewater collected**: This is a very important indicator, however we are not sure how it is measured. The M&E plan suggests this is the contribution pumped only from East and West Zarqa, and West Ruseifa. Are these volumes monitored at the pumping stations?

15. **Residential population connected to sewer system**: This indicator is mislabeled, as it should be: "% of network water subscribers connected to sewer system." Population requires an assumption about household size of those households connected; in addition some households may not have water connections.

As-Samra WWTP

16. **Treated wastewater used in agriculture**: This is a highly aggregated indicator, and its value and definition is unclear. First, the water in question is actually mixed water from the King Talal Dam (KTD). Second, this is a percentage of total irrigation water use in the Middle and Northern Jordan Valley (JV), and thus will vary considerably with hydrological variability. It would therefore perhaps be more useful to report as a quantity (volume) of treated wastewater used in agriculture, alongside a volume of water from the King Abdullah Canal (KAC) used in irrigation. Third, the indicator should be disaggregated to Northern and Middle JV areas, if possible.

17. **Quality of As-Samra effluent meets standard**: If monitoring is daily, the # of days not in compliance is fine; however, if several tests are conducted each day, it might be better to simply report the number of tests that passed and failed, respectively, by water quality indicator (BOD, turbidity, etc.).

18. **Volume of WWTP effluent discharged from As-Samra**: A time series for this indicator would be more useful than a single baseline value, since hydrological variability may influence water consumption and therefore wastewater production and treatment. Also note that year 1 levels are well above the reported baseline.

19. **Agriculture use of treated wastewater** (area): This indicator measures the cumulative annual amount of land irrigated with at least some treated wastewater. There are several problems with such a measure. First, different areas may be irrigated in different seasons, and the cumulative measure does not reflect this. Second, as with indicator 15 above (treated wastewater used in agriculture), it is the volume and not the land area, that is most relevant, and the treated wastewater referenced is in fact KTD water, not just treated wastewater. If expansion of area accompanies less water, that would show up as a success using this indicator. Finally, the cumulative nature of the indicator does not capture seasonal variations in the volume of treated wastewater supplied.

20. **"The actual substitution calculation"**: It is unclear what is being measured and how it is being measured. This would not appear to be an IE indicator.

21. **Expansion of As-Samra treatment plant**: What is the parameter being measured?

Annex B: Utility Performance Indicators

Version 1: Detailed

Basic Information   Utility________________Year_ _________

Basic Information:	Q1	Q2	Q3	Q4
Zone or Location				
Report prepared by				
Date prepared				
Estimated Population in Service Territory				
Estimated Population Served - Water				
Estimated Population Served - Sewer				
Active Customers (Accounts)				
Number of Water Meters				
Number of Connections to Lines				
Km of Secondary and Tertiary Lines, km				
New Water lines added, km				
New Sewer lines added,km				
Connections Added				
Connections disconnected				
Number of Bills Issued				
Water Volume of Billing System Bills				
Total Value of Bills				
GNI Income				

Water Balance **Utility______________ Year __________**

Water Balance	**Q1**	**Q2**	**Q3**	**Q4**
A. System Input				
Water Imports				
Own Sources				
Total System Input				
1. Billed Metered Consumption				
Billed by Billing System				
Tanker Sales				
Water Exports				
Total				
2. Billed Unmetered Consumption				
3. Billed Authorized Consumption (1+2)				
B. Non- Revenue Water A-3				
4. Unbilled Metered Consumption				
5. Unbilled Unmetered Consumption				
6. Unbilled Authorized Consumption 4+5				
C. Water Losses				
Physical Losses				
Administrative Loss				

Operational Indicators Utility______________ Year __________

Operational Indicators	**Q1**	**Q2**	**Q3**	**Q4**
Hours of water service / week				
Average water supply pressure				
Pump breakdowns				
Average pump repair time				
Number of pipe leaks / bursts				
Average leak repair time				
Planned water pipe replacements, km				
Number of sewer blockages				
Average repair time				
Planned Sewer lines replaced, km				
Meters repaired / replaced				
Energy consumption for pumping				
% Water quality tests passed at plants				
% Water quality test passed at houses				
Number of Complaints - water quality				
Number of Complaints - water quantity				
Number of Complaints - sewer				
Number of Complaints - Billing / Payment				
Number of Complaints - Total				

Financial Statement **Utility__________________ Year _____________**

Financial Statement	**Total**	**Water**	**Sewer**	**Water tankers**	**Desert wells**
Revenues					
Operational revenues					
Non operational revenues					
Other Revenues (Subsidy, etc)					
Total revenues					
Expenditures					
Wages and salaries					
Electricity					
Water imports					
Wastewater treatment plant					
Vehicle maintenance					
Network maintenance					
A&G expenses					
Other expenses					
Total expenditures					
Operating Income					
Cost Recovery Ratio					
Total Asset Value					
Depreciation					
Taxes, if any					
Debt Service					
Net Income					

Summary Indicators: **Utility__________________** **Year _____________**

Context	Q1	Q2	Q3	Q4
Zone or Location, if applicable				
Number of Customers (Subscribers)				
Distribution Line length, Km				
Average System Pressure, m				
Water Variable Cost, JD / m3				
Average Tariff, JD/m3				
Consumption, m3/month/customer				
Service Quality				
Water Coverage (Served / Total)				
Sewer Coverage (Served / Total)				
Hours of Service / Week				
Water Quantity Complaints/1000 water customers				
% of Water Quality tests passed				
Water Quality Complaints/1000 water customers				
Sewer Complaints / 1000 sewer customers				
Operations and Maintenance				
Physical Losses, L/ Customer/day				
Physical Losses, m3/km/day				
Bursts + Leaks/km/yr				
Average Leak Repair time				
Planned water pipe replacement %/yr				
Commercial Losses, L/Customer/Day				
Meter Replacement Rate, %/yr				
Pump Breakdown Rate				
Average pump repair time				
Energy Consumption kwhr/m3/m				
Sewer Blockage Rate				
Average Blockage Repair time				
Planned sewer pipe replacement %				
Finance				
Affordability , Average Water Bill / GNI				
Water Variable Costs / Average Tariff				
Sewer Cost / Average Tariff				
Billing Efficiency, %				
Collection Efficiency, %				
Operating Cost Recovery Ratio (OCCR), %				
Reinvestment expense / Asset Value				

Version 2: Abbreviated

Basic Information Utility________________Year_ _________

Basic Information:	Q1	Q2	Q3	Q4
Zone or Location				
Report prepared by				
Date prepared				
Estimated Population in Service Territory				
Estimated Population Served - Water				
Estimated Population Served - Sewer				
Active Customers (Accounts)				
Number of Water Meters				
Number of Connections to Lines				
Km of Secondary and Tertiary Lines, km				
Number of Bills Issued				
Water Volume of Billing System Bills				
Total Value of Bills				

Water Balance **Utility______________** **Year __________**

Water Balance	**Q1**	**Q2**	**Q3**	**Q4**
A. System Input				
Water Imports				
Own Sources				
Total System Input				
1. Billed Metered Consumption				
Billed by Billing System				
Tanker Sales				
Water Exports				
Total				
2. Billed Unmetered Consumption				
3. Billed Authorized Consumption (1+2)				
B. Non- Revenue Water A-3				
4. Unbilled Metered Consumption				
5. Unbilled Unmetered Consumption				
6. Unbilled Authorized Consumption 4+5				
C. Water Losses				
Physical Losses				
Administrative Loss				

Operational Indicators Utility______________ Year __________

Operational Indicators	Q1	Q2	Q3	Q4
Hours of water service / week				
Average water supply pressure				
Number of pipe leaks / bursts				
Number of sewer blockages				
Meters repaired / replaced				
Energy consumption for pumping				
% Water quality tests passed at plants				
% Water quality test passed at houses				
Number of Complaints - water				
Number of Complaints - sewer				
Number of Complaints - Billing / Payment				
Number of Complaints - Total				

<u>Financial Statement</u> **Utility__________________ Year _____________**

Financial Statement	Total	Water	Sewer	Water tankers	Desert wells
<u>Revenues</u>					
Operational revenues					
Non operational revenues					
Other Revenues (Subsidy, etc)					
Total revenues					
<u>Expenditures</u>					
Wages and salaries					
Electricity					
Water imports					
Wastewater treatment plant					
Vehicle maintenance					
Network maintenance					
A&G expenses					
Other expenses					
Total expenditures					
Operating Income					
Cost Recovery Ratio					

Summary Indicators: **Utility__________________** **Year _____________**

Context	Q1	Q2	Q3	Q4
Zone or Location, if applicable				
Number of Customers (Subscribers)				
Distribution Line length, Km				
Average System Pressure, m				
Water Variable Cost, JD / m3				
Average Tariff, JD/m3				
Consumption, m3/month/customer				
Service Quality				
Water Coverage (Served / Total)				
Sewer Coverage (Served / Total)				
Hours of Service / Week				
Water Complaints/1000 water customers				
% of Water Quality tests passed				
Sewer Complaints / 1000 sewer customers				
Operations and Maintenance				
Physical Losses, L/ Customer/day				
Physical Losses, m3/km/day				
Bursts + Leaks/km/yr				
Commercial Losses, L/Customer/Day				
Meter Replacement Rate, %/yr				
Energy Consumption kwhr/m3/m				
Sewer Blockage Rate				
Finance				
Affordability , Average Water Bill / GNI				
Water Variable Costs / Average Tariff				
Sewer Cost / Average Tariff				
Billing Efficiency, %				
Collection Efficiency, %				
Operating Cost Recovery Ratio (OCCR), %				

Annex C: Sampling Considerations

Water / Wastewater Activity

Introduction

The SI team conducted an analysis of calculations for the minimum sample required to produce reliable statistical estimates for MCA-Jordan's water sector program IE. We assume a clustered, quasi-randomized evaluation design estimated for a discrete treatment whose magnitude varies by cluster. We also assume that data collection will occur before and after implementation of the Compact activities. The purpose of the sample size estimates is to determine the minimum impact, Δ, that can be detected for a given number of clusters in the sample, *g*, and households in each cluster, *m*, for the evaluation sample. If the impact of the treatment is at least as large as Δ, we will be able to detect it 80 percent of the time in a sample of total size *N (=mg)*. If the treatment impact is less than Δ, we are less likely to detect it, although it is still possible.

The sample size calculations below are made for two scenarios. In the first scenario, the team will be able to use some socio-economic characteristics (SES) to identify comparison areas that are as similar as possible to the treatment areas *before* the intervention. Such SES data could be provided by prior household surveys in Zarqa conducted by the Department of Statistics (DoS) such as the 2004 Census or the 2010 Household Expenditure and Income survey. Concretely, receiving this data pre-baseline will allow the evaluation team to identify areas in Zarqa in which we have confidence that they will include a high number of observations that constitute good matches for the treatment observations. In the absence of this information, the evaluation team proposes to oversample by 50% in order to achieve the necessary number of matching units in the comparison group to conduct PSM, thereby compensating for the inability to conduct *ex ante* matching.

The optimal sample size depends, among other things, on the focal outcome variable. In determining the sample size we used several alternatives: water supply in both summer and winter; household water consumption according to quarterly bill; amount of quarterly water bill; household spending on treatment shop water; quantity of water purchased in water treatment shops; monthly expenditures on treating public network water; household reported health expenditures due to water; and monthly expenditures on tanker water. These indicators were chosen because they coincide with the project logic and were available from the 2009 Water Survey conducted by the DoS.

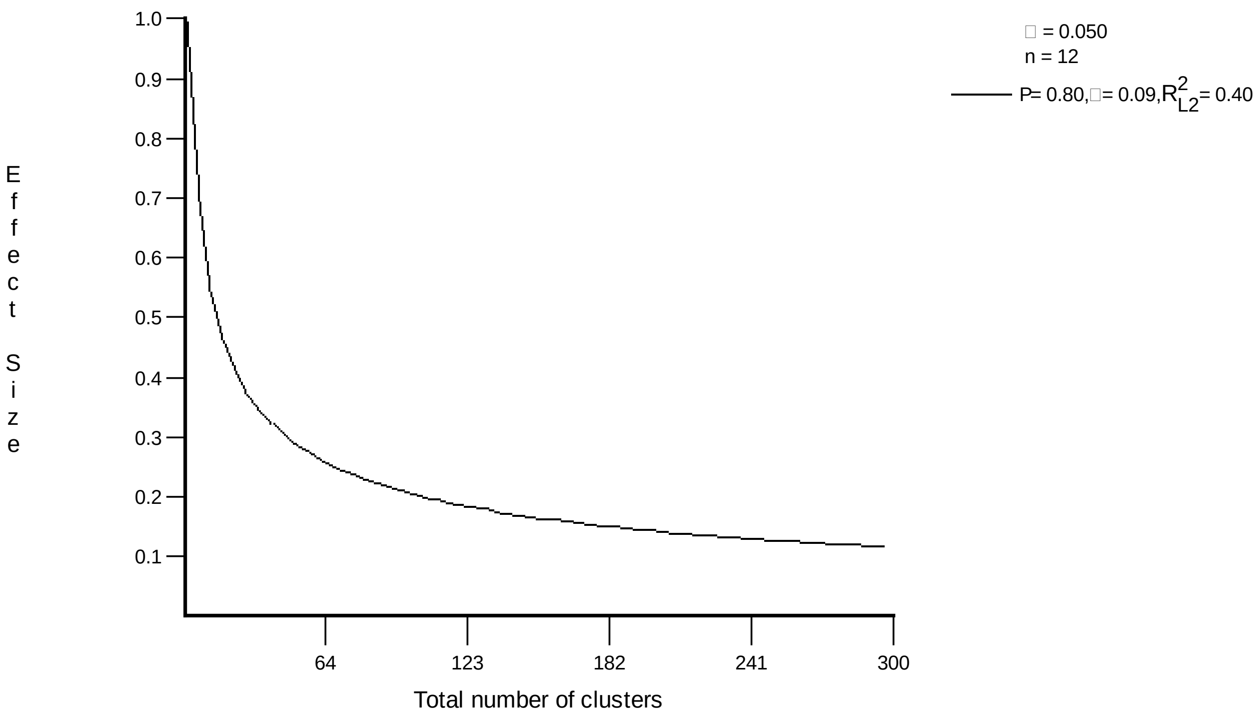
For the purpose of this sample size calculation, we are assuming that the treatment is discrete. As noted above, we plan on using difference-in-difference estimators with repeated observations on individual households. We assume Type I and II error rates of 5% and 20% respectively. We also assume that 90% of the variation in follow-up outcomes will be correlated with baseline outcomes. Finally, we use the 2009 Water Survey to compute intra-cluster correlation (ICC) for each relevant outcome variable. We define ICC as

(1) $\mathrm{ICC} = \frac{\tau^2}{\tau^2+\sigma^2}$,

or the proportion of overall variance explained by between group variance[15]. As can be seen in column (c) of Table 1 ICC varies quite significantly between the different outcome variables.

It is important to note that for all sample size estimations there is a tradeoff between decreasing the minimal detectable effect size and costs in terms of sample size. Figure 1 depicts this relationship for water use in winter, with the y-axis representing change in water use in standard deviations and the x-axis representing number of clusters needed with m=12. It is important to note the non-linear relationship between adding additional observations (here presented as additional clusters) and the decrease in the minimum detectable effect size.

Figure 1. Minimum Detectable Effect Size vs. Number of Clusters, Water Supply (Winter)



Results

Table 1 presents two scenarios to estimate the sample size needed to detect an effect of a given size. Panels (a) and (b) of Table 1 present the sample sizes needed to detect a 10% change on the respective outcome indicator between the treatment and the comparison group. Panels (c) and (d) of Table 1 present the sample sizes needed to detect a 5% change on these indicators. In each panel, we are separately estimating sample sizes for *m*=8, following the convention of 8 observations per Primary Sampling Unit (PMU) in the 2009 Water Survey, and for m = 12.

[15] The ICC can be estimated using a linear treatment model: Y_ij=α+BT+ v_j+ω_ij, where j indexes the cluster and i indexes the individual or household, Y is the outcome and T is the treatment. We assume that clusters are of identical sizes and v_j is IID with variance τ^2, and ω_ij is also iid, with variance σ^2. See Duflo, Glennerster and Kremer (2008), particularly 3921-3923.

These results show that a sample size of 3,440 would provide sufficient power to detect statistically significant changes of 10% magnitude in six important outcomes: water supply, water consumption, water bill, spending on treatment shop water, quantity of water purchased in treatment shops, and monthly expenditures on water. Differences in treating water from the public network as well as self-reported health expenditures because of water consumption would be harder to detect with a sample size of 3,440, especially for the latter. Reasonable power to detect such differences would require sample sizes beyond the capability of this evaluation. The same is true for changes of a magnitude of around 5%, shown in panels (c) and (d). While we may be able to detect some the changes, it is likely that we may not achieve the power to detect statistically significant effects.

A sample size of 2,500 would be sufficient for detecting 10% differences across treatment and control groups if the sample is comprised of 4 groups (roughly 625 households per group). However, in consideration of the likelihood of Zarqa-wide spillovers, the optimal final sample includes one additional group from peri-urban zones in Amman (roughly 625 households). Accounting for 10% attrition over the life of the evaluation, 313 households are added to the final sample, such that about **3,440 randomly-selected households** are required.

Different Activities

It is important to note from Section B.1 that we expect the benefits of the water and wastewater interventions to occur as depicted in Table 1. The sampling consideration outlined in the preceding paragraph is for *one* control/treatment group pair, for example comparing Group A with Group B, or Group C with Group D. Since the outcome indicators we look at are similar for both interventions, we expect sample sizes to be of approximately the same size for each treatment/comparison pair. If effects were to be estimated for Groups A, B and C separately, we would need three independent samples of the size calculated above in order to detect the impacts of the different treatment components. Moreover, since we expect the determinants of participation, which we will estimate using equation (1) above, to vary between the different activities, we may also require three comparison groups of similar size for each activity.

Exposure to water and wastewater network improvements

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

Table 1: Sample size estimates, water/wastewater intervention

(a) Δ = 10%; m=8

Outcome Variable	Mean	SD	ICC	g	n	Attrition	Loss without SES	N
				m=8				
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Water Supply Hours/Day (Winter)	17.2	6.8	0.41	54	426	63.9	213	702.9
Water Supply Hours/Day (Summer)	16	7.16	0.423	68	544	81.6	272	897.6
Water Consumption according to quarterly bill (cubic meters)	48.97	36.1	0.087	75	594	89.1	297	980.1
Quarterly Water Bill (JD)	16	15.9	0.049	116	922	138.3	461	1521.3
Spending on Treatment Shop Water (JD)	7.71	5.73	0.111	86	686	102.9	343	1131.9
Quantity Water purchased in Treatment Shops	180.9	125.5	0.088	69	548	82.2	274	904.2
Monthly Expenditure on Treating Water from Public Network (JD)	4.72	5.85	0.271	522	4172	625.8	2086	6883.8
Water related health expenditures in past 12 months (JD)	58	121.3	0.147	771	6162	924.3	3081	10167.3
Monthly expenditures on Tanker Water (JD)	30.48	25.06	0	59	470	70.5	235	775.5

(b) Δ = 10%; m=12

Outcome Variable	Mean	SD	ICC	m=12 g	n	Attrition	Loss without SES	N
	( )			( )			( )	
	(a)	(b)	(c)	(i)	(j)	(k)	(l)	(m)
Water Supply Hours/Day (Winter)	17.2	6.8	0.41	51	608	91.2	304	1003.2
Water Supply Hours/Day (Summer)	16	7.16	0.423	65	776	116.4	388	1280.4
Water Consumption according to quarterly bill (cubic meters)	48.97	36.1	0.087	60	716	107.4	358	1181.4
Quarterly Water Bill (JD)	16	15.9	0.049	88	1056	158.4	528	1742.4
Spending on Treatment Shop Water (JD)	7.71	5.73	0.111	72	1712	256.8	856	2824.8
Quantity Water purchased in Treatment Shops	180.9	125.5	0.088	55	658	98.7	329	1085.7
Monthly Expenditure on Treating Water from Public Network (JD)	4.72	5.85	0.271	478	5734	860.1	2867	9461.1
Water related health expenditures in past 12 months (JD)	58	121.3	0.147	663	7946	1191.9	3973	13110.9
Monthly expenditures on Tanker Water (JD)	30.48	25.06	0	40	470	70.5	235	775.5

(c) Δ = 5%; m=8

Outcome Variable	Mean	SD	ICC	g	n	Attrition	Loss without SES	N
				m=8				
	(-)			( )			( )	
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Water Supply Hours/Day (Winter)	17.2	6.8	0.41	210	1680	252	840	2772
Water Supply Hours/Day (Summer)	16	7.16	0.423	269	2152	322.8	1076	3550.8
Water Consumption according to quarterly bill (cubic meters)	48.97	36.1	0.087	297	2372	355.8	1186	3913.8
Quarterly Water Bill (JD)	16	15.9	0.049	461	3684	552.6	1842	6078.6
Spending on Treatment Shop Water (JD)	7.71	5.73	0.111	351	2806	420.9	1403	4629.9
Quantity Water purchased in Treatment Shops	180.9	125.5	0.088	273	2182	327.3	1091	3600.3
Monthly Expenditure on Treating Water from Public Network (JD)	4.72	5.85	0.271	1514	12110	1816.5	6055	19981.5
Water related health expenditures in past 12 months (JD)	58	121.3	0.147	3079	24630	3694.5	12315	40639.5
Monthly expenditures on Tanker Water (JD)	30.48	25.06	0	236	1888	283.2	944	3115.2

(d) Δ = 5%; m=12

Δ=5%

Outcome Variable	Mean	SD	ICC	g	n	Attrition	Loss without SES	N
				m=12				
	-----------( )-----------			-----------( )-----------			-----------( )-----------	
	(a)	(b)	(c)	(i)	(j)	(k)	(l)	(m)
Water Supply Hours/Day (Winter)	17.2	6.8	0.41	200	2392	358.8	1196	3946.8
Water Supply Hours/Day (Summer)	16	7.16	0.423	256	3070	460.5	1535	5065.5
Water Consumption according to quarterly bill (cubic meters)	48.97	36.1	0.087	239	2858	428.7	1429	4715.7
Quarterly Water Bill (JD)	16	15.9	0.049	352	4220	633	2110	6963
Spending on Treatment Shop Water (JD)	7.71	5.73	0.111	293	3506	525.9	1753	5784.9
Quantity Water purchased in Treatment Shops	180.9	125.5	0.088	222	2658	398.7	1329	4385.7
Monthly Expenditure on Treating Water from Public Network (JD)	4.72	5.85	0.271	1387	16642	2496.3	8321	27459.3
Water related health expenditures in past 12 months (JD)	58	121.3	0.147	2648	31766	4764.9	15883	52413.9
Monthly expenditures on Tanker Water (JD)	30.48	25.06	0	315	3774	566.1	1887	6227.1

Table 2: Sample size estimates, Water Smart Homes Activity

Δ=10%

Outcome Variable	Mean	SD	n	Attrition	N
	------------------------( )------------------------				
	(a)	(b)	(j)	(k)	(m)
Spending on Treatment Shop Water (JD)	7.71	5.73	770	115.5	885.5
Quantity Water purchased in Treatment Shops	180.9	125.5	676	101.4	777.4
Monthly Expenditure on Treating Water from Public Network (JD)	4.72	5.85	2878	431.7	3309.7
Water related health expenditures in past 12 months (JD)	58	121.3	6070	910.5	6980.5
Monthly expenditures on Tanker Water (JD)	30.48	25.06	940	141	1081

Annex D: New Method for Determining Water Loss Components in Water Distribution Systems

<u>Initial Test Application in Zarqa, Jordan</u>

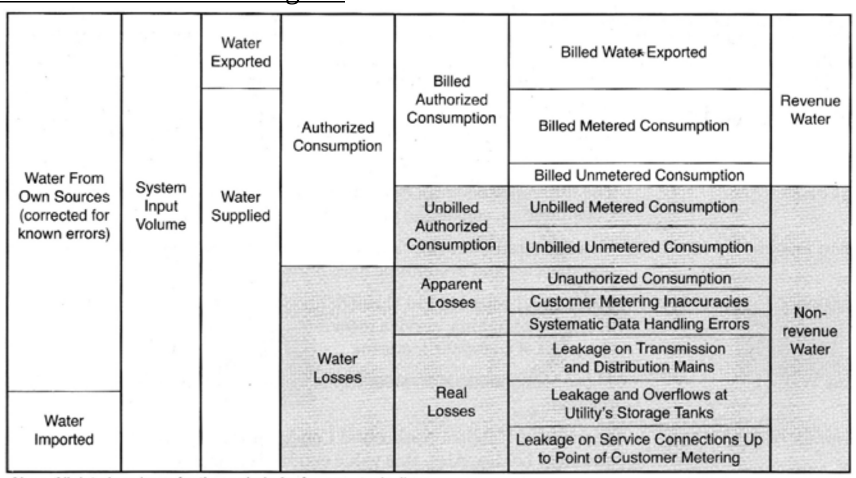
1. Introduction

One of the key challenges in assessing non-revenue water and water losses, and in planning programs that aim to reduce such elements, is a lack of information on the amounts of the different types of losses. Without knowledge of the magnitude of real losses or apparent losses, the most effective interventions cannot be determined, nor can their cost effectiveness be assessed.

Non-revenue water can be found from the system input volume less the billed authorized consumption. If unbilled authorized consumption is also subtracted out, water losses can be found. The basis of these terms and relationships are described in the IWA Water Balance shown below. It is not uncommon to have reasonably accurate estimates of the system input volume (bulk meter readings), and the billed authorized consumption (billing system records). Unbilled authorized consumption is often small and can typically be estimated roughly. Given these data, the total water losses can be found, but not any of its many components.

A common method to estimate the components of water losses is a top down analysis, where commercial losses are estimated from "rules of thumb", and the remaining losses deemed to be physical losses. This method is very rough, and assuming parameter values from developed countries tends to be inappropriate for most developing country situations where meters are old, or illegal connections are common. In such cases, the top down method becomes guesswork.

<u>IWA Water Balance Diagram</u>



NOTE: All data in volume for the period of reference, typically one year.

Note: If all exports are billed, all exports can be removed from the water balance, and associated mathematics, without affecting most indicators. If NRW is expressed as a percentage of system input volume, which is not recommended (but common), the percentage NRW will be affected.

The other method is a bottom-up analysis, which is considerably more laborious, but also more accurate. A bottom-up analysis requires water network engineering tests of minimum night flow, in combination with real loss component analysis, which is data-intensive and therefore costly.

Therefore both methods are problematic or subject to significant uncertainty, in situations where new projects are being planned and actual network performance data is limited.

Once installed, well designed, fully operational networks with pressure measurement and “smart metering” on all inlets and outlets to DMAs would provide sufficient information to develop zone-specific water balances, and to detect leaks and illegal connections. Unfortunately, at the start of a project, such detailed data and monitoring are usually not available.

This technical note describes another possible method for estimating the amount apparent losses and the components of real losses. Additional analyses can give estimates of the components of apparent losses. The method requires data on the system input volume and authorized billed consumption for different zones of a city network. The method has been applied the case of Zarqa, Jordan, with interesting and plausible results.

2. Water Supply Network Performance in Jordan

There is unfortunately, little reliable information on the magnitude of the components of the Water Balance in Jordanian cities. Some limited empirical information is presented below. Before reviewing these field data it is worth noting that most analyses assume that apparent losses and real losses are about equal. In fact the PMU uses 50% / 50% as a default value,

During the period 2008 – 2009, studies were conducted in 9 of the 330 districts in Amman, under a project supported by USAID. Three different private companies each conducted surveys and studies of three districts and made repairs of leaks discovered. No other corrective actions such as meter replacement, pressure management or others were undertaken. At the end of the study period the 9 districts were found to have a percentage NRW of 21% of the district supply, which was considerable lower that the Amman average of 33% at the same time. The text box below shows specific results, which show the component breakdown of the losses.

Amman	2009		Source: Patrick, et al, 2011		
1)	**9 Pilot districts out of 330**		**Results after surveys and leak repairs (NO METER REPLACEMENT)**		
	LOSSES AS A PERCENTAGE OF DISTRICT INPUT				
			Range	Average	
	Apparent Losses	Meter Error	6%	6%	Fairly consistent across districts
		Database error	1%	1%	Fairly consistent across districts
		Illegal Use	2-10%	6%	Not consistent across districts
	Real Losses	Leakage	5-10%	8%	Not consistent across districts
	Total Losses		14% - 27%	21%	
	Apparent Losses		9% - 17%	13%	
	Real Losses		5% - 10%	8%	
	Total Losses		14% - 27%	21%	
2)	**All 330 Districts**		**Losses =**	**33%**	
	ESTIMATED EXTRAPOLATION of LOSSES AS PERCENTAGE OF SYSTEM INPUT				
	Apparent Losses			15%	
	Real Losses			18%	
	Total Losses			33%	

In the 9 districts where real losses were reduced, they were only about a third of the losses, with meter error and illegal use remaining quite significant. A very rough extrapolation can be made to all the Amman districts, which results in a roughly even split between real losses and apparent losses.

During the period 2006 – 2008, JICA financed a project to conduct evaluations of DMAs in 11 locations in various Governorates in Jordan including Balqa, Zarqa, Madaba, Karak, Tafilah, and Ma'an. These efforts included DMA establishment, network surveys, partial mains replacement, connection and service line replacements, meter replacements / re-installation repair (those over 5 years old or improperly installed), and regularization of illegal connections. The project attempted minimum night flow tests, but difficulties in obtaining reliable indications of real losses due the intermittent flow conditions. Nonetheless, the NRW in the DMAs was reduced from a range of 40% - 60% to about 20% - 30% of input, at a modest cost. The publically available documents on this work do not provide sufficient information to determine exact water balances before and after the rehabilitation work.

However, more complete results from one DMA in Ma'an were analyzed more closely (Sukkar, et al 2009). The results are shown below, including basic information on the DMA, rehabilitation activities undertaken, overall program results and an estimated water balance.

Al Salalim DMA, Balqa Governorate — Analysis by the author, based on data in Sukkar, et 2009

Program to survey lines, repair leaks, replace mains and service lines, regularize illegal connections, realign and replace meters

DMA PARAMETERS					
DMA area	3.3	km2	Measured Consumption	5990	m3/week
Elevation Range	1068m - 850	m	Average Measur. Consumption/Customer	421	L/Cust/day
Population	11,950		Days of Storage Per Customer	8.6	days
Connections	1025		Average Pressure	55	m
Customers	2034		Length of Mains	40.7	km
Total Customer Storage Vol	7335	m3	Density of Connections	25.2	Conns / km of mains
Avg Customer Storage Vol	3.61	m3	Average Length of Service line	7	m per customer
Supply Schedule, days/week	2.5	Fri pm to Mon am	Pipe Material	DI for mains, PE for service lines	

ACTIVITIES CONDUCTED	Feb 2007 to Jan 2008					
Main Line surveyed, km	34.8	86%	of total mains length			
Leaks Repaired, #	106	3.0	leaks repaired per km of mains surveyed			
Length of Mains replaced, m	230	0.7%	of km of mains surveyed			
Service Lines replaced, #	138	13%	of connections			
Illegal Connections Rectified	13	1.3%	of known connections			
Meters Replaced / Repaired	580	29%	of customers	57%	of connections	
Approx night flow measurement at start of project	19	m3/hr =	1,140	m3/week		Project team report some uncertainty about the accuracy
Approx night flow measurement at end of project	14	m3/hr =	840	m3/week		of these tests, due to intermittent flow conditions

Program Results		Initial	Final	Change			Note
System Input	m3/week	10,850	12,480	1,630	15%	up	Increased consumption is from reduced commercial loss
Consumption	m3/week	5,990	10,003	4,013	67%	up	and perhaps real consumption increase due to more
NRW	m3/week	4,860	2,477	-2,383	-49%	down	readily available network water ??

Estimated Bottom Up Water Balance, before and after program

	Initial			Final			NOTES	
	m3/week	% of input	L/Conn/Day	m3/week	% of input	L/Conn/Day		
Consumption	5,990			5,990				
Converted Commercial Loss	0			1,637			27%	Increase over previous measured consumption
New consumption	0			2,376			31%	Increase over previous full consumption
Total observed consumption	5,990	55%	835	10,003			67%	Increase in revenue
Apparent Loss	3,720	34%	518	1,637	13%	225	57%	decrease due to meter program
Real Loss	1,140	11%	159	840	7%	116	27%	decrease due to rehabilitation program
Total Water	10,850	100%	677	12,480	20%	341	50%	decrease due overall program

Several observations can be made, including:

- The rehabilitation program was not exhaustive –less than 1% of the mains were replaced, only 13% of the service connections were replaced, and 29% of customer meters replaced. It could be that the network was in pretty good condition, but that seems unlikely at it had an NRW of 48% of input (or 677 L/Connection/Day, which is in Class C on the IWA real loss scale which ranks performance on a scale from A (very good) to D very poor)). The more likely explanation is that only the major water loss issues were addressed.
- Based on the rough night flow estimates, real losses were found to be much smaller than apparent losses in the situation before and after the rehabilitation program.
- There was a net increase in water consumption which cannot be fully explained based on regularization of illegal connections and improved metering. This increase may have resulted from changes in user perceptions of water quality from the rehabilitation.

The Aqaba Water Company has been working on studies to determine the amount of apparent losses and real losses in their network for several years. They have been assuming a 50% / 50% split between the two types of losses. In several weeks the results of their work will be released. Early indications are that the apparent losses may be even higher than real losses.

3. MCC Project in Zarqa

The readers of this technical note do not need an extensive introduction to the Zarqa water network project being financed by MCC, with some parallel support from GOJ and other bi-lateral donors. The project is designed to rehabilitate and restructure the network, to:

a) Decrease direct pumping of well water resources directly into the network, and increase the number and size of elevated storage tanks, to decrease average network pressures, reduce NRW, and create more efficient operations

b) Create Zones, Distribution Areas, and District Metered Areas, with appropriate metering and pressure management to facilitate efficient operations

The expected project impacts include increased use of piped network water, providing a) health benefits (especially for users who consume little water currently), b) household financial savings due to less use of tankers and water shops, and c) reduction in the use of expensive imported water such as supply from Disi.

3.1 Water Balance Estimates for Zarqa from Consultant Report 2008

The table below shows the initial information needed to compute a water balance for Zarqa for 2008. Note that this report uses the term UFW, instead of Water Losses. The total water losses were estimated at 71,985 m3/day, and total NRW at 74,947 m3/day, values close to 60% of system input.

IWA Water Balance Categories	Water Volume	
	Total m^3/annum	Av. Day m^3/d
System Input - Production	50,408,378	137,728
System Input - Imports	3,677,248	10,047
Export - Azraq Conservation Area	727,207	1,987
Export - to Amman from Khaw	7,047,988	19,257
Export - to North - Mafraq from Qunaya	124,350	340
Export - to West - Jerash and Balqa from Um Rumanneh	272,027	743
Export - to Desert	146,035	399
System Input Total	**45,768,019**	**125,049**
Domestic Billed Metered Consumption	17,437,114	47,642
Non-Domestic Billed Metered Consumption	900,440	2,460
Billed Metered Consumption Total	**18,337,555**	**50,103**
Unbilled Unmetered Consumption - Zarqa Backwash	931,513	2,545
Unbilled Metered Consumption - Agriculture	133,333	364
Unbilled Metered Consumption - Tankers	18,969	52
Unbilled Consumption Total	**1,083,815**	**2,961**
Unaccounted-for Water: UFW	**26,346,649**	**71,985**
UFW %	57.6%	
Non-Revenue Water: NRW = UFW + Unbilled Cons.	**27,430,464**	**74,947**
NRW %	59.9%	

Table 2-27: Zarqa Governorate Imports, Exports and Water Production.

The Consultant also prepared a top-down analysis to estimate water loss components, shown below. The first result, based on a series of assumptions, is that the apparent losses total only 8,923 m3/day. As per the top-down method, the real losses can be computed from the difference between the total losses 71,985 m3/day and those apparent losses yielding a result of 63,062 m3/day. In comparison with the results from Amman and other cities in Jordan, the ratio of apparent losses to real losses in Zarqa seems far too low. A re-examination of the water balance calculations is in order.

The information / assumptions used to estimate the components of apparent losses need a second look.

a. Unknown / unmetered connections. The figure of 10% of known and metered consumption is a pure estimate. For the case of Amman illegal use was estimated at 6% of system input which translates to about 8% of known consumption. General IWA guidance suggests an estimate of 8% (Seago, et al 2005)

Top down water balance analysis from NOD report for 2008

Parameter	Unit	Value	At Current Supply Continuity (37 hrs/wk)
Total UFW - Average Day	m3/d		71,985
Total UFW	% of SI		57.6%
System Input Total	m3/d		125,049
Supply Continuity (water pressure in pipes for avg 4 hrs longer duration than water-on duration)	system hrs/wk	=37+4	43
Administrative (Apparent) Losses			
Unknown/unmetered connections	% of known&metered	10%	5,366 m3/d
Meters stopped connections	% of known&metered	2%	1,052 m3/d
Meters avg. under-registration for working customer meters	% underregistration av	5%	2,505 m3/d
Systematic data handling errors	% underrecording av	0%	-
Total Administrative Losses	m3/d		8,923 m3/d
Total Administrative Losses	% of SI		7.1%
Physical (Real) Losses			
Total Physical Losses	m3/d		63,062
Total Physical Losses	% of SI		50.4%

b. Connections with stopped meters. The 2% figure is based on a survey of 100 connections in one section of the city. It is unclear how reliable this estimate is.

c. Meter under-registration. The 5% figure was selected under the assumption that high pressure conditions would mean that meters were operating at high flow, which is their most accurate operating condition. However, the Consultant reports that there are many areas in the city where pressures are low, and flow rates through meters will be very low leading to high accuracy. The report indicates that many of the meters were installed in 1992, and have been operating under intermittent conditions for over 20 years. Several studies have indicated that a 1% drop in accuracy per year could be expected under such conditions. (Male et al, 1985)

d. Data handling errors. The estimate of 0% is highly unlikely in a Jordanian city. Amman data indicates 8%. The analysts for Amman indicate that they suspect illegal use in other districts is much higher than in the 9 districts that were studied. Reports indicate that the billing system at Zarqa has experienced some difficulties. General IWA guidance suggests an estimate of 10% (Seago, et al 2005)

Using new estimates of the components of apparent loss, the water balance would be as shown below. The total apparent losses come to 21,279 m3/day (or 17% of input and 30% of losses). The resulting real losses come to 50,706 m3/day (40.5% of input and 70% of losses) While these estimates may be more plausible they are not based on strong empirical evidence. The amount of apparent and real losses both before and after the project have important implications for the project economics. Improved water balances can be determined from a combination of night flow tests and DMA inflow / outflow analysis, once the rehabilitations are complete,

But a better baseline value would be useful for project impact assessment and for project planning. In addition, a method to estimate a water balance in non-intervention areas would be useful to assess project impact. The next section of this paper illustrates such a new method.

Revised Approximate Top Down Water Balance for Zarqa

		m3/day	% of input	% of Losses	L/ Conn /Day	IWA Class
System Input		125,049	100.0%			
Billed Authorized Consumption		50,103	40.1%			
Non-Revenue Water		74,946	59.9%			
Unbilled Authorized Consumption		2,961	2.4%			
Water Losses		71,985	57.6%			
Apparent Losses						
Illegal Connections	8%	4,298	3.4%	6.0%		
Meter error	22%	11,674	9.3%	16.2%		
Data Handling error	10%	5,306	4.2%	7.4%		
Total Apparent Losses		21,279	17.0%	29.6%	233	D
Real Losses		50,706	40.5%	70.4%	555	C
Total Losses		71,985	57.6%	100.0%	789	D

4. New Method for Water Loss Component Estimation

A new method has been developed to estimate apparent losses, and components of real losses, to:

a. re-examine the assumptions of the project design, and

b. formulate a basis for ongoing performance monitoring.

Additional calculations could be used to get rough estimates of the magnitude of the components of apparent losses.

The method, which is illustrated through a pilot application in Zarqa, consists of two analyses of water losses in different zones or distributions areas (DAs) in relation to connection density. By looking at the extreme limits of the relationship between the indicators and connection density, estimates of the components of the losses can be found. Note that the unbilled authorized consumption, which is part of NRW, but not water losses, is small and is ignored for simplicity.

4.1 Input data for Zarqa

The table below, from an Annex of the Investment Master Plan, provides basic data on the DAs and Zones for Zarqa, for 2008. The water losses in m^3/km of mains and in liters per connection per day can be found from these data. Note that the table lists customers, not connections, but connections for each Zone are listed elsewhere in the Master Plan.

Quantities 14 Zarqa WSS **Population, Average System Input, UFW and Infrastructure Statistics by Distribution Area 2008 and 2030**

WSA / Distribution Area		Population		AD System Input		UFW			Infrastructure 2008		Infrastructure 2030	
		2008	2030	2008	2030	2008	2030	2030	Pipe Length	Connections	Pipe Length	Connections
ID	NAME						Without project	With project				
				m3/d	m3/d	m3/d	m3/d	m3/d	m	no	m	no
Az1	Azraq	7,396	10,154	1,937	1,147	746	563	222	48,756	991	56,749	1,351
AzTot		**7,396**	**10,154**	**1,937**	**1,147**	**746**	**563**	**222**	**48,756**	**991**	**56,749**	**1,351**
Du1	Hallabat	3,213	4,704	414	532	216	261	103	17,440	281	21,050	402
Du2	Dulail	33,568	46,481	4,271	5,252	2,203	2,578	1,016	170,373	2,939	205,640	3,967
Du3	Tafeh	1,009	4,144	296	468	234	230	91	43,425	88	52,414	354
DuTot		**37,790**	**55,329**	**4,981**	**6,252**	**2,653**	**3,069**	**1,209**	**231,237**	**3,309**	**279,104**	**4,722**
Ka1	King Abdullah City High	-	276,589	105	31,255	105	15,341	6,046	23,910	-	384,889	27,659
Ka2	King Abdullah City Low	-	159,419	174	18,015	174	8,842	3,485	39,658	-	638,380	15,942
KaTot		**-**	**436,008**	**278**	**49,270**	**278**	**24,183**	**9,531**	**63,568**	**-**	**1,023,268**	**43,601**
Ta1	Tatweer	8,501	16,888	1,697	1,908	1,112	937	369	81,755	1,613	124,899	2,896
TaTot		**8,501**	**16,888**	**1,697**	**1,908**	**1,112**	**937**	**369**	**81,755**	**1,613**	**124,899**	**2,896**
Za1	Zarqa High	26,308	31,677	4,163	3,580	2,520	1,757	692	104,246	4,405	124,167	5,226
Za2	Zarqa Mid - Batrawi	253,728	313,932	38,991	35,475	23,146	17,412	6,863	740,444	42,485	881,943	51,788
Za3	Zarqa North	3,490	53,977	555	6,100	337	2,994	1,180	14,405	584	17,158	8,904
ZaTot		**283,526**	**399,586**	**43,708**	**45,154**	**26,003**	**22,163**	**8,735**	**859,095**	**47,474**	**1,023,268**	**65,918**
Aw1	Awajan High	61,196	76,882	9,291	8,688	5,301	4,264	1,681	238,006	8,978	299,180	10,896
Aw2	Awajan Low	37,024	47,069	5,666	5,319	3,252	2,611	1,029	154,168	5,432	193,793	6,671
Aw4	Awajan West	22,922	46,344	3,530	5,237	2,035	2,570	1,013	100,430	3,363	126,243	6,568
Aw5	Awajan North	28,028	55,386	4,323	6,259	2,496	3,072	1,211	124,475	4,112	156,469	7,850
AwTot		**149,171**	**225,680**	**22,809**	**25,502**	**13,084**	**12,517**	**4,933**	**617,079**	**21,885**	**775,685**	**31,985**
Ru1	Russaifah North West	4,704	47,676	616	5,387	355	2,644	1,042	20,420	557	25,258	5,480
Ru2	Russaifah High	96,064	129,853	11,961	14,674	6,619	7,202	2,839	274,294	11,372	339,288	14,926
Ru3	Russaifah Low	69,265	86,039	8,713	9,723	4,861	4,772	1,881	217,970	8,199	269,618	9,890
Ru4	Hitteen Tower	12,157	15,753	1,497	1,780	821	874	344	30,963	1,439	38,300	1,811
Ru5	Hitteen Reservoir	121,259	171,153	15,234	19,341	8,491	9,493	3,741	377,169	14,354	466,539	19,673
RuTot		**303,448**	**450,472**	**38,022**	**50,904**	**21,147**	**24,985**	**9,847**	**920,815**	**35,921**	**1,139,003**	**51,779**
No1	Khaw	1,356	5,752	219	650	126	319	126	7,922	192	9,725	791
No2	Hashmeya South	623	1,568	101	177	58	87	34	3,627	88	4,453	216
No3	Hashmeya Town	26,139	32,974	4,194	3,726	2,399	1,829	721	144,974	3,692	177,964	4,533
No4	Hashmeya Rural	3,471	6,942	562	784	323	385	152	20,367	490	25,002	954
No5	Sukhna	20,040	28,772	3,249	3,251	1,872	1,596	629	118,673	2,831	145,677	3,956
NoTot		**51,629**	**76,008**	**8,325**	**8,589**	**4,779**	**4,216**	**1,662**	**295,564**	**7,293**	**362,821**	**10,449**
We1	Merhib	1,454	2,258	381	255	256	125	49	29,426	266	37,452	397
We2	Berein	2,647	3,976	676	449	448	221	87	49,647	485	63,187	699
We3	Um Rumanneh	3,449	5,242	870	592	573	291	115	62,065	631	78,992	922
We4	Berein North	1,043	1,759	276	199	186	98	38	21,715	191	27,637	309
We5	Sarout/Alouk	3,304	4,978	822	562	538	276	109	57,038	605	72,594	875
We6	Qunaya	966	1,661	266	188	183	92	36	22,554	177	28,705	292
WeTot		**12,864**	**19,874**	**3,291**	**2,246**	**2,184**	**1,102**	**434**	**242,445**	**2,355**	**308,566**	**3,495**
TOTAL for WSAs		**854,325**	**1,690,000**	**125,049**	**190,973**	**71,985**	**93,736**	**36,944**	**3,360,316**	**120,841**	**5,093,364**	**216,196**

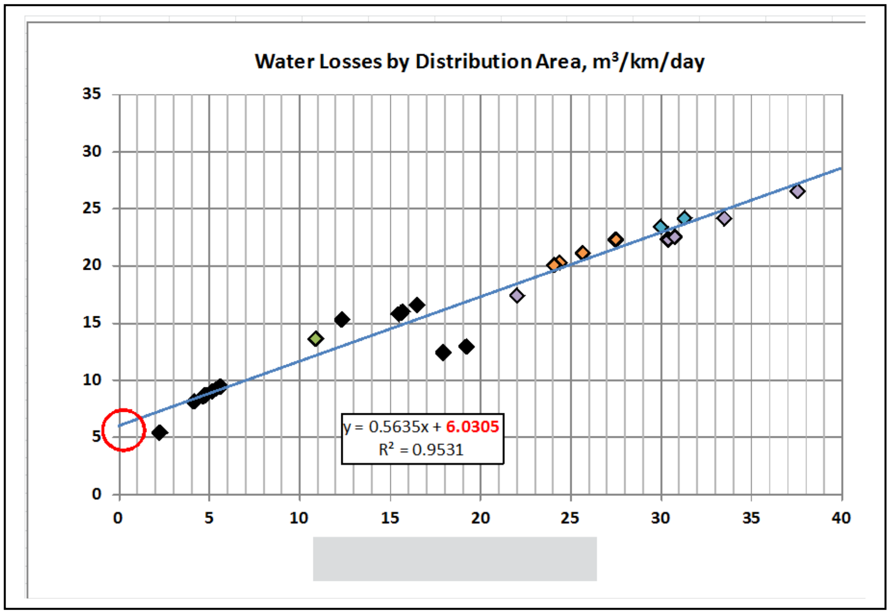
In addition to UFW effects, there are other significant level of service issues incorporated into the above statistics:

At 2008 => 37 hours per week average supply, 55.8 lcd average domestic use
With Project at 2030 => 70 hours per week average supply, 85 lcd average domestic use
Without Project at 2030 => 18 hours per week average supply, 53 lcd average domestic use

Alternative scenarios to the above can be generated, depending on constraints - quantity of water available, required minimum duration of supply, required minimum unit quantity of supply, etc.

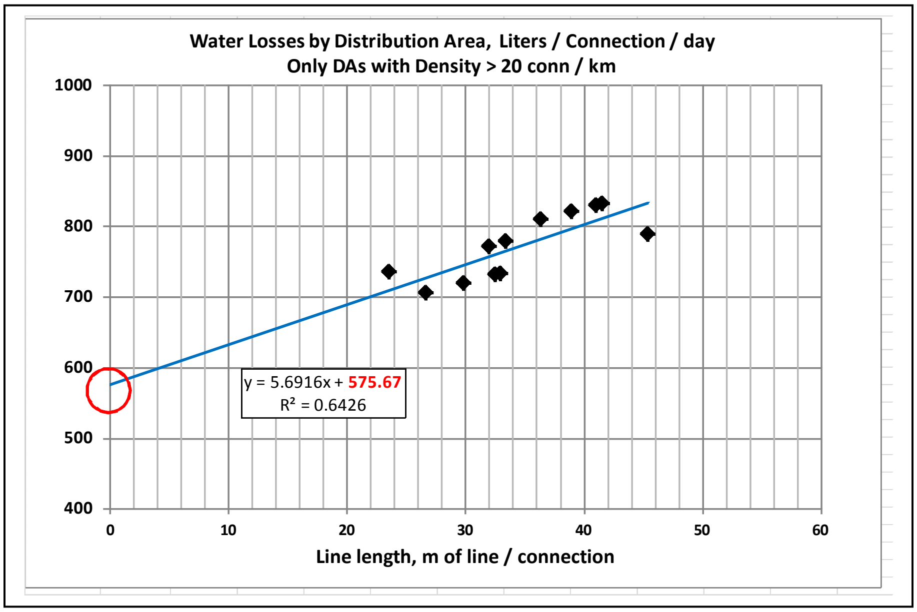
4.2 Limit Analysis

The graph below provides a plot of water losses in m³/km/day, for different DAs in Zarqa. The colored points are those in priority areas of the city, as determined by MCA and the engineering consultant.



The intercept in diagram above represents the losses if the connections per km went to zero, which would be like a situation like a zone with just pipeline and no other sources of water loss. It gives an estimate of the real line losses, that is the background losses and burst losses.

The second graph provides a plot of water losses in liters / connection / day, for only those DAs in Zarqa, whose connection density is above 20 connections / km, according to IWA protocol.



The intercept in diagram above represents the losses if the length of pipeline per connection went to zero, which would be like a zone with lots of connections and no pipeline. It gives an estimate of the apparent losses, real losses at the service connection junction, and real losses in the service line.

The values of these two intercepts facilitate a new water balance calculation, as illustrated in the Table below.

BASIC INFORMATION			
Total Losses		71,985	m3/day
Mains Length		3,360	km
Connections		91,290	
Estimate Pressure		40	m
Hours / day		5.9	
Avg Serv line/Conn		5.6	m
WATER BALANCE CALCULATIONS			
1. Analysis of Line Losses			
Pure Line Loss =	Intercept =	6.031	m3/km/day
Total Losses		21.422	m3/km/day
Line loss / Total		28%	
Total Mains Line Losses		20,262	m3/day
UARL Line Losses	w/UARL = 18 l/km/m	2,419	m3/day
Infrastructure Leakage Index (ILI) =		8.4	
ILI if water was flowing continuously		34.3	
2. Other Real Losses			
Tank Overflow	default estimate	0	m3/day
Serv Conn UARL	0.8 l/conn/day/m	2,921	
Service Connection Loss =		24,468	m3/day
Serv Line UARL	25 l/km/day/m	639	
Service Line Loss =		5,352	m3/day
3. Sum of Real Losses			
Total Mains Line Losses =		20,262	m3/day
Service Connection Loss =		24,468	m3/day
Service Line Loss =		5,352	m3/day
Total Real Loss =		50,083	m3/day
4. Resulting Amount of Apparent Losses		21,902	m3/day
5. Analysis of Connection Losses			
Total Connection Loss from Intercept		576	L/Conn/Day
Total Connection Loss		52,553	m3/day
Real Losses at Connections		29,820	m3/day
New Estimate of Apparent Losses		22,733	m3/day
Difference from Earlier Estimate		830	m3/day
% Difference from Earlier Estimate		3.8%	
This difference could really be in Apparent Losses			
Revised estimate for Real Losses		49,252	

The charts below show a first Bottom-Up Water Balance, for the whole system.

Water Balance - Bottom Up and Top Down					% of Cons	% of SIV	L/Conn/Day
System Input Volume 125,048 (Imports included, but exports removed)	Authorized Consumption 53,064	Billed Authorized Consumption, w/o exports	50,103			40.1%	549
		Unbilled Auth Cons	2,961			2.4%	32
	Water Losses 71,986	Apparent Losses 22,733	*Meter Error*	*10,021*	*20%*	18.2%	249
			Stopped Meters	*1,002*	*2%*		
			Data Handling	*5,010*	*10%*		
			Illegal Use	*6,700*	*13%*		
		Real Losses 49,252	Line Leakage	19,927		39.4%	540
			Tank Overflow	0			
			Service Conns	24,062			
			Service Line	5,264			

Then further adjustments are made to develop a new Water Balance combining Bottom-up and Top Down Analyses

Water Balance - Bottom Up All volume values in m3/day					% of Losses	% of SIV	L/Conn/Day
System Input Volume 125,048 (Imports included, but exports removed)	Authorized Consumption 53,064	Billed Authorized Consumption, w/o exports	50,103			40.1%	549
		Unbilled Auth Cons	2,961			2.4%	32
	Water Losses 71,986	Apparent Losses 21,903	Meter Error			17.5%	240
			Stopped Meters				
			Data Handling				
			Illegal Use				
		Real Losses 50,083	Line Leakage	20,262	40%	40.1%	549
			Tank Overflow	0	0		
			Service Conns	24,468	49%		
			Service Line	5,352	11%		

4.3 Key results

It is obvious that apparent losses are much more than 7% of system input volume – actually about two to three time that amount.

Thus, consumers appear to be using more water than currently estimated – so the relative price of network water to them may increase if apparent losses are reduced, which could influence their behavior by decreasing consumption (substitution effect), or could reduce their well-being through an income effect. This has important implications for the economics of the project. Furthermore, the amount of water that can be recovered from leak reduction is probably lower than anticipated, which also has an impact on the economics of the project.

The method can also be used to determine a water balance FOR EACH Distribution Area, but it needs to be tested in areas with good night flow test results, and then refined.

4.4 Possible Adjustments / Enhancements

If we use a limit analysis for Liters /Connection/Day for all points (not just those where the connection density is above 20 connections per km), the intercept increases, resulting in a new Water Balance, with higher apparent losses

Water Balance - Bottom Up and Top Down					% of Cons	% of SIV	L/Conn/Day
System Input Volume 125,048 (Imports included, but exports removed)	Authorized Consumption 53,064	Billed Authorized Consumption, w/o exports	50,103			40.1%	549
		Unbilled Auth Cons	2,961			2.4%	32
	Water Losses 71,986	Apparent Losses 30,066	*Meter Error*	*10,021*	*20%*	24.0%	329
			Stopped Meters	*1,002*	*2%*		
			Data Handling	*5,010*	*10%*		
			Illegal Use	*14,033*	*26%*		
		Real Losses 41,919	Line Leakage	16,960		33.5%	459
			Tank Overflow	0			
			Service Conns	20,479			
			Service Line	4,480			

This implies either more meter error, more illegal use, or more other apparent losses.

In addition, the base data for the DAs could be adjusted to account for meter error in the consumption measurements, which would refine the resulting water balance even more.

The method should be the subject of a brief scientific peer-review by qualified water network engineers and NRW specialists.

<u>Summary</u>

At the time of approval of the original Evaluation Design Report (EDR), the Millennium Challenge Corporation (MCC) indicated that there was insufficient information to approve all elements proposed in Component 3: Impacts on WAJ Zarqa (Impact/Performance Evaluation). Given the importance of Component 3 to the overall evaluation proposal and evaluation logic, it was agreed that a more detailed proposal should be submitted as an annex to the EDR once that additional information was collected and organized.

The following presents a detailed outline of the activities proposed to address the three main questions related to utility performance, summarized as:

4. **Impacts on utility cost recovery**: Does the net cost recovery of the utility improve due to the Compact, and is this related to service improvements?
5. **Operations and maintenance**: What is the impact of the Compact on the budget and execution of O&M?
6. **Service improvements**: At the utility level, are there measurable changes in service delivery quality trends in Zarqa relative to those of other municipal utilities in Jordan?

In answering these questions, three discrete sets of activities (referenced throughout as "elements") are proposed.

<u>Element A: Enhanced analysis of standard indicators of system-wide and local-scale utility performance – Performance Evaluation</u>

This element of Component 3 was approved by MCC. However, we have updated the description to better correspond to the effort we envision, which is reorganized slightly from the version included in the original EDR, and consists of four main activities:

1. Establishing a protocol for utility data collection that does not duplicate other reporting methods. In parallel with this, we are cross-checking the data where possible to identify inconsistencies in the information, and to make adjustments to the data collection protocol as needed.
2. Improving the understanding of utility performance (particularly for NRW) through tracking of indicators for the pre-investment period covering 2008-2013, and conducting basic trend analysis of key indicators alongside analyses of progress on network restructuring and other Compact activities.
3. Including local-scale intensive monitoring of performance indicators. A sub-set of indicators **that can be spatially disaggregated** (at least to the DMA level) could be collected for specific project zones that are deemed comparable and are subject to different levels of treatment with Compact interventions.
4. Carrying out comparative analyses of Zarqa and other Jordanian utilities' performance indicators, controlling for underlying differences in starting points and trends (due to geographic, population/scale, climatic, water availability, or institutional factors) across utilities.

Element B: Engineering tests to better understand actual subscriber consumption as well as the components of Non-Revenue Water (NRW) in treatment and control areas – Impact Evaluation

It is important to attempt to assess meter accuracy to better measure changes over time in consumption in treated and control areas included in the household survey, and thereby obtain additional understanding of NRW. The proposal aims to conduct water meter testing in treatment (areas benefitting directly from MCC investment) and control (zones not receiving such improvements) areas that vary with respect to meter types, sizes, ages, water pressure, and perhaps a few other important parameters. Thus, we will be able to apply corrections to the water consumption data in areas covered by the meter testing, and hopefully, as described in this report, to also apply corrections to areas outside of the meter testing zone.

Element C. Other data analysis – Impact/Performance Evaluation

The original EDR also included a proposal to analyze geocoded secondary data on water quality, customer complaints, and sewer overflows. The first of these (water quality) was deemed infeasible based on field visits conducted in October 2013, while the latter two (geocoded customer complaints and sewer overflows) are collected in a detailed GIS database by WAJ-Zarqa, which should enable us to incorporate it into the analysis at a future date. This system includes customer addresses associated with each complaint.

Timeline Considerations

Most critically, the timeline for Element B of Component 3 should be closely aligned with that for Component 1, since these provide complementary information that is critical for answering general integrating questions about the welfare implications of the Compact. More specifically, it is critical to assess the accuracy of baseline network water consumption and storage data collected through Component 1. Since meters will be replaced over the course of Compact implementation, it is imperative that baseline meter accuracy be measured close to the time of the surveys, which were started in March 2014.

Resource and Budget Requirements

Based on our assessment of current obligations, we are working under the assumption that roughly $10,000 can be allocated to fund some parts of this task from the Evaluator's "Other Direct Costs" budget line this year. Required resources for carrying out this work at baseline and endline can be grouped into two main categories: 1) Equipment and 2) Human resources and consumables needed for carrying out the tests. Since we are proposing tests that will feed directly into the evaluation design, we believe that the Independent Evaluator should pay for the latter out of the "Other Direct Costs" budget line. The equipment that is purchased, however, will be useful for other activities and tests conducted by WAJ-Zarqa, and we therefore do not consider that it makes sense for the Independent Evaluator to purchase these materials.

Annex E: Component 3: Impacts on Water Authority of Jordan-Zarqa – Details of Impact/Performance Evaluation

At the time of approval of the original Evaluation Design Report (EDR), the Millennium Challenge Corporation (MCC) indicated that there was insufficient information to approve all elements proposed in **Component 3**: Impacts on WAJ Zarqa (Impact/Performance Evaluation).

Reminder of rationale for Component 3, and evaluation questions

This component was mostly developed in light of the possibility that some of the benefits of the Jordan Compact may not be directly reflected in welfare changes measured among households and enterprises in Zarqa, nor among the farmers who may receive additional flows of treated wastewater for their irrigation activities. Rather, many of the benefits of the investments may be captured by the local water utility, the WAJ-Zarqa, or by other larger government institutions responsible for water delivery in Jordan, including the central WAJ, the Jordan Valley Authority (JVA), or the Ministry of Water and Irrigation (MWI). Benefits captured by these institutions could in turn lead to reductions in public debt in Jordan and free up capital for other productive economic activities nationwide.

While attributing such economy-wide changes is not a tractable question for the Impact Evaluation to address (there is no appropriate design to implement appropriate experimental or quasi-experimental control for such changes), the set of activities proposed in Component 3 should help to clarify impacts on households (via the meter testing element), and should elucidate whether positive changes are occurring at the utility and neighborhood levels (via the longitudinal tracking of utility performance measures included in Element A).

Our first proposal in the EDR was to conduct enhanced and comparative analysis of standard indicators of system-wide utility performance (Component 3 Element A). Thus, we could at least provide an approximate sense of the scale of improvements in water management – through analysis of trends in performance indicators – that might be related to the Compact. In addition, the comparative analysis with other WAJ units that were less likely to be affected by the infrastructure and other MCC investments in Zarqa would help to indicate whether the observed trends could really be linked to the Compact, rather than other general dynamic influences and service improvements occurring Jordan-wide (e.g., general water supply and demand forces in Jordan; institutional reform and corporatization of utilities).

In addition, we proposed to couple these utility-scale analyses with local-scale measurements and engineering tests (hereafter referred to as meter accuracy tests, given the revised design proposed in this annex) designed to assess the validity of some of the key outcome measures (namely estimates of network water consumption) monitored using the household surveys in treatment and control areas (Component 3 Element B).[16]

Thus, Component 3 was developed to address the following primary evaluation questions:

[16] There are many types of engineering test – night flow tests to estimate the leakage part of NRW, and meter tests to determine the meter error part of NRW and to correct water consumption data obtained in the household surveys

- **Impacts on utility cost recovery**: Does the net cost recovery of the utility improve due to the Compact, and is this related to service improvements?
- **Operations and maintenance**: What is the relationship between the Compact and the budget and execution of O&M?
- **Service improvements**: At the utility level, are there measurable changes in service delivery quality trends in Zarqa relative to those of other municipal utilities in Jordan?

Of course, beneath these questions lie a series of sub-questions related to specific indicators of utility performance, including those dealing with the reliability of water delivery, magnitude of lost revenues due to NRW (physical and administrative components), the cost savings from reduced pumping requirements and more efficient operations, the reduced financial and aesthetic losses from avoided repairs to the distribution network or for management of sewer overflows.

Finally, Component 3 Element B (meter testing) will help to clarify a key variable – household-level consumption of network water – that is required for answering one of the **principal** evaluation questions: **Impacts on water consumption**: Does the WNP change the quantity of water consumed at the household level?

Timeline and implementation of Component 3

Given the initiation of Compact activities in Zarqa and the launch of baseline surveys, the timing of certain elements of Component 3 has become critical. At the most basic level, the initiation of Component 3 should be closely aligned with that for Component 1, since these provide complementary information that is critical for answering general integrating questions about the welfare implications of the Compact, as discussed in the EDR.

More importantly, however, the Component 1 surveys will collect baseline data on network water consumption and water storage in matched treatment and control areas located in Zarqa (and Amman), and it is of fundamental importance to assess the accuracy of these consumption data. Since the water consumption data that will be collected in the baseline survey will largely – though not exclusively – be obtained from water meters and billing records, errors in metered consumption may be systematically different (either high or low) in treatment areas relative to control areas, due to a range of factors discussed in further detail below. Unless meter (and other) errors are independent of consumption, which seems unlikely, such differences will introduce an unknown bias into our estimates of the impact of the compact (since relative consumption changes could be over- or under-estimated). The fact that meters will be replaced over the course of the Compact further complicates things, since errors in measured consumption could change (reduce) through the compact investments themselves, with potentially complicated and differential effects on water bills and on water sourcing behaviors. Thus, we need to understand meter accuracy alongside of measurement of those consumption levels and water management behaviors, which will only be precisely measured at the time of the surveys. If we were to conduct meter testing at a later time, any errors detected could be confounded by seasonal or time-varying water management practices, or more seriously, by the Compact itself.

Given the importance of Component 3 to the overall evaluation proposal and evaluation logic (as depicted in Figure D.1 of the EDR), it was agreed during discussions of the EDR that a more detailed proposal should be submitted as an annex once the additional information required for approval was collected and organized. More specifically, Element A (enhanced data collection and analysis of WAJ-Zarqa) was approved by MCC, while additional information was requested regarding Element B (engineering tests and further analysis of non-revenue Water (NRW) at WAJ-Z).[17] In addition, MCC requested additional guidance regarding the indicators to be used in Element A, many of which are not currently included in the Indicator Tracking Table (ITT). In order to facilitate the drafting of this augmented proposal, MCC approved additional effort and travel by the Social Impact team, and specifically the Senior Technical Advisor, Alan Wyatt.

This document includes the requested information, and is organized as follows. First, in this section EE.1, we review key aspects of Component 3 as it was presented in the EDR, updating it based on developments that have occurred since the submission of the EDR. Section EE.2 then lists the indicators related to the performance of WAJ-Z that will be tracked in Element A, categorized into three specific groups: 1) current ITT indicators; 2) indicators that are not being reported in the ITT but that can be computed from data being collected as part of the ITT process; and 3) indicators that do not figure in the ITT process but that are critical to better understanding the evolution of the utility's performance over time. Section EE.3 presents the rationale and revised proposal for Element B, which is now focused on meter testing in a subset of treated and matched control areas being monitored in the household survey. Section EE.4 concludes. **We hope that MCC will consider the time-sensitive nature of the proposal, given the need to harmonize the effort detailed in Element B with the field data collection activities that have been initiated under Component 1.** The specific rationale for this coordination with Component 1 data collection activities is discussed further in the timeline at the end of Section EE.2.

EE.1. Review of EDR Component 3

In the original EDR, we proposed to include 3 elements in Component 3, as summarized in Table EE.1, and described below.

Table EE.1 Summary of Proposed Activities in Component 3

Component	Evaluation Methodology[a]	Timing
Component 3: Impacts of Compact on NRW, and changes in relative performance of WAJ-Zarqa	Element A: **Performance Evaluation.** Augmented tracking of utility performance Element B: **Impact Evaluation.** Small number of meter accuracy tests in areas included in Component 1. Element C: **Impact/Performance Evaluation.** Other geo-coded data collection over areas included in Component 1 (and across Zarqa).	Ongoing data collection

[a] MCC distinguishes between two types of evaluations, 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

[17] Element C was also not approved, but it is not urgent at this time as it involves the analysis of secondary data. In addition, this revised proposal reassigns Element C activities to Element 1.

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.

EDR Element A: Enhanced analysis of standard indicators of system-wide and local-scale utility performance – Performance Evaluation

This element consists of augmentation of the utility-scale measures included in the ITT of the MCA-J's M&E with additional indicators that will enable generation of a more complete picture of the performance of the water and sewer networks under utility management, as well as operational efficiency, the degree of utility cost recovery, and overall financial sustainability.

The effort we envision, reorganized slightly from the version included in the original EDR, consists of four main activities:

7. Establishing a protocol for utility data collection, in close collaboration with the M&E unit of the MCA-J (and the ITT process), that allows base data to be gathered in a way that does not duplicate other reporting methods. In parallel with this, we are cross-checking the data where possible to identify inconsistencies in the information, and to make adjustments to the data collection protocol as needed. The field visits in October 2013 revealed that development of data accuracy ratings would not be possible in this context. Data accuracy ratings are very useful and are increasingly being used by US-based and international organizations. But they require significant information on the methods used to collect and verify the data, which are not available in Jordan at this time (*The development of the data collection protocol is in progress, we are currently discussing the indicators with MCC and MCA-J, including whether they can be incorporated into the ITT – see details in Section EE.2 for the complete list of the proposed indicators.*)
8. Improving the understanding of utility performance (particularly for NRW) through tracking of indicators for the pre-investment period covering 2008-2013, and conducting basic trend analysis of key indicators alongside analyses of progress on network restructuring and other Compact activities. (*This activity is in progress, with MCA-J, we have developed the initial system for analyzing the indicators that were collected during the pre-investment baseline period.*)
9. Including local-scale intensive monitoring of performance indicators. In keeping with the basic scientific premises of an IE, a sub-set of indicators **that can be spatially disaggregated** (at least to the DMA level) could be collected for specific project zones that are deemed comparable and are subject to different levels of treatment with Compact interventions. This would require some effort to improve the connection between WAJ-Zarqa GIS and billing system databases. (*This activity is in preparation, once the WAJ-Z indicators and sampling strategy for Component 1 have been finalized, we will prepare a data request to submit to the WAJ for such disaggregated data, and explore the feasibility of its collection, which largely depends on whether additional LOE could be allocated to Alan Wyatt. Preliminary assessments of the availability of disaggregated data are summarized in Section EE.2.*)
10. Carrying out comparative analyses of Zarqa and other Jordanian utilities' performance indicators, controlling for underlying differences in starting points and trends (due to

geographic, population/scale, climatic, water availability, or institutional factors) across utilities. While it is not clear if sufficient information is available to conduct these comparative analyses on all parameters, SI Team members have performed such comparative analyses on a subset of indicators in other contexts, without large extra data collection efforts. (*The details of this activity are still under discussion. We have collected limited data for WAJ governorate departments and the corporatized utilities. Once the WAJ-Z indicators have been finalized, if MCC agrees, we will prepare a data request for other utilities covered by WAJ and/or the PMU in Jordan.*)

In addition to this, if funds and data allow, the IE team could address the topic of "economic level of losses", which is suggested as a focus in the Investment Master Plan. This topic, which actually relates to a financial criterion for a water supplier, has been widely discussed by water analysts and researchers interested in NRW, and has been applied to the case of Aqaba in Jordan (Wyatt and Alshafey 2012). The basic idea is that a utility can increase spending on operations and maintenance (O&M), thereby reducing losses over time, but that eventually, the marginal cost of saving NRW will exceed the marginal returns of savings. So, an optimal, target level of NRW can be determined. The modeling work will also lead to O&M guidelines on the optimal water meter replacement frequency and optimal frequency of leak detection campaigns. For the case of Zarqa, given the high cost of alternative water sources (e.g., imported Disi water), such a modeling exercise could be highly informative tp determine if resource allocation for O&M, and O&M practices will keep the water supply systems operating at a high efficiency, assuring long terms benefits to the local population. The model could be applied in Zarqa. It is unclear whether this analysis is desired by MCC at this time, and it is not critical to the success of the impact evaluation; however we ask that MCC and MCA-J describe to us their interest in seeing it done by our team.

EDR Element B: Meter accuracy tests to better understand actual subscriber consumption in treatment and control areas – Impact Evaluation.

As discussed in the original EDR, the evaluation team has concerns over the baseline assumptions about non-revenue water (NRW) and household water consumption used in the development of the Jordan Compact, assumptions that have bearing on the potential benefits of the project. We therefore feel it is important to attempt to assess meter accuracy to better measure changes over time in consumption in treated and control areas included in the household survey, and thereby obtain some additional understanding of NRW. The most critical aspect of this is obtaining better estimates of water consumption by subscribers, given that increasing water availability is a key objective of the Compact.

The conventional method for determining the real losses and apparent losses involves the pairing of records of the frequency, flow and duration of leaks and bursts in a system with a network operational test – called *minimum night flow analysis*. During night flow analysis, utility personnel measure water consumption on several successive nights and determine the minimum flow (often around 2-3 am, when both consumption and apparent losses are likely to be low; surveys must be conducted to identify and characterize commercial or other users consuming water late at night). An accurate assessment of night time real losses can then be made from minimum night flow less known night time consumption. Night flow tests are sometimes paired with (simpler) meter testing to double check the magnitude of all the components of NRW. In addition, pressure transducers are used for water pressure measurements throughout the test period, so that pressure corrections can be applied to compute daytime real losses. Such tests are challenging logistically, and expensive: they require trained operational personnel and public involvement. In the case of water systems with intermittent supply, the water supply schedules may have to be modified, causing disruption to users inside and outside the test area. They are particularly difficult in baseline conditions with networks in poor condition.

Given these challenges, and based on Alan Wyatt's review of the pre-Compact engineering analyses conducted by the engineering firm Nicholas O'Dwyer and of other studies on meter accuracy (detailed below) as well as discussions with engineers at MCA-J and at WAJ, we have significantly modified the original EDR proposal for Element B. Our revised proposal – which is limited to meter testing around the time of the Component 1 surveys is detailed in Section EE.3 below. To be clear, the revised proposal **will not** include minimum nightflow analysis and associated surveys, yet will still provide critically important data on actual water consumption by subscribers enrolled in the household survey in Zarqa, as well as some useful information on various components of NRW (particularly better understanding of the balance of real and apparent losses). The proposal aims to conduct water meter testing in treatment (areas benefitting directly from MCC investment) and control (zones not receiving such improvements) areas that vary with respect to meter types, sizes, ages, water pressure, and perhaps a few other important parameters. The goal of this stratified testing of meter accuracy is to develop a mathematical tool to "correct" the measured water consumption in all census blocks located in Zarqa, though the extent to which out-of sample extrapolation will be possible will need to be assessed (details follow in Section EE.3). At the very least, we will be able to apply corrections to the water consumption data in areas covered by the meter testing.

<u>Element C. Other data analysis – Impact/Performance Evaluation</u>

The original EDR also included a proposal to analyze secondary data on water quality, customer complaints, and sewer overflows as a third element of Component 3. The first of these (water quality) was deemed infeasible based on field visits conducted in October 2013, while the latter two (customer complaints and sewer overflows) are collected in a detailed GIS database by WAJ-Zarqa, that will be folded into the disaggregated analysis proposed in Element A. This system includes customer addresses associated with each complaint. *We will prepare a data request to access such data in the future, as part of Element A.*

<u>EE.2. Details of indicators to be monitored in Element A: Enhanced data collection and analysis of WAJ-Z</u>

The specific objectives of Element A are to:

1. Generate a more complete picture of the evolving performance of the water and sewer networks under utility management over time, as well as operational efficiency, the degree of utility cost recovery, and overall financial sustainability.

2. Where possible, obtain spatially disaggregated data that can be utilized to enhance the insights and analyses conducted under the impact evaluation.

This augmented tracking of the performance (service quality, efficiency and financial) of the water utility in Zarqa consists of 4 activities, detailed below.

Activity 1: Establishing a protocol for utility data collection (*In progress)*

The M&E Unit of MCA-J is already collecting a variety of utility performance indicators at the level of the WAJ-Zarqa utility, and these are useful for the evaluation. Nonetheless, we believe it important to augment these measures with additional indicators that will enable generation of a more complete picture of the performance of the water and sewer networks under utility management, as well as operational efficiency, the degree of utility cost recovery, and overall financial sustainability. These additional metrics are consistent with typical norms for utility management / monitoring best practice, as well as with the current reporting and analysis conducted by the Jordanian water utilities (Aqaba, Yarmouk and Miyahuna) currently reporting to the Project Management Unit (PMU) of the WAJ, responsible for privatization of water utilities.

To facilitate interpretation and harmonization with the Monitoring and Evaluation (M&E) activities of the MCA-J, and in an attempt to establish a protocol for utility data collection (activity 1 of this element as described above in Section EE.1), we have organized the indicators to be tracked in Element A into three types (Table EE.2; with input data for computing the indicators listed in Table EE.3):

1. Current ITT indicators;
2. Indicators that are not being reported in the ITT but that can be computed from data being collected as part of the ITT process; and
3. Indicators that do not figure in the ITT process but that are critical to better understanding the evolution of the utility's performance over time.

It is our recommendation that the M&E unit at MCA-J work with WAJ-Zarqa to collect this full set of indicators on a quarterly basis, and incorporate them into a revised ITT or other M&E tracking system. Though we understand that MCC and MCA-J may want to keep the number of indicators included in the ITT process to a manageable number, **it is our recommendation that all of these indicators be integrated into it.** There are two reasons why we feel this is important. The first is completeness. The ITT indicators currently provide a very limited view of utility performance, omitting key information that is tracked for understanding utility efficiency and financial status, under typical norms for utility management / monitoring best practice. The second relates to transparency. Many of the ITT indicators are not directly measurable, and come out of calculations (ratios, multiplications, etc.) using measurable data (using indicators of type 2 above). It would be best to allow consumers of the information included in the ITT to understand how these aggregated indicators were obtained from the data provided by the utility.

In particular, we would draw special attention to one of the most important indicators associated with the Water Network Project: NRW. Since the objectives of the project include both increasing network water consumption and reducing NRW (both of which are inputs to the calculation of %

NRW), the percentage measure of NRW included in the current ITT becomes misleading. This problem was detailed in the original EDR. Thus, instead of % NRW, we recommend using the IWA recommended NRW indicators - Liters per subscriber per day and m^3 / kilometer of mains / day.

Our team has begun to conduct and display such analyses with the data currently available. We will continue to develop these methods for visualizing the service quality, efficiency and financial performance of WAJ-Z over time, and will report on the first set of these analyses at the appropriate time (i.e., when deliverables related to Component 3 are due). Additional data are needed to compute all indicators on a system-wide basis, as detailed in tables EE.2 and EE.3. This activity is led by Alan Wyatt. Table EE.4 outlines inputs required for this and other Activities under Element A of Component 3.

Table EE.2 Summary of System-Wide Indicators to be Included in WAJ-Z Performance Evaluation

Category		Indicator	**Type 1** In ITT Now	**Type 2** Not in ITT, but data already collected	**Type 3** Not in ITT, data not currently already collected	Can be determined at zone level?
Contextual Information		Year	**Y**			
		Quarter	**Y**			
		Number of Water Subscribers		**Y**		**Y**
		Number of Sewer Subscribers		**Y**		**Y**
		Number of Water Connections			**Y**	**Y**
		Number of Water Meters			**Y**	**Y**
		Water Distribution Line length, Km		**Y**		**Y**
		Sewer Collection Line length, km		**Y**		**Y**
		Average Quarterly Water & Sewer Bill			**Y**	**(Y)**
		% completion of water network project		**Y**		**Y**
		% completion of sewer network project		**Y**		**Y**
Service Quality	1	Residential Water Use, lpcd	**Y**			**(Y)**
	2	Network Water Use, m^3/Subscriber-quarter		**Y**		**(Y)**
	3	Hours of Water Supply / Week	**Y**			**Y**
	4	Water Complaints/1000 Water Subscribers		**Y**		**(Y)**
	5	Percent of Water Subscribers w/Sewer Conn	**Y**			**Y**
	6	Sewer Complaints/1000 Sewer Subscribers		**Y**		**(Y)**
	7	Volume of wastewater collected, Mm3/qtr	**Y**			
	1	Total NRW, L/Subscriber/day		**Y**		**Y***

Operations and Maintenance	2	Total NRW, m3/km/day		**Y**		**Y***
	3	Disi Water Imported / Total System Input			**Y**	
	4	Energy Consumption kwhr/m3 pumped			**Y**	
	5	Bursts and Leaks / km water mains		**Y**		**Y**
	6	Sewer Blockage Events / km of sewer line		**Y**		**Y**
	7	Water Meters Replaced / Total		**Y**		**(Y)**
Finances and Cost Recovery	1	Operating Cost Coverage	**Y**			
	2	Billing and Collection Efficiency, %			**Y**	**(Y)**
	3	Variable O&M costs / Sales revenue			**Y**	
	4	Salary Cost / Total Cost, or per m3 water SI			**Y**	
	5	Electricity Cost / Total Cost, or per m3 water SI			**Y**	
	6	Maintenance Expense / Total Cost			**Y**	
	7	Outstanding Debt (Arrears)	**Y**			

Notes: (Y) will be possible if interconnection between billing system and GIS is improved; Y* once strategic metering is in place

Colors indicate contextual information (grey), service quality (blue), O&M (orange), and finances and cost recovery (green)

Table EE.3 Input data required for computing Table EE.2 indicators

Data Parameter	**Data source / Report**	**Suggested timing of reporting**
Number of Water Subscribers	WAJ Zarqa Subscriber / Water Report	End of quarter
Number of Sewer Subscribers	WAJ Zarqa Subscriber / Water Report	End of quarter
Number of Water Connections	WAJ Zarqa Subscriber / Water Report	End of quarter
Number of Water Meters	**New data needed from WAJ Billing System**	End of quarter (?)
Water Distribution Line length, Km	**New Info needed from WAJ Zarqa or PMC**	End of quarter
Sewer Collection Line length, km	**New Info needed from WAJ Zarqa or PMC**	End of quarter
Average Quarterly Water & Sewer Bill	**New Info needed from WAJ Zarqa**	Average over quarter (?)
% completion of water network project	**New info needed from PMC**	End of quarter
% completion of sewer network project	**New info needed from PMC**	End of quarter
Residential Measured Water Use	WAJ Zarqa Subscriber / Water Report	Total over the quarter
Non-Residential Measured Water Use	WAJ Zarqa Subscriber / Water Report	Total over the quarter
Total Measured Water Use	WAJ Zarqa Subscriber / Water Report	Total over the quarter
Hours of Water Supply / Week	**Source Unknown…WAJ Zarqa ?**	Average over quarter
Total Water Complaints	WAJ Zarqa Complaints Report	Total over the quarter

Total Sewer Complaints	WAJ Zarqa Complaints Report	Total over the quarter
Volume of wastewater collected, Mm3/qtr	As Samra Data Reports	Total over the quarter
Water produced (m3)	WAJ Zarqa NRW Report	Total over the quarter
Imported drinking water (m3)	WAJ Zarqa NRW Report	Total over the quarter
Imported water from Disi pipeline	**New data needed**	Total over the quarter
Total System INPUT	WAJ Zarqa NRW Report	Total over the quarter
Exported treated drinking water (m3)	WAJ Zarqa NRW Report	Total over the quarter
Total NRW	WAJ Zarqa NRW Report	Total over the quarter
Energy Consumption kwhr for pumped	**New data needed**	Total over the quarter
Bursts and Leaks	WAJ Zarqa Complaints Report	Total over the quarter
Sewer Blockage Events	WAJ Zarqa Complaints Report	Total over the quarter
Water Meters Replaced	WAJ Zarqa Complaints Report + PMC	Total over the quarter
Water & Sewerage Service Revenues	WAJ Zarqa or WAJ HQ Financial Data	Total over the quarter
Water Export Revenues	WAJ Zarqa or WAJ HQ Financial Data	Total over the quarter
Sewer Fee Revenues	WAJ Zarqa or WAJ HQ Financial Data	Total over the quarter
Other Revenues	WAJ Zarqa or WAJ HQ Financial Data	Total over the quarter
Total Revenues	WAJ Zarqa or WAJ HQ Financial Data	Total over the quarter
Billed Water Volume or Billed Value	**New data needed from WAJ Billing System**	Total over the quarter
Outstanding Debt (Arrears)	**New data needed (WAJ HQ?)**	Total over the quarter
Total Operating Costs	WAJ Zarqa or WAJ HQ Financial Data	Total over the quarter
Salaries Costs	WAJ Zarqa or WAJ HQ Financial Data	Total over the quarter
Electricity Costs	WAJ Zarqa or WAJ HQ Financial Data	Total over the quarter
Fuel Costs	WAJ Zarqa or WAJ HQ Financial Data	Total over the quarter
Chemicals Costs	WAJ Zarqa or WAJ HQ Financial Data	Total over the quarter
Water Import/Purchase Costs	WAJ Zarqa or WAJ HQ Financial Data	Total over the quarter
Desalination costs	WAJ Zarqa or WAJ HQ Financial Data	Total over the quarter
Network and Plant Maintenance Costs	WAJ Zarqa or WAJ HQ Financial Data	Total over the quarter

Notes: Colors indicate contextual information (grey), service quality (blue), O&M (orange), and finances and cost recovery (green)

Table EE.4 Element A Activities and Details (Status, effort, final products)

Activity	Topics	Status	Data Sources	Principal implementer	Frequency	SI LOE	Final product
1	Protocol for utility data collection	Underway. Availability of certain data being investigated	WAJ-Zarqa WAJ-HQ As Samra Others	MCA-J with training and periodic assistance from SI Team Alan Wyatt	Quarterly	10 days plus 1 trip in 2014 4 days in yrs 3 and 4 10 days plus 1 trip in 2017	Modified ITT including utility parameters
2	Historical and ongoing analysis on NRW	Historical analysis nearly complete	WAJ-Zarqa	SI Team –	Quarterly	2 days per year plus 3 days for final report	Evaluation annex focusing on NRW performance
3	Local-scale monitoring of performance indicators	Awaiting completion of household survey data collection	WAJ-Zarqa GIS and WAJ-Zarqa MMS	SI Team Alan Wyatt	Baseline and endline	5 days in 2014 5 days in 2017	Contribution to HH survey reports
4	Comparative Analysis of Zarqa to other utilities	Pending	ITT WAJ-Zarqa PMU	SI Team Alan Wyatt	Annual or just endline	TBD	Analysis of comparative utility performance improvement at endline

Our team has begun to conduct and display such analyses with the data currently available. We will continue to develop these methods for visualizing the service quality, efficiency and financial performance of WAJ-Z over time, and will report on the first set of these analyses at the appropriate time (i.e., when deliverables related to Component 3 are due). Additional data are needed to compute all indicators on a system-wide basis, as detailed in tables EE.2 and EE.3. This activity is led by Alan Wyatt. Table EE.4 outlines inputs required for this and other Activities under Element A of Component 3.

Activity 2: Improving the understanding of utility performance through tracking of indicators over time (*In progress)*

The second activity is improving the understanding of utility performance (particularly for NRW) through more detailed tracking of selected indicators for the pre-investment period covering 2008-2013, and conducting ongoing trend analysis of these key indicators alongside analyses of progress on network restructuring and other Compact activities. This type of pre-investment data is only available for the basic IWA water balance and can facilitate tracking of several NRW indicators. This analysis will follow IWA water balance definitions and protocols – unlike previous MCC and MCA efforts. As noted above, this analysis will use the IWA recommended NRW indicators - Liters per subscriber per day and m^3 / kilometer of mains / day. This activity has been nearly completed for the period 2008 – 2013. A small number of hours will be needed to extend this trend analysis in future years, as outlined Table EE.4.

Activity 3: Including local-scale intensive monitoring of performance indicators (*In preparation, pending discussion on indicators in Component 1)*

The third activity consists of local-scale data collection that is aligned with the surveys included in Component 1. Pending agreement on and a process being established for data collection procedures for Activity 1, w**e are preparing a request to WAJ-Z for provision of this local-scale data**, based on information Alan Wyatt collected while in Jordan in October 2013 (indicated in the right-most column in Table EE.2). The locations for this intensive monitoring still need to be determined based on the results of the household survey sampling procedure. Local-scale utility data will provide complementary data that can be triangulated with data from the household surveys to track progress and changes over time in the areas affected by the compact, and will increase operational monitoring knowledge at WAJ-Z. The analyses that will be performed will be similar to those conducted on system-wide indicators, with an eye to extrapolating to other zones in the city if possible. The LOE estimates for Alan Wyatt for this activity are listed in Table EE.4.

Activity 4: Carrying out comparative analyses of Zarqa and other Jordanian utilities' performance indicators (*In preparation)*

The fourth and final activity in Element A is to carry out comparative analyses of Zarqa and other Jordanian utilities' performance indicators, controlling for underlying differences in starting points and trends (due to geographic, population/scale, climatic, water availability, or institutional factors) across utilities. At this time, we have not prioritized this comparative assessment because we judged that the data, if available, could be collected at a later time. If not available, pursuing a request to WAJ-Amman and the PMU at this time would be a distraction for the main evaluation. **This activity is therefore pending at this time and will be re-visited in January 2015, at which time any additional required effort and resources will be further discussed.**

EE.3. Proposal for Element B

EE.3.1. Background

The estimates of baseline and current network water consumption by subscribers that underlie the calculations in the investment case for the Jordan Compact are largely based on water meter readings and/or assumptions about the percentage of NRW in Zarqa. The EDR and subsequent discussions with many project participants have raised questions about the accuracy of these estimates. Analysis presented in the EDR suggests that the NOD study greatly under-estimated apparent losses and over-estimated real losses. The EDR recommended that *"In addition, the IE should measure the 2 main components of NRW – Apparent (Administrative) Losses and Real (Physical) Losses. These different types of losses have different implications for project finances, the behavioral responses of and benefits flowing to potential beneficiaries, and the ultimate* *changes and distribution of project economic outcomes. A good baseline and accurate ongoing measurement of the separate components of NRW – unbilled, authorized consumption, real losses, apparent losses – is therefore necessary to track the impacts of the Compact investments, and especially for teasing out their effects on utility cost recovery and consumer well-being."*

During a field visit by Alan Wyatt in October 2013, he carefully reviewed all of the available pre-Compact engineering studies, including the Investment Master Plan (NOD 2010), the DMA Prioritization Study (NOD 2012), the Pilot Leak Detection Study (NOD 2012) and the Batrawi Distribution Area Drawings (NOD 2012). It became clear that a complete re-assessment of the water balance would require an expensive series of engineering tests and data collection / analysis which were outside the scope of the Impact Evaluation. Efforts to measure water meter accuracy, on the other hand, are essential to obtaining an accurate understanding of the effects of the Compact on network water consumption – the so-called substitution benefits. Indeed, perhaps the single most important outcome for the evaluation to attribute accurately to this investment program is a change in network water consumption. As an added benefit, zone meter testing (described below) will provide localized information on the level of real losses and illegal consumption.

The source of the problem with measuring water consumption is that many of the meters in Zarqa are very old, and have suffered wear due to various effects of intermittent water supply and roof tanks. In general meters, decline in accuracy with age, but under-registration is exacerbated by low flows associated with roof tanks, high mineral content of the water, air entrainment, and pressure transients. All these sources of error are amplified in intermittent supply situations. However, it is possible that some specific locations within intermittent supply zones could experience high pressures, high flow rates and low meter error. Such locations could even have lower meter error than water systems with continuous supply and roof tanks. Many other areas in the zone would likely have lower pressures and flow rates. The results of the NOD Hydraulic Modeling confirm this basic situation. As part of the pre-Compact design work, Nicholas O'Dwyer only did limited meter surveys (and no tests) in 2008 and found a mix of meter brands and a rate of 2% of completely non-functional meters (NOD, 2008).

Based on this information, and on a review of the evidence from sites in Jordan as well as in other countries (see summary in Table EE.5), we are confident that meter error will be high in most places, perhaps low in a few places and be subject to wide spatial variation, especially given the dominant use of cheaper and less accurate Class B meters in Zarqa.[18] One test conducted in 2000 in Zarqa as a part of a set of JICA-funded water utility studies showed that the registered consumption was only half of the actual consumption.[19] Conversations with Dr Walid Sukkar, a Jordanian NRW expert with extensive experience in the " Middle Governorates" indicated that meter error should vary greatly across different parts of Zarqa, due to the mix of elevations, pressures, and flow rates. Given that one of the main expected impacts of the Compact is to increase water consumption, it is absolutely critical to accurately determine actual consumption

[18] There are four classes of meters, labeled A through D. Class A meters are least accurate and cheapest, and Class D meters are most accurate and most expensive.
[19] Personal communication from NRW Manager at WAJ-Zarqa, October 2013

at baseline and endline. The IE team has therefore recommended a series of water meter accuracy tests in selected areas where household surveys are to be conducted.

The evidence discussed above means that network water consumption data will be measured with varying accuracy across our sample. This raises the possibility of two potential problems. At best, the variable accuracy of meters will introduce non-systematic measurement error (that is unrelated to treatment assignment) that will tend to attenuate estimates of impact (Hutcheon et al., 2010). More seriously, however, measurement error could introduce bias (of unknown magnitude and severity) into our estimates of impacts. There are a variety of potential mechanisms through which such bias could occur, but we describe two particularly likely possibilities here:

1. **Underestimation of impacts**: If treatment areas happen to have older meters (as might be expected given that they were selected for network investments), it would not be surprising if meter error in treated blocks is greater, which would introduce a downward bias in the consumption measures in treatment zones, relative to their matched controls. As a result, the difference-in-difference (DiD) change (or increase) in consumption resulting from the investments will also be underestimated.

2. **Overestimation of impacts:** If older meters in treatment zones are more likely to be replaced (indeed meter replacement is one of Compact activities), baseline consumption in such areas would be underestimated relative to controls, whereas endline consumption would be overestimated relative to controls (which would then have older meters), such that the DiD change in consumption resulting from the investments will be overestimated. In addition, if meter error decreases in water consumption (which is widely considered to be true), the DiD estimate will also increase in consumption.[20]

Given these concerns, and particularly those over systematically different and time-varying meter error (either high or low) in treatment areas relative to control areas, we believe that it is critical to conduct meter accuracy tests in a subset of both treatment and control blocks. The time-varying unobservable changes in meter accuracy, which will influence water bills, water sourcing behaviors, and a whole suite of downstream impacts, are precisely the types of threats that could derail the evaluation. Thus, we need to understand meter accuracy alongside of measurement of consumption and water management behaviors, which will only be precisely measured at the time of the surveys. If we were to conduct meter testing at a later time, any errors detected could be confounded by seasonal or time-varying water management practices, or because of meter replacements, by the Compact itself.

Table EE.5 Meter accuracy studies reviewed by the Evaluation team

Study Author	Date	Location	Meter types	Main findings

[20] We note that WAJ has recently instituted a policy of replacing water meters that are more than 5 years old. The coverage of this replacement program is incomplete at this time, which introduces another layer of uncertainty related to meter accuracy.

Omar	2007	Amman, Jordan	Class B and C	A small sample of new Class B meters under registered by 4% compared to new Class C meters under intermittent supply and by 12% under continuous supply with roof tanks. A small sample of existing Class B meters (age unknown) under registered by 5% compared to new Class C meters under intermittent supply and by 25% under continuous supply with roof tanks. Pressures and flow rates unknown. "Rolled" meters under-register 8% to 14% depending on degree of "roll"
Patrick, R. et al	2010	Amman, Jordan	Class B	NRW was reduced in 9 out of 330 districts in Amman, through water audits and leak repairs. Average meter under-registration was close to 10%. Inlet pressures and flows unreported. 35-58% of the meters were rolled
Sukkar, et al	2011	Balqa, Jordan	Class B	A detailed hydraulic analysis of fill rates of roof tanks in various parts of Balqa showed a range of fill times as high as 30 hours – indicating very low flows and high meter error.
AlShafey	2013	Aqaba, Jordan	Class B	Roof tanks and low water flows cause an average meter error of 23% for existing class B meters (continuous supply). Pressure about 2 bar
Flores, J.. et al	2009	Spain	Varying	Meter error varies greatly depending on meter specs: size, class, type and age. Overall aggregate is 14.3% under-registration.
Fantozzi, et al		Palermo, Italy	Class C	Palermo: New Class C meters, roof tanks and intermittent supply 14% to 45% under-registration
Mutikanga, H.	2009	Kampala, Uganda	Class B-D	Aggregate meter error of 21% with a range from 4% (new Class D) to 72% (15-year old Class B)
Thornton, J.	2013	Sao Paulo, Brazil	Class C	New Class C meters under-registed water flow in to roof tanks by 11% to 16 % under continuous supply situations due to

EE.3.2. Evaluation questions to be answered by Element B

Given the issues summarized above, we have determined that the data collection activities of Component 3 Element B have become critical for answering one of the most important evaluation questions listed in Table B.1 of the EDR, as well as several of the utility-specific questions (as listed below). Through question 1, Element B therefore falls squarely in the domain of the Impact Evaluation, while providing complementary information about NRW components that is important for the Performance Evaluation of WAJ-Zarqa.

1. **Impacts on water consumption (Component 1 question):** Does the WNP change the quantity of water consumed at the household level?
2. **Impacts on utility cost recovery**: Does the net cost recovery of the utility improve due to the Compact, and is this related to service improvements?
3. **Operations and maintenance**: What is the impact of the Compact on the budget and execution of O&M?
4. **Service improvements**: At the utility level, are there measurable changes in service delivery quality trends in Zarqa relative to those of other municipal utilities in Jordan?

EE.3.3. Specific Objectives

Through the meter testing exercise, the Evaluation Team aims specifically to:

1. Conduct a series of tests of residential water meters in selected treatment and control areas where household surveys are being conducted, to obtain more accurate measures of network water consumption, both at baseline and at endline.
2. Once baseline measurements are complete and have been reviewed and assessed, determine the extent to which these tests of selected water meters or small groups of meters can be applied for out-of-sample estimates in other survey zones and to the broader Zarqa system.
3. Where possible, estimate real losses and illegal consumption in network / household clusters where meter testing and household surveys are being conducted.
4. Following baseline testing, determine the extent to which these tests of other NRW components can be applied to the system-wide utility monitoring process.

The Evaluation Team expects that an additional benefit of Element B – beyond the measurement objectives that are most central to the Impact Evaluation – will be the building of capacity among WAJ-Zarqa staff, who will learn to conduct such tests and to analyze their results, in collaboration with the IE Team and Aqaba Water. Regular testing of meter accuracy is an essential but low cost component of good water utility management and directly leads to increased revenue. WAJ-Zarqa have already expressed a keen interest to learn from the experiences of Aqaba Water.

EE.3.4. Methodology

To achieve these specific objectives and improve our confidence in water consumption estimates to be used in the Evaluation, we propose to conduct a series of meter accuracy tests that will allow us to better characterize both the magnitude and variation in meter errors. Thus, these measurements should cover, to the extent possible, the range of situations in Zarqa with respect to meter type, meter age, water supply pressure, roof tank volume, and consumption. In addition, given our expectations of differential rates of meter replacement and changes in consumption during the Compact, the measurements should cover both treatment and control zones included in the household surveys. Ensuring sufficient variation in these variables should allow the Evaluation Team to develop “calibration curves” for real consumption in all survey zones, as detailed in Section EE.3.5.

Given the intermittent water supply situation in Zarqa, water meter testing must be conducted “in-situ”. The basic approach to testing meter accuracy is relatively straightforward, but practical considerations of installation of new meters in congested urban areas can pose challenges. First, technicians temporarily install a highly accurate water meter “upstream” of a WAJ subscriber meter, or group of subscriber meters. They then measure consumption over a period of about 10 days – or roughly 2 water supply cycles. Aqaba Water (AW) has been conducting a series of such water meter accuracy tests in various parts of its urban network, to improve its metering program, increase cost recovery and better assess its water balance. AW is interested to work with WAJ-

Zarqa to achieve the objectives listed above, and will provide training during piloting of the meter testing program as well as analysis of results.

Two different methods are common for in-situ water meter testing in intermittent supply networks. The second has higher accuracy, but a higher difficulty of application. We anticipate that a combination of both methods will be required in Zarqa, as informed by practical constraints in the network, and have anticipated that roughly 40% of the zones will require Method 1 and 60% of zones will require method 2 Details on costs and implementation procedures follow further in Section EE.6 and EE.7.

1. Small zone testing: Small zones or DMAs are temporarily created by a) installing and reading one or more small size (1"-2") high accuracy reference meter(s) at the zone inlet(s) and b) concurrently measuring consumption recorded at the individual subscriber meters. Manufactured meter boxes can be used to eliminate the need for major civil works, other than excavation. Some pipe work is needed for preparing the inlet pipe with needed joints, extensions or reducers for installing the meter.

Prior to starting any testing, a rapid leak detection survey is conducted and any major leaks are repaired (by WAJ)[21]. Next, a pressure test is performed during water supply time to insure that the zone is isolated – and that the measurement at the inlet(s) records all the water going into the zone. Third, all subscriber connections are shut-off for a short period, and the reference meter(s) is used to measure any remaining leakage or illegal use in the zone. Finally, the subscribers are allowed to receive water normally over one or more cycles. Meter readings are taken and recorded by meter readers. In Zarqa, the entire test for a particular zone would be conducted over a period of about 10 days.

This first approach is intended to be small scale and is not intended to be indicative of system wide leakage levels, yet quantifying any leakage -or illegal use- in the zone is important for calculating errors in meters. The difference between the recordings of the main reference meters (we have budgeted for 3 per monitoring site) and the subscriber meters represents NRW. The measured leakages and illegal uses will be deducted, to determine meter error. This method has the advantage of assessing meter error in multiple subscribers at one time, and is most useful where meter type and age are relatively homogeneous.

2. Individual meter testing: High accuracy reference meters (10 per site) can also be installed in series with existing subscriber meters, where suitable and convenient locations can be found, which we anticipate to be 60% of the monitored locations. The reference meters are used to record consumption over several cycles, and the total difference between the two meters is calculated. In this case, the new meter must be sufficiently displaced from the old meter to not generate flow disturbances. This second approach is useful where zones have a large variation in meter types, ages, pressures etc.; where the zone inlet piping is difficult to

[21]Note: If a major leak is found, WAJ-Zarqa will have essentially no choice but to fix the leak, even if it is a temporary fix until they can obtain any additional resources they requires. In the worst case scenario another location could be chosen to meter testing. Small leaks will not impede the meter testing work.

access; or where multiple inlets enter a zone. It provides improvements in accuracy of measurement of individual meter error, but of course require more testing locations / activity than the first method.

EE.3.5. Analysis and Application of Local Results

Analysis of the meter testing results will rely on multivariate analysis of the factors that influence meter error, as shown in Equation 1 below:

$$\chi_{ij}^{*} - \chi_{ij}^{n} = \beta_0 + \beta_{1k} \cdot Z_{jk} + \beta_{2l} \cdot X_{ijl} + + \varepsilon_{ij} + \gamma_i \quad (1)$$

In this equation, χ_{ij}^{*} and χ_{ij}^{n} refer to high-accuracy and normal meter readings covering a household (or group of households) *i* located in zone *j*; Z_{jk} is a vector of *k* zone-specific variables that influence meter error; X_{ijl} is a vector of *l* household-specific (or group-specific) variables that influence meter error; ε_{ij} is an zone-level error term; and γ_i is a household-specific (or group-specific) error term. The coefficientsβ_0, β_{1k}, and β_{2l} are estimated using multivariate regression methods, in order to estimate the factors that drive meter error. If possible, the model will then be used to extrapolate based on measures of X_{ijl} and Z_{jk} to other survey areas where meter testing is not conducted – this will be assessed after baseline. These factors will include variables such as meter type, meter age, meter size, water supply pressure, and the shape and volume of the consumption profile. Additional details on these variables are discussed in detail below.

Meter type: Zarqa predominately uses multi jet velocity type meters. However, the meters may vary by manufacturer, and whether they have been refurbished or are new. The available meter type information in the billing database can be used to determine whether this delineation can be used in a useful manner for extrapolating the test results.

Meter age: Meters under study will be checked for age by using available data in the billing system. This data may have limitations in terms of accuracy due to entry errors. Another cause of concern is the time period covered by the records, since the most immediate method of accessing information uses the current status of the meter, where the dates relate to the installation of the meter for a last customer, and not previous history of using the meter. Further data could be found by more thorough examination of the database, linking the meter to more than one customer. There is also the issue of meter refurbishing, where the meter is equipped with new mechanism while the chassis and hence meter number remains the same. In this case, the meter age can be assumed to correspond to the latest refurbishment.

Meter size: Whether to test different meter sizes remains to be examined. Examining the billed amount for each consumption type may signify that residential meters are the major source of metering errors by volume. However, there is no restriction from performing similar testing to customers with larger meters if needed when choosing a representative sample is being performed.

Network pressure: Network pressure drives tank filling speed among other factors, such as internal leakage and use points that lie before the customer tank. During tests, logged network pressure can be available, yet a system-wide average pressure values are not available. The use of estimations made by utility staff is widely used for providing such value, but more accurate results can be reached by AW constructing simple hydraulic models, in addition to taking instantaneous readings from select points around the network for calibration.

Customer consumption: Billed consumption of the test customers can be compared with the average for Zarqa. Customer consumption, along with pressure and meter size, is indicative of which flows the meter is working under.

Tank volume: The average water tank volume for customers, with consideration for the use of cascading the supply to other storage tanks while the main tank is being filled, can be used with conjunction of pressure and average consumption to draw indication of average water flows through the meter.

WAJ-Zarqa has a billing system with information on subscribers and meters, a GIS with subscribers and network configuration and a maintenance management system which tracks meter replacement. Unfortunately, the interconnection between these systems is poor. It is not possible, at the current time to be able to determine for example the average meter age in a small GIS polygon that is used for the household survey purposes. However, technicians from Aqaba have been able to establish good connections between their information systems, and may be able to help WAJ-Zarqa improve the interconnection between its information systems, facilitating more accurate extrapolation of local meter tests to larger areas. At the very least, the household survey will collect meter numbers from all survey participants; the data in the billing system for these households can thus be readily used for our analysis.

EE.3.6. Measurement program

The first step in design of the measurement program will be to choose locations for meter testing based on the sampling frame for Component 1. This will be done through consultations between Alan Wyatt, Marc Jeuland, AW personnel experienced in meter testing, the M&E Unit at MCA-J and WAJ-Zarqa engineers. In preparation for the sample design for this activity, personnel from Aqaba Water would visit Zarqa, to gather information on the spatial distribution of meter attributes (age, type, pressure etc) and to overlay such information on the survey zones. Such an assessment, performed in collaboration with WAJ-Zarqa personnel will be needed to determine the number of test locations / meters and the type of measurement techniques to be used. Careful planning is necessary to increase the chances that meter testing results can be applied to all household survey zones.

Once the sample is defined to ensure variation across treatment groups and the variables of interest (described above), and following review and consultation with the Independent Evaluator and relevant parties at MCA-J and MCC, AW will help WAJ-Zarqa to carry out pilot testing of each of the two proposed methods in several zones. After these pre-tests, full-scale measurements will commence, timed shortly after completion of the household surveys. Results from the baseline meter accuracy tests will be included in the baseline reports associated with Component 1, while endline test results will be presented in the Component 1 endline report.

EE.3.7. Estimated Resource Requirement and Budget

Based on our assessment of current obligations, we are working under the assumption that roughly $10,000 can be allocated to fund some parts of this task from the Evaluator's "Other Direct Costs" budget line at baseline (this year). Ideally, more resources than this would be available for this activity which would allow sampling across a wider area and improve our ability to apply the results across other zones included in the household survey; the budget estimate in Table EE.6 however has been developed in consultation with contacts at Aqaba Water with this resource constraint in mind. The implications of this budget constraint are explained in detail in the discussion that follows Table EE.6. In addition, we do not believe that it makes sense for the Independent Evaluator to purchase the equipment required for the tests that can be used for other purposes (specifically, the single high accuracy meter and meter box as well as smaller highly accurate water meters), as noted below, which should more logically become the property of WAJ-Zarqa for use during the evaluation period.

Table EE.6 Budget calculations for Component 3 Element B baseline data collection

Summary	Total	SI budget line	Other budget
1. Detailed Planning of Measurement Locations (Travel costs, accommodations and daily compensation for Aqaba engineer; compensation of time for Zarqa engineer)	$1,422	$1,422	$0
2. Training/Pilot testing (Equipment; Excavation; Travel costs, accommodations and daily compensation for Aqaba engineer; Compensation of time for Zarqa engineers, technicians, laborers and meter readers; Vehicle)	$4,996	$2,496	$2,500
3. Full-scale measurements (Additional excavation; Equipment; Compensation of time for Zarqa engineers, technicians, laborers and meter readers; Vehicle)	$5,240	$5,240	$0
4. Data analysis and any follow-up activity (Compensation for Aqaba and WAJ-Zarqa engineers)	$720	$720	$0
5. Presentation of results to the IE Team and MCC / MCA / PMC (Travel and accommodations; Compensation for Aqaba and WAJ-Zarqa engineers)	$494	$494	$0
6. Contingencies (~20%)	$2,628	$2,128	$500
7. Supervision and support from Alan Wyatt (baseline and endline)	6 days	6 days	n.a.
Total	**15,500**	**$12,500**	**3,000**

Notes:
Fixed costs = $7632 (Meters + meter box; Aqaba water support for pilots; Zarqa engineer LOE)
Variable costs = $5240 (Consumables, technicians, laborers, meter readers, fittings for residential meters, leak detection instrument rental)
Cost per additional zone in full-scale test = $5240/16 ≈ **$330**

Summary of resource requirements: Required resources for carrying out this work at baseline and endline can be grouped into two main categories: 1) Equipment and 2) Human resources and consumables needed for carrying out the tests. Since we are proposing tests that will feed directly into the evaluation design, we believe that the Independent Evaluator should pay for the latter out of the “Other Direct Costs” budget line. The equipment that is purchased, however, will be useful for other activities and tests conducted by WAJ-Zarqa, and we therefore do not consider that it makes sense for the Independent Evaluator to purchase these materials.

The equipment needs are the following:
- Highly Accurate water meters (Volumetric or ultrasonic flowmeters) for zone measurements.
- Smaller, highly accurate water meters for testing individual meters
- Meter boxes and pipe fittings to insert new meters into the water line
- Pressure sensors and Data Loggers
- Leak detection equipment – WAJ-Zarqa has some but additional equipment may be needed.
- Material to repair any major leaks found in small zones.

The human resources and consumables needed at baseline are:
- Oversight provided by the Senior Technical Advisor (Alan Wyatt). Level of effort is detailed above.
- Travel, lodging and subsidence expenses for engineers(s) from Aqaba Water. They will provide training and conduct piloting for the baseline during 4 days in Zarqa, as well as 3 days of support for analysis and one day for presentation of results to the IE team and to MCC/MCA-J. Training needs at endline are uncertain at this time, and will depend on whether WAJ-Zarqa incorporates the meter testing methodology into its standard activities.
- Engineers to select measurement locations, oversee technical work, and if warranted develop hydraulic models for network and meter-tank systems. These personnel are included in the training/piloting and full-scale measurements line items specified above.
- Vehicle and work crews to install and remove meters. These costs are included in the training/piloting and full-scale measurements line items specified above.
- Meter readers to record measurements. These personnel are included in the training/piloting and full-scale measurements line items specified above.

As shown in the notes below Table EE.5, there are substantial “fixed costs” in planning, pilot testing, analysis and presentation of results; the “variable costs” of additional tests (mainly additional labor and vehicle time for each full scale test that is conducted) is relatively low at $330. The fixed costs in the budget above are around $7632, and the total variable costs for 16 full scale tests – enough to allow us to sample in 8 matched pairs of treatment and control zones from the WNP and combined WNP/WWNP where water consumption is most likely to change – are about

$5240.[22] The average cost of each of the full scale tests is thus about $330. We estimate that the number of meters tested could be roughly doubled for about $5000 more (to 32 zones), and tripled to 48 for about $10000 more. We strongly recommend increasing the scope of testing to 48 full scale tests because this would dramatically improve our ability to understand and extrapolate the results of our analysis. The Evaluator's budget, however, would not be sufficient for this. That is why we have proposed the scale that is described here.

[22] It is important to consider the implications of this sample size given the description of the methodology included in Section EE.3.5. Of the six variables identified to be potentially important there, we would expect to find significant variation across users within a zone in terms of consumption and water storage (and possibly meter size, if working in areas with larger water users), but meter age, meter type, and network pressure are less likely to vary. Thus we need to sample different types of zones to cover variation in these types of characteristics, and 8 zones will very likely be insufficient to achieve this variation.

REFERENCES

Albert, J.; Alikhan, S.; Etter, L.; Jeuland, M. and Wyatt, A. (2013). “Impact Evaluation Design Report: Jordan Compact – Water Sector.” Social Impact: Arlington, VA.

AlShafey, M. (2013) Unpublished meter test results conducted by the author at Aqaba Water Co, provided via direct personal communication

Flores, J. Diaz, I. (2009) Accurate Assessment of Under Metering in Madrid Water Supply, IWA Water Efficiency Conference; Sydney Australia

Hutcheon, Jennifer A, Arnaud Chiolero, and James A Hanley. (2010). "Random measurement error and regression dilution bias." British Medical Journal 340 (7761).

Mutikanga, H.E, et al. (2009) Assessing Meter Accuracy for reduction of Non-Revenue Water IWA Water Efficiency Conference; Sydney Australia

Nicholas O’Dwyer Consulting Engineers (2010), Preparation of the Feasibility Study and Environmental and Social Impact Assessment for Zarqa Governorate Water System Restructuring and Rehabilitation – Investment Master Plan, Amman, Jordan

Nicholas O’Dwyer Consulting Engineers (2010), Zarqa Governorate Water System Restructuring and Rehabilitation – Investment Master Plan, Batrawi Distribution Area Drawings Typical Detail Drawings, Amman, Jordan

Nicholas O’Dwyer Consulting Engineers (2012) Zarqa Governorate Water System Restructuring and Rehabilitation – District Metering Area (DMA) Prioritization Summary Report, Amman, Jordan

Nicholas O’Dwyer Consulting Engineers (2012) Preparation of the Feasibility Study and Environmental and Social Impact Assessment for Zarqa Governorate Water System Restructuring and Rehabilitation – Final Pilot Leak Detection Report, Amman, Jordan

Omar, M. (2007) Improving Water Distribution and NRW Management in Amman, Jordan, Presented at NRW TC Meeting, Jakarta, Indonesia

Patrick, R. et al (2011) Amman Non-Revenue Water Reduction Program. IWA Water Efficiency Conference; Dead Sea, Jordan

Sukkar et al (2011) Minimizing Leakage in Jordan by Incorporating an Innovative Modelling Method that Simulates Pressure Reduction in an Intermittent Water Supply System. IWA Water Efficiency Conference; Dead Sea, Jordan

Thornton, J. (2013) Unpublished meter test results conducted by the author at various locations for SABESP, Sao Paulo, Brazil, provided via direct personal communication

Wyatt, A. and M. Alshafey (2012). "Non-revenue water: financial model for optimal management in developing countries: application in Aqaba, Jordan." Water science and technology: water supply 12(4): 451-462.

Annex F: Component 4: Evaluation Design for the Water Smart Homes Activity

At the time of approval of the original Evaluation Design Report (EDR), there was insufficient information about the Water Smart Homes (WSH) Activity of the Jordan Compact to produce an evaluation design. It was determined then that an evaluation component could be proposed at a later date, but that the scope of this evaluation would have to be commensurate with the small scale of the investment relative to the main infrastructures. This annex describes such a design.

Description of the intervention. According to the design report from EcoConsult, the "Household Infrastructure and Knowledge Improvement Intervention" was originally conceived as targeting "household water practices to improve water potability and increase water efficiency at the household level." The Intervention was to achieve its objectives through an education and awareness program, and provision of grants for infrastructure replacement and support to the poor, in particular to families receiving grant from the National Aid Fund (NAF).

The communications strategy that was eventually implemented was developed following a needs assessment conducted by the implementing consultant Cowater, which helped to "segment target audiences, develop messages, and identify most appropriate and effective communication channels that would most likely lead to behavior change." The campaign focused on water quality and water conservation, and disseminated messages through interpersonal channels (religious "waethat", NAF, and women NGO networks; tribal "madafa" networks, and school education), and distribution of differentiated communication materials (interpersonal messages, posters/banners, news and social media channels, pamphlets, toolkits and water bottles, and educational videos).

The infrastructure program meanwhile selected recipients from a survey of roughly 11,000 NAF beneficiaries living in Zarqa governorate who were located using information in NAF enrollment lists and an intensive field campaign. Based on the results of these surveys, 5,198 households were deemed eligible for infrastructure support, and were ranked in priority based on a composite eligibility score that included factors related to baseline infrastructure conditions, land tenure, and household socio-economic variables. In the end, 3,958 of NAF households benefited from the improvements; some of these were newly enrolled based on the evolution of the NAF lists between the original field survey and the time of implementation.

A final component of the WSH intervention was the women plumbers training program, which was developed to generate employment and business opportunities for women in Zarqa Governorate, taking into account cultural, market and environmental constraints. Trainees were deemed eligible based on age (between 25 and 40 years), security and health clearance, minimum education (up to 9^{th} grade), and family acceptance of the program. Approximately seventy women applied to participate, and 30 were chosen following interviews. These women received basic vocational training to the level of semi-skilled plumber, and assistance in general small business management.

Evaluation questions. The following are key evaluation questions:

1. Did the WSH messaging campaign lead to significant changes in household water handling and storage, that manifested in a) improved water potability and b) improved water efficiency at the household level?
2. Did the WSH activity lead to infrastructure improvements among NAF beneficiaries?
3. Did NAF beneficiaries experience economic benefits from these infrastructure and behavioral changes, specifically through reduced coping costs, increased household productivity, and/or improved health and well-being?
4. Did the plumbers training portion of the WSH intervention lead to long term capacity and employment opportunities for the women who participated?

Overview of design elements. In addition to conducting a performance evaluation of the WSH program that would rely on reports and data collected at MCA-J, we considered three data collection elements:

1. Using the household survey sample to test for recall of messages disseminated through the various communication channels used by WSH;
2. Conducting qualitative work with women plumbers who were trained by the WSH program to deliver plumbing services, and with National Aid Fund (NAF) beneficiaries who received infrastructure improvement support; and
3. A quantitative data collection exercise (survey) designed around the discontinuity or eligibility cutoff for the WSH infrastructure support.

Ultimately, we recommended (and MCC approved) inclusion of the first two elements listed above. The third element was not recommended due to concerns about lack of statistical power (primarily) and our ability to construct an appropriate sample (secondary). As such, the data collection will not be sufficient to provide definitive causal evidence on the effects of this program: selection into the intervention, confounding, and other threats preclude clean identification of impacts.

Design details. This section presents the details of the proposed design.

Element 1: Recall of WSH messages. (Recommended and retained)
The WSH messaging campaign was aimed at promoting water conservation behaviors and investments in Zarqa, and targeted a variety of different communication channels, as discussed above. Given the breadth of the campaign, we decided to test whether households in the primary household survey sample for the impact evaluation recognized and made changes following exposure to such messages. Specifically, we will test for differences in recall of a random selection of WSH vs. other messages in Zarqa and Amman, and then ask households who remember seeing specific messages whether they have made changes as a result of this exposure (issues related to question 1 above). The other messages will include a collection of materials previously used in other campaigns in Jordan, as well as in other neighboring countries. This message recall will be folded into back-checks of the seasonal survey data with a random sub-sample of the overall household sample, and will be repeated with the complete sample at endline.

The analysis will then test for differences (using simple t-tests and multivariate models that correct for geographic clustering and control for potential confounders) in the recall and behavior changes induced between households in Zarqa who were exposed most directly to the messaging campaign, and households in Amman, who were not. While it is possible that spillovers would result in Amman and that such households would also recall the WSH messages, we would expect that:

- Recall of the specific WSH messages will be higher in Zarqa, where the campaign was centered, and
- Recall of the non-WSH messages should be similar across locations.

Data Collection	*Timing*	*Sample Unit/ Respondent*	*Sample Size*	*Relevant instruments/modules*	*Exposure Period (months)*
Seasonal 2 backchecks	11/2016-12/2016	Household/ Head of household or spouse	345	Backcheck survey/questions on recall of 4 promotional messages; water consumption behavior	Not applicable
Endline Household Survey	4/2018-6/2018	Household/ Head of household or spouse	3850 (estimated)	Household survey/ questions on recall of 4 promotional messages; water consumption behavior	Not applicable

Element 2: Qualitative work with women plumbers and NAF beneficiaries of the infrastructure improvements. (Recommended and retained)

We will conduct qualitative and semi-quantitative interviews with all (~30) women plumbers who completed the training under the WSH program, to collect basic demographic information about them, and to understand how participation in the program influenced their own employment opportunities, whether they are continuing to use and market their plumbing skills, how their customer networks are or are not evolving, and whether they are satisfied with the effects the program had on their own lives (issues related to question 4 above). In parallel, we will conduct focus groups and semi-quantitative interviews with 100 NAF beneficiaries of the infrastructure support program, to understand how this support improved their water supply situation and status, whether the effects were long-lasting, whether they led to measurable impacts on their well-being, and how (issues related to questions 2 and 3 above). We will also interview a comparison group of ~50 NAF beneficiaries who did not receive such improvements or attention, noting of course that these groups are not strictly comparable.

The purpose of these interviews is to better inform our understanding of the process and outcomes associated with the WSH infrastructure activity and women plumbers training and capacity building, to supplement the information already available in progress reports and in the infrastructure audits that were completed by the MCA-J.

NAF respondents will be invited randomly from the lists of NAF households constructed and shared by Cowater, and will be asked to participate in the focus groups / semi-structured interviews. Data will be collected by enumerators skilled in qualitative data collection, and interviews will be audio-recorded (subject to informed consent) for translation into English and further analysis. Reported results will also be cross-checked with information recorded in the Bill of Quantities for this intervention and the MCA-J infrastructure audit.

Data Collection	*Timing*	*Sample Unit/ Respondent*	*Sample Size*	*Relevant instruments/modules*	*Exposure Period (months)*
WSH qualitative work	2/2018	Female plumber (individual)	30	Qualitative and semi-quantitative interviews/demographics, participation in program, employment status, customer networks, satisfaction with training	Not applicable
WSH qualitative work	2/2018	NAF households/ head of household or spouse	150	Semi-quantitative interviews and focus groups/questions on WSH infrastructure support, trends in water supply and status, trends in well-being	Not applicable

Element 3: Survey with NAF beneficiaries on each side of the eligibility cutoff. (Not recommended and not retained)

We were able to collect data from the WSH needs assessment data, which made clear the eligibility criteria for selection into the program. Using these data, we determined that it would be possible to sample roughly 400 households evenly split on each side of the cutoff, which would maintain some similarity across groups and the possibility of using a regression discontinuity design.

Unfortunately, power calculations based on this feasible number did not suggest that this sample size would be sufficient for anything but the most basic changes (e.g., change in home storage volume and sanitation rating). These calculations revealed that detectable effect sizes for a number of key parameters would have to be very large (Table F1).

Table F1. Detectable effect sizes (at 10% significance and 80% power) for key variables with a sample of 400 NAF beneficiaries

Variable	**Detectable effect size**
Recall of WSH message	23%
Report of water conservation	93%
Leaks were fixed	69%
Storage containers cleaned more frequently	78%
Diarrheal disease prevalence	62%
Report water shortage	41%
New storage containers were obtained	258%
Change in storage volume	14%
Sanitation rating	9%

Given these calculations, we deemed it inadvisable to continue with a rigorous evaluation of the WSH program. This conclusion was only strengthened by our determination that a) the eligibility criteria had not been rigorously applied (replacement households were enrolled who did not satisfy the criteria or were not fully assessed against it); and b) the address information for many untreated NAF households was incomplete.

Challenges. The results of the WSH-related data collection and analyses will have to be interpreted with caution, given the nature of the proposed design. The lack of inclusion of a clear control group, especially for evaluation of the effects of plumber training and the infrastructure improvements, will preclude causal attribution to the WSH project. The evidence provided on these aspects of the program will thus largely be anecdotal. On the other hand, the use of the Amman sample for the testing of message recall does constitute a viable control group that is primarily threatened by spillovers and non-Compact confounders. It is plausible that households in Amman are more widely exposed to water saving messages for numerous reasons, including the more streamlined utility that operates there, the generally higher level of education of the population, and its greater exposure to international programs and donors. These differences could confound our measures of impact. Households in Amman could also be indirectly affected by the WSH campaign, through relatives living in Zarqa, their own travel there, or other mechanisms facilitating the diffusion of communications across Governorate boundaries. Finally, the sample construction for the qualitative work on NAF infrastructure beneficiaries may introduce bias, since the sample is small and participants will have to agree to participate in focus groups and data collection.

Timeline for these activities. As noted above, the first element will be conducted during seasonal survey back-checks and also during the endline household survey, in the summer of 2018. The qualitative work is planned for the second half of 2017. All survey instruments will undergo the usual Institutional Review Board (IRB) reviews, and participants' confidentiality and privacy will be maintained. A single report will be produced following analysis of the data

Annex G: Proposed evaluation items that were not approved for funding

At the time of presentation of midline results, several data collection activities were suggested by the evaluation team but not approved. Specifically, the proposal was to engage in two separate end line household surveys (in winter 2017 and summer; only summer 2018 was approved), to include two agricultural surveys in order to have complete data across years (only summer 2018 was approved; the equivalent activity in the summer of 2017 was not), and to conduct additional surveys (not approved) in a subsample of households in early 2017 to test for recall of Water Smart Homes (WSH) activity marketing messages. The rationale for each of these is described below:

- Two seasonal (winter 2017 and summer 2018) end line surveys for the household survey would offer two primary advantages: 1) It would allow us to correct and fully isolate changes in outcomes across both summer and winter seasons. Only the first seasonal round (2015) occurred during the winter; all other survey rounds have taken place in the summer. As a result, we will only be able to comment on changes in summer consumption with confidence. Acknowledging that water consumption and availability are highly seasonal, particularly in Jordan, we caution against extrapolation of changes in the summer over an annual basis, therefore, though we note that summer is the most critical season for water scarcity, and therefore likely to be the season during which impacts are most strongly felt by project beneficiaries.", 2) It would allow us to deliver information on water quality to a random subset of households to rigorously test how impacts are conditioned by erroneous perceptions of water quality. On the latter point, we noted at mid-line that households appeared to not have confidence in the quality of water delivered by the network, and that this might impede realization of the critical secondary substitution away from expensive non-network water sources. In discussions of this issue, the Evaluation Management Committee argued that doing this additional work through the survey was not appropriate.
- The summer 2017 agricultural survey would have allowed us to construct a panel of 8 consecutive farming seasons (since we ask about the prior summer and winter cultivation activities in each survey), which would have provided us with a very rich dataset through which to observe the evolving farming situation in the Jordan Valley and highlands along the Zarqa River. Instead, one year (two seasons) will be missing from the dataset, decreasing statistical power and our ability to account for interyear variability in weather and other growing conditions, and in the behavior of food markets.
- Finally, also based on mid-line results that pointed to very low recall (<10%) of WSH activity messages in our sample, we decided to not include a data collection task specific to this activity. We agreed to ask households about such messages again in the end line survey, though we expect even lower recall at that time due to the lapse in time since the intervention activities.

We also considered doing a more extensive survey of the NAF beneficiaries who received infrastructure support as part of the WSH activity, but did not recommend this option owing to a) the difficulty in constructing a viable counterfactual sample for this analysis; and b) the limitations in statistical power that a rigorous design (in our case a regression discontinuity design) would face, for measuring final outcomes like improved income and decreased household expenditures. See Annex F for additional details.