Impact evaluation of water infrastructure investments: Methods, challenges and demonstration from a large-scale urban improvement in Jordan

Marc A Jeuland[1,1], Jennifer Orgill-Meyer[2,2], Seth Morgan[1,1], Daniel Hudner[3,3], Mateusz Pucilowski[4,4], Alan Wyatt[5,5], Mohammed Shafei[6,6], James Cajka[5,5], and Jeff Albert[7,7]

[1]Duke University
[2]Franklin and Marshall
[3]Mercy Corps
[4]Social Impact
[5]RTI International
[6]Independent Consultant
[7]Aquaya Institute

January 20, 2023

Abstract

Impact evaluation (IE) of large infrastructure presents numerous challenges, and investments in urban piped water and sanitation are no exception. Here we present methods for more systematic assessment of the implications of such interventions, discussing tradeoffs between validity, relevance and practicality that arise from alternative approaches. Then, to more clearly illustrate the many issues that typically arise in such IEs, we draw on an example application in Zarqa, Jordan, where the Millennium Challenge Corporation invested about US$275 million to upgrade and extend piped water and sewer networks, as well as increase the capacity of the country's largest wastewater treatment plant. The theory of change for the intervention took a systems view of impacts: the project aimed to improve water supply to urban areas while maintaining flows to irrigators through enhanced wastewater reuse. The case adds valuable evidence on the impacts of large infrastructure investments and illustrates well the challenges of capturing spillovers, mitigating study contamination, maintaining statistical power, and determining overall welfare effects, in situations involving diverse market and nonmarket impacts. These limitations notwithstanding, the case highlights the high value of conducting IEs, and why applied researchers should not give up on pragmatic and interdisciplinary collaborations to evaluation in the face of complex interventions.

Tables and figures

Table 1. Summary of evaluation options, with focus on main internal validity threats, relevance, and practical considerations that are of particular importance for network water supply and sanitation

Method	Description and comments	Threats to validity of causal inference	Relevance of evaluation evidence	Practical / logistical considerations
Experimental				

Method	Description and comments	Threats to validity of causal inference	Relevance of evaluation evidence	Practical / logistical considerations
Randomized Controlled Trial (RCT) [*Duflo et al.*, 2007]	RCTs are generally not feasible for network water infrastructure, as such interventions are clustered, directional, and designed to serve population at scale or to address known (selected) system deficiencies. Some complementary interventions (information campaigns) can be evaluated using this approach. Smaller-scale rural infrastructure (e.g., condominial sewerage, village-scale piped water) can be evaluated with cluster RCTs, or step-wedge RCTs.	*Confounding* due to unbalanced randomization *Spillovers* (violation of the stable unit treatment value assumption, or SUTVA), whereby some units benefit as a result of other units' uptake. Vulnerable to *selective attrition*	Typically *artefactual*, w/ limited evaluation questions Treatment effect can be *representative* "Gold standard" for causal researchers Results are *not conditioned* by assumptions *Statistical power* is a design feature, but usually sufficient for a few pre-identified outcomes	*Cost*: High, especially when powered for multiple outcomes or interventions *Contamination risk*: Moderate, as pressure to help "untreated" units increases over time *Coordination*: Mainly pertains to maintaining integrity of randomization *Interpretation*: Intuitive and highly transparent *Pre-intervention data needs*: Low to none *Flexibility to adapt*: Very low
Experimental encouragement design [*Katz et al.*, 2001]	Subsidies or other assistance to customers can generate exogenous variation in the take-up of infrastructure connections, for use as an instrumental variable for isolating impacts. The resulting local average treatment effect is specific to those who respond to the encouragement [*Heckman et al.*, 2006].	Same as above	Same as above, except that the treatment effect only applies to the population that responds to the encouragement	
Quasi-experimental				

Method	Description and comments	Threats to validity of causal inference	Relevance of evaluation evidence	Practical / logistical considerations
Natural experiment [*J Angrist et al.*, 2002]	Some infrastructure placements are determined by geographic or other factors that are "as good as random" in determining exposure to improvements, such that they provide researchers with "natural experiments" [*Cerdá et al.*, 2012], that give rise to comparable treatment and control groups. Another version of this is an interrupted time series analysis where a time-dependent event (e.g., rehab of one part of a water network) gives rise to a sharp change that only affects some households or others.	*Confounding* by geographic / other factors determining exposure may also confound outcomes *Spillovers* (i.e., violation of SUTVA) outside of treatment area	Evidence arises directly from the *real world* Treatment effect is *representative* but contingent on natural experiment conditions Generally accepted by researchers Results are *not conditioned* by assumptions *Statistical power*: Difficult to anticipate ex ante	*Cost*: Low to moderate, depending on data collection needs *Contamination risk*: Low *Coordination*: Moderate; mainly in combining with other methods (DiD) to strengthen validity *Interpretation*: Intuitive but not always transparent *Pre-intervention data needs*: Low to none *Flexibility to adapt*: Impossible *Other*: Natural experiment can be hard to anticipate

Method	Description and comments	Threats to validity of causal inference	Relevance of evaluation evidence	Practical / logistical considerations
Difference-in-differences (DiD) [*Card and Krueger*, 2000]	In this approach, impacts are estimated by subtracting out the trend in an unexposed sample, which represents the counterfactual, from that in an exposed sample. Such samples are created using variation in spatial targeting or other eligibility criteria, which are common for network water infrastructure extension or rehabilitation. The validity of the comparison relies on pre-treatment trends being similar in the groups, and can be enhanced using matching or econometric models that control for differences in baseline covariates.	*Confounding by* time-varying unobservables *Spillovers* (i.e., violation of SUTVA) Vulnerable to *selective attrition*	Evidence arises directly from the *real world* Treatment effect is usually *representative* (unless combined w/other methods) Generally accepted by researchers, subject to showing parallel trends Results are *not conditioned* by assumptions *Statistical power* is a design feature	*Cost*: Moderate to high, depending on data collection needs *Contamination risk*: Moderate to high *Coordination*: Moderate; mainly in combining with other methods (matching) to strengthen validity *Interpretation*: Intuitive and transparent *Pre-intervention data needs*: Moderate to high (parallel trends) *Flexibility to adapt*: Moderate

Method	Description and comments	Threats to validity of causal inference	Relevance of evaluation evidence	Practical / logistical considerations
Matching or synthetic control [*Abadie and Gardeazabal*, 2003; *Rosenbaum and Rubin*, 1985]	These methods are best when combined with DiD analysis, but can be used to improve comparability when targeting is correlated with baseline characteristics. Various matching approaches enhance comparability by sampling untreated observations that can approximate the treatment counterfactual. For example, propensity score matching (PSM) finds treated and untreated observations that have a similar probability of being treated, from a regression of participation on observables. Synthetic control uses a time series of pre-intervention observations to "train" an algorithm that identifies weights for a pool of observations with similar counterfactual trends as one or more treated units.	*Confounding* by unobservables (Conditional Independence Assumption), worse when match quality is low *Spillovers* (i.e., violation of SUTVA)	Evidence arises directly from the *real world* Treatment effect only applies to units with suitable comparisons (common support region) Researchers are often skeptical that the CIA has been met Results are *conditioned* by assumptions of the matching algorithm *Statistical power* is a design feature	*Cost*: Moderate to high, depending on data collection needs *Contamination risk*: High *Coordination*: Moderate; mainly in combining with other methods (DiD) to strengthen validity *Interpretation*: Intuitive, but matching may lack transparency *Pre-intervention data needs*: Moderate (matching) *Flexibility to adapt*: Moderate

Method	**Description and comments**	**Threats to validity of causal inference**	**Relevance of evaluation evidence**	**Practical / logistical considerations**
Instrumental variables (IV) [*J D Angrist and Krueger*, 2001]	An instrumental variable is a factor that predicts exposure to or participation in an intervention, but that does not affect outcomes directly through channels other than that effect on participation. This creates exogenous variation in the intervention that can be leveraged to determine its impacts. The impact measure is a local average treatment effect that measures the effect of the intervention on those ("compliers") whose participation is affected by the instrument. Program placement rules or constraints may give rise to valid instruments.	*Confounding:* For many interventions and outcomes, there are few plausibly "exogenous" assignments of this type, at least in a statistical sense *Spillovers* (i.e., violation of SUTVA)	Evidence arises directly from the *real world* Treatment effect (LATE) is not representative, and not always for the most relevant population Researchers are often skeptical about exclusion restriction Results are *conditioned* by exogeneity assumptions *Statistical power* is often reduced by 2-stage estimation	*Cost*: Low to moderate, depending on data collection needs *Contamination risk*: Not applicable *Coordination*: Low *Interpretation*: Unintuitive, lacks transparency *Pre-intervention data needs*: Low *Flexibility to adapt*: High *Other*: Suitable IV may not exist

Method	Description and comments	Threats to validity of causal inference	Relevance of evaluation evidence	Practical / logistical considerations
Regression discontinuity (RD) [*Imbens and Lemieux*, 2008; *Thistlethwaite and Campbell*, 1960]	RD exploits discontinuities in eligibility for an intervention with respect to an assignment variable. For example, population thresholds, or a poverty line threshold for subsidy eligibility.	*Confounding*: Eligibility rule violations or manipulation, or "fuzzy" discontinuities that are difficult to characterize well *Spillovers* (i.e., violation of SUTVA) Vulnerable to *selective attrition*	Evidence arises directly from the *real world* Treatment effect is limited to units very near the discontinuity Generally accepted by researchers Results are *conditioned* on proximity to eligibility cutoff Statistical power may be limited	*Cost*: Low to moderate, depending on data collection needs *Contamination risk*: Moderate, depending on rigor with which eligibility is assessed *Coordination*: Low *Interpretation*: Intuitive, but transparency may be lacking due to definition of the RD bandwidth *Pre-intervention data needs*: Low *Flexibility to adapt*: Low
Other				
Ex post regression	Statistical comparison of treated and untreated units, with statistical control for observed differences between the groups. Also commonly called "observational" comparisons.	*Selection:* Units that participate are systematically different than those that do not *Confounding* by unobservables *Spillovers* (i.e., violation of SUTVA)	Evidence arises directly from the *real world* Treatment effect is usually representative Causal researchers are typically highly skeptical of results Results are *conditioned* on controls *Statistical power*: Difficult to anticipate ex ante	*Cost*: Low to moderate, depending on data collection needs *Contamination risk*: Not applicable *Coordination*: Low *Interpretation*: Intuitive, but transparency may be lacking (contingent on choice of controls) *Pre-intervention data needs*: None *Flexibility to adapt*: High

Method	Description and comments	Threats to validity of causal inference	Relevance of evaluation evidence	Practical / logistical considerations
Counterfactual modeling [*Balke and Pearl*, 2013]	Complex water resources systems evolve stochastically according to both human and environmental influences. This approach leverages systems understanding from socio-hydrological or hydro-economic models to conduct "with" and "without" simulations of interventions, for construction of model-based comparisons [*Srinivasan*, 2015].	*Confounding* by behavioral or other system-level factors not accounted for	Evidence is *artefactual*; model may diverge from real world observations Treatment effect is usually representative, but may not align with policy-maker priorities and needs Not widely used by causal social science researchers, who are wary of over-calibration Results are *conditioned* on model assumptions *Statistical power*: Not applicable	*Cost*: Low *Contamination risk*: Not applicable *Coordination*: Low *Interpretation*: Not intuitive and not always transparent (requires interdisciplinary expertise) *Pre-intervention data needs*: Moderate to high, depending on calibration needs *Flexibility to adapt*: High *Other*: Required model effort is substantial

Table 2. Summary of study populations and data collection methods deployed

Survey element	Survey type	Sampling frame	Sample selection	Stratification / comparison group	Representation	Sample size
Household	4-wave Panel	Zarqa and Amman (from Jordan Dept. of Statistics (DoS))	Survey geocodes selected based on *ex ante* matching of treated and control zones, using Census data Random sampling within sample geocodes Replacements selected from sample geocodes	WNP only – WNP only control WWNP only – WWNP only control WNP+WWNP – WNP+WWNP control Distinct control groups from: Zarqa Amman	Representative of sample geocodes at baseline, based on comparisons to Census and other sources	3359 3416 3596 3662

Survey element	Survey type	Sampling frame	Sample selection	Stratification / comparison group	Representation	Sample size
Enterprise	2-wave Panel	Zarqa and Amman (from DoS + household referrals)	Same geocodes as household sample Random selection within sample geocodes Referrals for informal enterprises Replacements selected from closely neighboring enterprises	Same as household (though analysis uses all controls for each group to maximize statistical power)	Representative of sample geocodes at baseline Informal enterprises likely under-represented (due to low referral rates)	345 418
Farm	3-wave Panel	Jordan Valley and highlands (from DoS)	Survey zones selected based on expected differences in exposure to treated wastewater Random selection in sample zones Replacements selected within zones	Five locations: Highlands u/s KTD (— river flow) JV1 North (— Non-Compact WW) JV2 Mid-North (— Compact WW) JV3 North-Central (— Compact WW) JV4 South-Central (little change in WW)	Representative of sample zones	551 539 539

Survey element	Survey type	Sampling frame	Sample selection	Stratification / comparison group	Representation	Sample size
Refugee	Single cross-section	UNHCR registration list for Zarqa and Amman	Priority survey geocodes selected according to household sample, with augmenting based on treatment status outside hh geocodes Random sampling by treatment status Referrals for unregistered refugees	Treatment status: WNP only WWNP only Both WNP and WWNP Controls in Zarqa Control Amman	Representative of registered population in sample areas Unregistered population likely under-represented (due to low referral rates)	1617
Water vendor	Single cross-section	Shops: Ministry of Health list + canvassing Tankers: Canvassing	Full sampling from canvassed locations	None	Representative of water vendors in Zarqa and East Amman in 2018	320
Meter testing	Repeat cross-section	Meter listing in selected zones	Zones selected for variation in JC status, elevation, pressure, throughput (for survey 1), and JC status and meter replacement (for survey 2) Random sample of meters within selected zones	Compact and non-Compact zones	Not representative	37 223

Survey element	Survey type	Sampling frame	Sample selection	Stratification / comparison group	Representation	Sample size
Water loss testing	Single cross-section	Canvassing of land plots in selected areas	"Well isolated" zones selected (as suggested by utility) Comparison of meter registered data to bulk meter inflow Random sub-sample of meters evaluated to adjust for meter error	Meter error testing sub-sample stratified by meter replacement status	Not representative; only relevant to "well isolated" zones	1797
Key informant interviews	Single cross-section	Listing of key JC stakeholders	Contact to all listed stakeholders Replacements included as suggested by stakeholders	None	Representative of institutions, but likely not all perspectives	22

Table 3. Summary of main impacts on household behaviors and outcomes

Outcome	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample
	(1) WNP	(2) WWNP	(3) Both	(4) WNP	(5) WWNP	(6) Both
Water supply						
Reported water pressure rating[1]	**-0.38***** (0.15)	n.a.	**-0.49***** (0.14)	**-0.63***** (0.12)	n.a.	**-0.84***** (0.14)
Reported perception of network water quality	**+0.63**** (0.30)	n.a.	-0.29 (0.34)	+0.32 (0.27)	n.a.	-0.58 (0.36)

Outcome	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample
Assessed water quality (E. coli count)[2]	-0.049 (0.053)	n.a.	0 (0)	n.a.	n.a.	n.a.
Hours piped water, for days w/water	+0.86 (0.61)	n.a.	+0.37 (0.64)	**+1.15**** (0.46)	n.a.	**+1.83***** (0.47)
Reported water shortage, past month	**-0.10*** (0.05)	n.a.	-0.05 (0.06)	**-0.12***** (0.04)	n.a.	**-0.098*** (0.05)
Network water use – Utility sample[3]	**+2.9***** (0.66)	n.a.	**+2.9*** (1.5)	+0.52 (0.59)	n.a.	**+1.9*** (1.1)
Network water use – Survey sample		n.a.	+6.1 (3.9)	+5.1 (3.5)	n.a.	+3.7 (4.3)
Expenditure on water from vendors (JD/month)	-5.1 (5.0)	n.a.	-7.1 (5.5)	-6.8 (4.4)	n.a.	**-9.7*** (5.8)
Expenditure on water, all sources (JD/month)	-3.6 (5.8)	n.a.	-6.2 (5.7)	-4.0 (5.1)	n.a.	-10.6 (6.3)
Wastewater management						
Use of stand-alone cesspits	n.a.	**-0.13***** (0.04)	**-0.07*** (0.04)	n.a.	**-0.14***** (0.05)	**-0.11*** (0.04)
Sewer connection	n.a.	**+0.14***** (0.05)	**+0.17***** (0.05)	n.a.	**+0.09*** (0.05)	**+0.12**** (0.05)
Expense for septic tank evacuation	n.a.	-1.0 (4.7)	-15.1 (11.3)	n.a.	-7.7 (5.6)	-18.2 (11.2)
Sewer backup prevalence	n.a.	-0.02 (0.01)	-0.004 (0.01)	n.a.	**-0.02** (0.01)**	-0.002 (0.01)
Overall welfare						
Expenditure (JD/month)	-2.2 (32.6)	+30.9 (39.5)	+15.6 (40.8)	+5.3 (32.8)	**+76.8*** (39.0)	-4.0 (47.1)

Outcome	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample
Net income	-22.2 (29.2)	-5.7 (27.1)	-8.8 (33.7)	**-66.6*** (38.7)	-42.1 (35.8)	**-131***** (46.6)
Assets	+0.03 (0.03)	+0.04 (0.04)	+0.04 (0.04)	+0.04 (0.03)	+0.02 (0.04)	-0.03 (0.04)
Sample size for comparison	1,914	1,443	1,389	2,359	1,559	1,418

Notes: All estimates are difference-in-differences estimates for coefficient $\varkappa_t$ for period t=1 (after the intervention). Standard errors are shown in parentheses. Statistical significance is denoted as follows: *** p<0.01; ** p<0.05; * p<0.1. The specification controls for time and household fixed effects, and includes additional time-varying controls for the number of refugees arriving in the sample area (a demand shock) and a household-specific wealth index yield very similar results (for alternative results omitting the controls and fixed effects, see Appendix 3). The subsample comparisons are as follows: WNP – Water Network Project treatment zones and matched control zones; WWNP – Wastewater Network Project treatment zones and matched control zones; Both – Water and Wastewater Network Project treatment zones and matched control zones.

[1] Measured on a four point scale (1 = excellent; 4 = poor)

[2] Water samples were only collected and analyzed in Zarqa

[3] The regressions for this outcome do not control for the time-varying factors because we use the full utility database, rather than restricting to the survey sample.

Table 4. Summary of main impacts on small enterprise behaviors and outcomes

Outcome	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample
	(1) WNP	(2) WWNP	(3) Both	(4) WNP	(5) WWNP	(6) Both
Water supply						

Outcome	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample
Piped water is primary source	-0.05 (0.10)	n.a.	-0.01 (0.10)	-0.07 (0.11)	n.a.	+0.09 (0.12)
Hours piped water, for days w/water	-1.8 (2.5)	n.a.	0.25 (2.5)	**-8.1***** (3.0)	n.a.	-3.7 (3.0)
Water consumption (m^3/month)	+16.9 (16.8)	n.a.	+24.7 (19.2)	+12.5 (16.5)	n.a.	+25.9 (21.1)
Reported water interruption	+0.03 (0.11)	n.a.	-0.03 (0.12)	+0.03 (0.14)	n.a.	+0.05 (0.15)
Expenditure on water from vendors (JD/month)	+0.21 (0.46)	n.a.	-0.35 (0.49)	-0.54 (0.54)	n.a.	**-1.50**** (0.61)
Expenditure on water, all sources (arcsin, JD/month)	+0.22 (0.28)	n.a.	-0.02 (0.31)	**-0.57*** (0.33)	n.a.	-0.01 (0.34)
Wastewater management						
Use of some wastewater system	n.a.	-0.12 (0.08)	+0.02 (0.08)	n.a.	-0.16 (0.10)	+0.05 (0.08)
Sewer connection	n.a.	**-0.18**** (0.09)	-0.09 (0.09)	n.a.	-0.11 (0.10)	+0.02 (0.09)
Cost of wastewater management (arcsin, JD/month)	n.a.	-0.22 (0.32)	**-0.46*** (0.26)	n.a.	+0.05 (0.41)	-0.04 (0.37)
Overall welfare						
Expenditure (arcsin, JD/month)	**-0.46***** (0.15)	-0.02 (0.15)	**+0.36*** (0.18)	**-0.34**** (0.17)	+0.07 (0.18)	**-0.39*** (0.21)
Asset value (arcsin, JD)	+0.25 (0.33)	+0.09 (0.36)	**-0.77***** (0.28)	+0.24 (0.30)	-0.16 (0.31)	-0.31 (0.31)

Outcome	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample
Land value (arcsin, JD)	+0.57 (0.35)	+0.28 (0.45)	-0.50 (0.37)	+0.60 (0.43)	+0.41 (0.53)	-0.50 (0.34)
Sample size for comparison	246	229	216	156	139	239

Notes: All estimates are difference-in-differences estimates for coefficient $\varkappa_t$ for period t=1 (after the intervention). Standard errors are shown in parentheses. Statistical significance is denoted as follows: *** p<0.01; ** p<0.05; * p<0.1. The specification controls for time and enterprise fixed effects, as well as the following other time-varying factors: reported complaints about sewer overflows; respondent years with enterprise, number of total employees, reported obstacles to growth, and frequency of water interruptions. Alternative specifications without controls yield very similar results (see Appendix C). The subsample comparisons are as follows: WNP – Water Network Project treatment zones and matched control zones; WWNP – Wastewater Network Project treatment zones and matched control zones; Both – Water and Wastewater Network Project treatment zones and matched control zones.

Table 5. Summary of main impacts on farm behaviors and outcomes

Outcome	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample
	(1) JV1	(2) JV2 (Treatment)	(3) JV3 (Treatment)	(4) JV4	(5) Highlands (Treatment)
Wastewater use in irrigation	**+0.10**** (0.04)	**+0.14***** (0.04)	**+0.15***** (0.04)	**-0.47**** (0.04)	**+0.08*** (0.04)
Perceived water quality[1]	**-0.81***** (0.3)	**-0.53*** (0.31)	-0.14 (0.31)	**+1.40***** (0.31)	0.03 (0.34)
Irrigated area (dunum)	**+6.0*** (3.4)	**+8.4**** (3.5)	**-9.0**** (3.6)	**-19.9***** (3.5)	**+12.5***** (3.8)
Farm revenue (JD/yr)	**+91823*** (48757)	-60459 (51173)	-58896 (51336)	**-91772*** (51908)	**+186422***** (55180)
Farm profit (JD/yr)	+67362 (47385)	-64974 (49696.95)	-46449 (49854)	-50554 (50473)	**+153102***** (53656)

Outcome	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample
Farm land value (JD)	**+62124***** (18012)	+25717 (19918)	**-61272***** (19275)	-6484 (21277)	**-42200*** (22611)

Notes: All estimates are difference-in-differences estimates for coefficient $\varkappa_t$ for period t=1 (after the intervention). Standard errors are shown in parentheses. Statistical significance is denoted as follows: *** p<0.01; ** p<0.05; * p<0.1. The specification controls for time and farm fixed effects. The subsamples are as follows: JV1 is furthest north in the Jordan Valley, and represents a set of farms that were mostly unaffected by the Compact since their water supply is independent of the Zarqa system; JV2 and JV3 represent areas where flows of recycled wastewater newly arrived (JV2) and increased substantially (JV3); JV4 represents an area that already had substantial flows of recycled water prior to the investment; Highlands farms, finally, are located along the Zarqa River and also received access to more steady water supply.

[1] Measured on a ten point scale (1 = poor; 10 = excellent)

Table 6. Summary of utility performance indicators, relative to other urban utilities in Jordan

Result	Indicator	Utility
Reduced NRW	Indicator: Total losses	Indicator: Total losses
	Total losses; network (%)	Aqaba
		Amman
		Zarqa
	Total losses; network (L / Subscriber / Day)	Aqaba
		Amman
		Zarqa
	Indicator: Pipe breaks/bursts per km of mainlines	Indicator: Pipe breaks/bursts per k
	Main bursts/ 100km	Aqaba
		Amman
		Zarqa
	Service leaks / 1000 connections	Aqaba
		Amman
		Zarqa
Increased revenue to utility	Utility revenue (2015 JD/m^3 sold)	Aqaba
		Amman
		Zarqa
Increased cost recovery by utility	Billing efficiency (%)	Aqaba
		Amman
		Zarqa
	Collection efficiency (%)	Aqaba
		Amman
		Zarqa
	Operating Cost Recovery Ratio (OCRR)	Aqaba
		Amman

Result	Indicator	Utility
		Zarqa

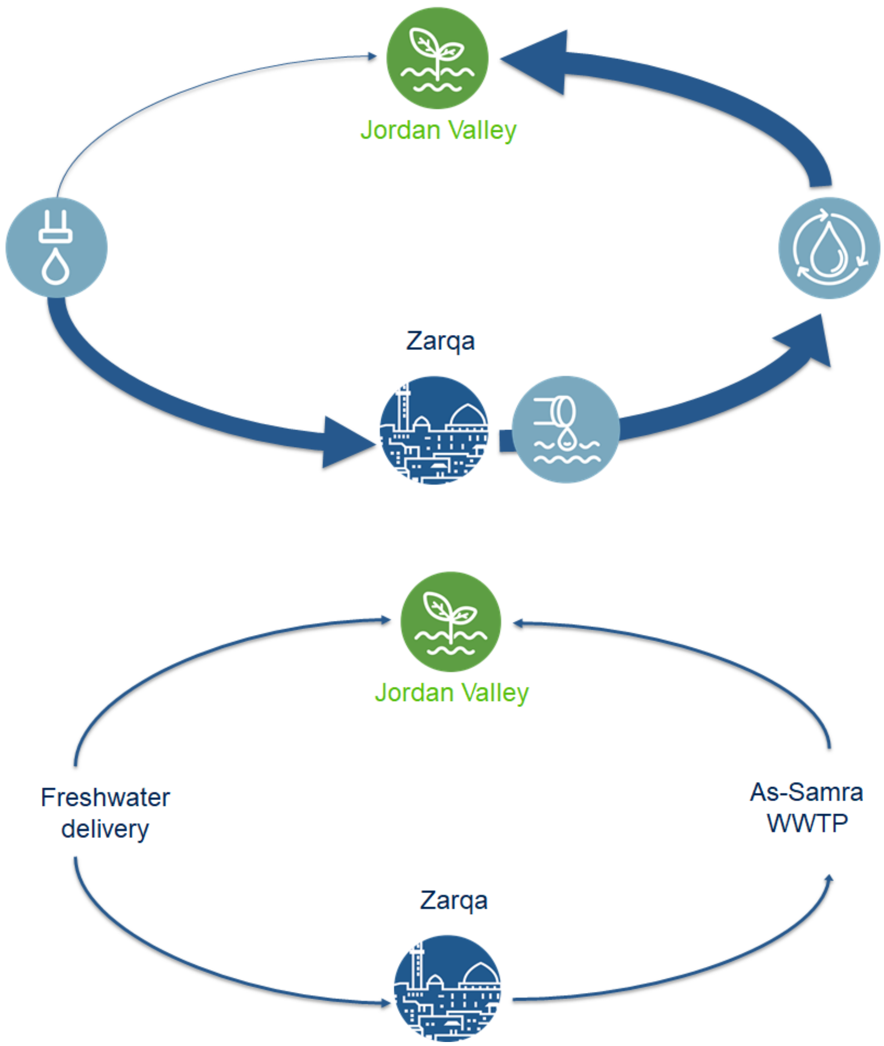


Figure 1. Qualitative depiction of (Top) pre- and (Bottom) ex pected post- Jordan Compact situations

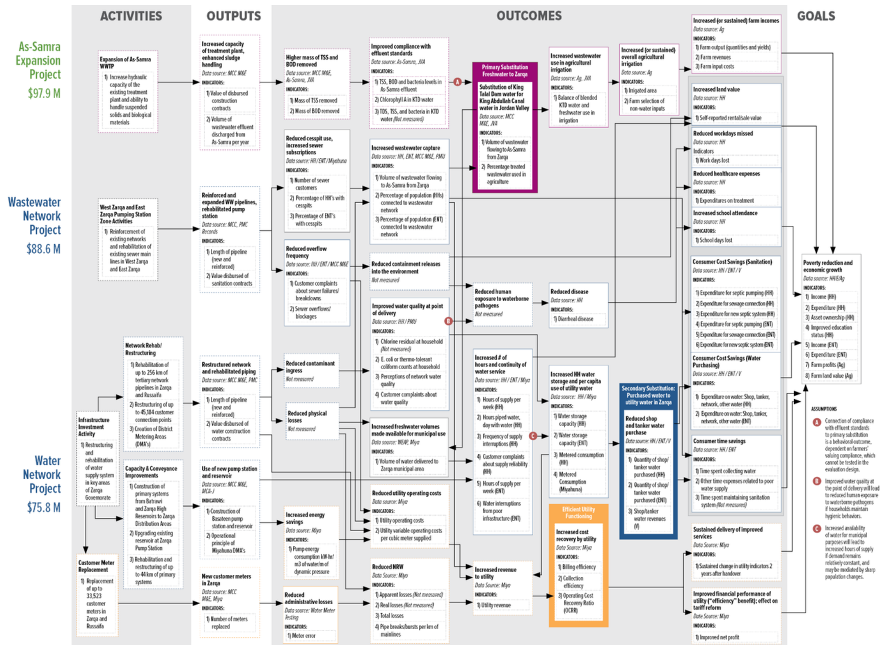


Figure 2. Full program theory of change, as elicited through participatory stakeholder consultations

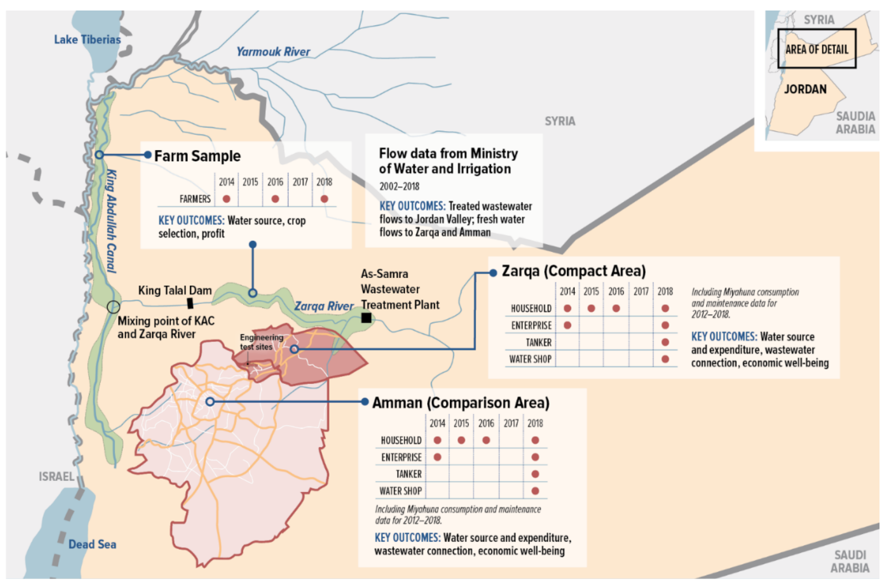


Figure 3. Locations and timing of data collection activities to support the evaluation

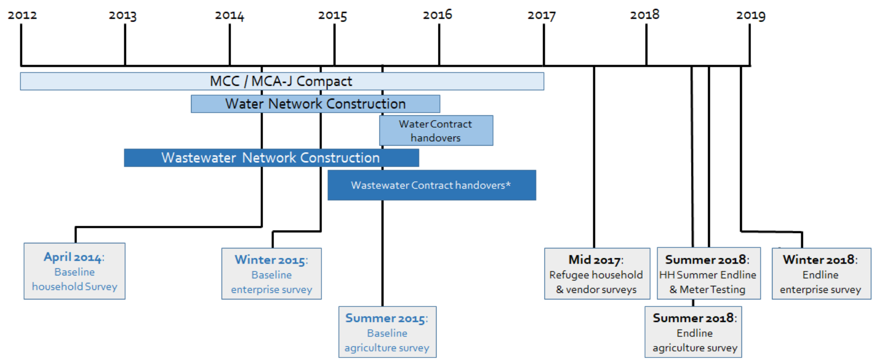


Figure 4. Timing of data collection relative to infrastructure intervention [Note that baseline surveys were conducted prior to any infrastructure operations, except for a few small wastewater Contract handovers preceding the baseline agriculture survey]

Tables and figures

Table 1. Summary of evaluation options, with focus on main internal validity threats, relevance, and practical considerations that are of particular importance for network water supply and sanitation

@ >p(- 8) * >p(- 8) * >p(- 8) * >p(- 8) * >p(- 8) * @ **Method** & **Description and comments** & **Threats to validity of causal inference** & **Relevance of evaluation evidence** & **Practical / logistical considerations**
Experimental & & & &
Randomized Controlled Trial (RCT) [*Duflo et al.*, 2007] & RCTs are generally not feasible for network water infrastructure, as such interventions are clustered, directional, and designed to serve population at scale or to address known (selected) system deficiencies. Some complementary interventions (information campaigns) can be evaluated using this approach. Smaller-scale rural infrastructure (e.g., condominial sewerage, village-scale piped water) can be evaluated with cluster RCTs, or step-wedge RCTs. &

- *Confounding* due to unbalanced randomization
- *Spillovers* (violation of the stable unit treatment value assumption, or SUTVA), whereby some units benefit as a result of other units' uptake.
- Vulnerable to *selective attrition*

&

- Typically *artefactual*, w/ limited evaluation questions
- Treatment effect can be *representative*
- "Gold standard" for causal researchers
- Results are *not conditioned* by assumptions
- *Statistical power* is a design feature, but usually sufficient for a few pre-identified outcomes

&

- *Cost*: High, especially when powered for multiple outcomes or interventions
- *Contamination risk*: Moderate, as pressure to help "untreated" units increases over time
- *Coordination*: Mainly pertains to maintaining integrity of randomization
- *Interpretation*: Intuitive and highly transparent
- *Pre-intervention data needs*: Low to none

- *Flexibility to adapt*: Very low

Experimental encouragement design [*Katz et al.*, 2001] & Subsidies or other assistance to customers can generate exogenous variation in the take-up of infrastructure connections, for use as an instrumental variable for isolating impacts. The resulting local average treatment effect is specific to those who respond to the encouragement [*Heckman et al.*, 2006]. &

- Same as above

&

- Same as above, except that the treatment effect only applies to the population that responds to the encouragement

&

Quasi-experimental & & & &

Natural experiment [*J Angrist et al.*, 2002] & Some infrastructure placements are determined by geographic or other factors that are "as good as random" in determining exposure to improvements, such that they provide researchers with "natural experiments" [*Cerdá et al.*, 2012], that give rise to comparable treatment and control groups. Another version of this is an interrupted time series analysis where a time-dependent event (e.g., rehab of one part of a water network) gives rise to a sharp change that only affects some households or others. &

- *Confounding* by geographic / other factors determining exposure may also confound outcomes
- *Spillovers* (i.e., violation of SUTVA) outside of treatment area

&

- Evidence arises directly from the *real world*
- Treatment effect is *representative* but contingent on natural experiment conditions
- Generally accepted by researchers
- Results are *not conditioned* by assumptions
- *Statistical power*: Difficult to anticipate ex ante

&

- *Cost*: Low to moderate, depending on data collection needs
- *Contamination risk*: Low
- *Coordination*: Moderate; mainly in combining with other methods (DiD) to strengthen validity
- *Interpretation*: Intuitive but not always transparent

- *Pre-intervention data needs*: Low to none
- *Flexibility to adapt*: Impossible
- *Other*: Natural experiment can be hard to anticipate

Difference-in-differences (DiD) [*Card and Krueger*, 2000] & In this approach, impacts are estimated by subtracting out the trend in an unexposed sample, which represents the counterfactual, from that in an exposed sample. Such samples are created using variation in spatial targeting or other eligibility criteria, which are common for network water infrastructure extension or rehabilitation. The validity of the comparison relies on pre-treatment trends being similar in the groups, and can be enhanced using matching or econometric models that control for differences in baseline covariates. &

- *Confounding by* time-varying unobservables
- *Spillovers* (i.e., violation of SUTVA)
- Vulnerable to *selective attrition*

&

- Evidence arises directly from the *real world*
- Treatment effect is usually *representative* (unless combined w/other methods)
- Generally accepted by researchers, subject to showing parallel trends
- Results are *not conditioned* by assumptions
- *Statistical power* is a design feature

&

- *Cost*: Moderate to high, depending on data collection needs
- *Contamination risk*: Moderate to high
- *Coordination*: Moderate; mainly in combining with other methods (matching) to strengthen validity
- *Interpretation*: Intuitive and transparent
- *Pre-intervention data needs*: Moderate to high (parallel trends)
- *Flexibility to adapt*: Moderate

Matching or synthetic control [*Abadie and Gardeazabal*, 2003; *Rosenbaum and Rubin*, 1985] & These methods are best when combined with DiD analysis, but can be used to improve comparability when targeting is correlated with baseline characteristics. Various matching approaches enhance comparability by sampling untreated observations that can approximate the treatment counterfactual. For example, propensity score matching (PSM) finds treated and untreated observations that have a similar probability of being treated, from a regression of participation on observables. Synthetic control uses a time series of pre-intervention observations to "train" an algorithm that identifies weights for a pool of observations with similar counterfactual trends as one or more treated units. &

- *Confounding* by unobservables (Conditional Independence Assumption), worse when match quality is low
- *Spillovers* (i.e., violation of SUTVA)

&

- Evidence arises directly from the *real world*
- Treatment effect only applies to units with suitable comparisons (common support region)
- Researchers are often skeptical that the CIA has been met
- Results are *conditioned* by assumptions of the matching algorithm
- *Statistical power* is a design feature

&

- *Cost*: Moderate to high, depending on data collection needs
- *Contamination risk*: High
- *Coordination*: Moderate; mainly in combining with other methods (DiD) to strengthen validity
- *Interpretation*: Intuitive, but matching may lack transparency
- *Pre-intervention data needs*: Moderate (matching)
- *Flexibility to adapt*: Moderate

Instrumental variables (IV) [*J D Angrist and Krueger*, 2001] & An instrumental variable is a factor that predicts exposure to or participation in an intervention, but that does not affect outcomes directly through channels other than that effect on participation. This creates exogenous variation in the intervention that can be leveraged to determine its impacts. The impact measure is a local average treatment effect that measures the effect of the intervention on those ("compliers") whose participation is affected by the instrument. Program placement rules or constraints may give rise to valid instruments. &

- *Confounding:* For many interventions and outcomes, there are few plausibly "exogenous" assignments of this type, at least in a statistical sense
- *Spillovers* (i.e., violation of SUTVA)

&

- Evidence arises directly from the *real world*
- Treatment effect (LATE) is not representative, and not always for the most relevant population
- Researchers are often skeptical about exclusion restriction
- Results are *conditioned* by exogeneity assumptions
- *Statistical power* is often reduced by 2-stage estimation

&

- *Cost*: Low to moderate, depending on data collection needs
- *Contamination risk*: Not applicable
- *Coordination*: Low
- *Interpretation*: Unintuitive, lacks transparency
- *Pre-intervention data needs*: Low
- *Flexibility to adapt*: High
- *Other*: Suitable IV may not exist

Regression discontinuity (RD) [*Imbens and Lemieux*, 2008; *Thistlethwaite and Campbell*, 1960] & RD exploits discontinuities in eligibility for an intervention with respect to an assignment variable. For example, population thresholds, or a poverty line threshold for subsidy eligibility. &

- *Confounding*: Eligibility rule violations or manipulation, or "fuzzy" discontinuities that are difficult to characterize well
- *Spillovers* (i.e., violation of SUTVA)
- Vulnerable to *selective attrition*

&

- Evidence arises directly from the *real world*
- Treatment effect is limited to units very near the discontinuity
- Generally accepted by researchers
- Results are *conditioned* on proximity to eligibility cutoff

- Statistical power may be limited

&

- *Cost*: Low to moderate, depending on data collection needs
- *Contamination risk*: Moderate, depending on rigor with which eligibility is assessed
- *Coordination*: Low
- *Interpretation*: Intuitive, but transparency may be lacking due to definition of the RD bandwidth
- *Pre-intervention data needs*: Low
- *Flexibility to adapt*: Low

Other & & & &
Ex post regression & Statistical comparison of treated and untreated units, with statistical control for observed differences between the groups. Also commonly called "observational" comparisons. &

- *Selection:* Units that participate are systematically different than those that do not
- *Confounding* by unobservables
- *Spillovers* (i.e., violation of SUTVA)

&

- Evidence arises directly from the *real world*
- Treatment effect is usually representative
- Causal researchers are typically highly skeptical of results
- Results are *conditioned* on controls
- *Statistical power*: Difficult to anticipate ex ante

&

- *Cost*: Low to moderate, depending on data collection needs
- *Contamination risk*: Not applicable
- *Coordination*: Low
- *Interpretation*: Intuitive, but transparency may be lacking (contingent on choice of controls)
- *Pre-intervention data needs*: None
- *Flexibility to adapt*: High

Counterfactual modeling [*Balke and Pearl*, 2013] & Complex water resources systems evolve stochastically according to both human and environmental influences. This approach leverages systems understanding from socio-hydrological or hydro-economic models to conduct "with" and "without" simulations of interventions, for construction of model-based comparisons [*Srinivasan*, 2015]. &

- *Confounding* by behavioral or other system-level factors not accounted for

&

- Evidence is *artefactual*; model may diverge from real world observations
- Treatment effect is usually representative, but may not align with policy-maker priorities and needs
- Not widely used by causal social science researchers, who are wary of over-calibration
- Results are *conditioned* on model assumptions
- *Statistical power*: Not applicable

&

- *Cost*: Low
- *Contamination risk*: Not applicable
- *Coordination*: Low
- *Interpretation*: Not intuitive and not always transparent (requires interdisciplinary expertise)
- *Pre-intervention data needs*: Moderate to high, depending on calibration needs
- *Flexibility to adapt*: High
- *Other*: Required model effort is substantial

Table 2. Summary of study populations and data collection methods deployed

@ >p(- 12) * >p(- 12) * >p(- 12) * >p(- 12) * >p(- 12) * >p(- 12) * >p(- 12) * @ **Survey element** & **Survey type** & **Sampling frame** & **Sample selection** & **Stratification / comparison group** & **Representation** & **Sample size**

Household & 4-wave Panel & Zarqa and Amman (from Jordan Dept. of Statistics (DoS)) &

- Survey geocodes selected based on *ex ante* matching of treated and control zones, using Census data
- Random sampling within sample geocodes
- Replacements selected from sample geocodes

&

WNP only – WNP only control

WWNP only – WWNP only control

WNP+WWNP – WNP+WWNP control

Distinct control groups from:

- Zarqa
- Amman

& Representative of sample geocodes at baseline, based on comparisons to Census and other sources &

1. 3359
2. 3416
3. 3596
4. 3662

Enterprise & 2-wave Panel & Zarqa and Amman (from DoS + household referrals) &

- Same geocodes as household sample
- Random selection within sample geocodes
- Referrals for informal enterprises
- Replacements selected from closely neighboring enterprises

& Same as household (though analysis uses all controls for each group to maximize statistical power) & Representative of sample geocodes at baseline

Informal enterprises likely under-represented (due to low referral rates) &

1. 345
2. 418

Farm & 3-wave Panel & Jordan Valley and highlands (from DoS) &

- Survey zones selected based on expected differences in exposure to treated wastewater

- Random selection in sample zones
- Replacements selected within zones

&

Five locations:

- Highlands u/s KTD (↑ river flow)
- JV1 North (↑ Non-Compact WW)
- JV2 Mid-North (↑ Compact WW)
- JV3 North-Central (↑ Compact WW)
- JV4 South-Central (little change in WW)

& Representative of sample zones &

1. 551
2. 539
3. 539

Refugee & Single cross-section & UNHCR registration list for Zarqa and Amman &

- Priority survey geocodes selected according to household sample, with augmenting based on treatment status outside hh geocodes
- Random sampling by treatment status
- Referrals for unregistered refugees

&

Treatment status:

- WNP only
- WWNP only
- Both WNP and WWNP
- Controls in Zarqa
- Control Amman

& Representative of registered population in sample areas

Unregistered population likely under-represented (due to low referral rates) & 1617
Water vendor & Single cross-section & Shops: Ministry of Health list + canvassing

Tankers: Canvassing &

- Full sampling from canvassed locations

& None & Representative of water vendors in Zarqa and East Amman in 2018 & 320
Meter testing & Repeat cross-section & Meter listing in selected zones &

- Zones selected for variation in JC status, elevation, pressure, throughput (for survey 1), and JC status and meter replacement (for survey 2)
- Random sample of meters within selected zones

& Compact and non-Compact zones & Not representative &

1. 37
2. 223

Water loss testing & Single cross-section & Canvassing of land plots in selected areas &

- “Well isolated” zones selected (as suggested by utility)
- Comparison of meter registered data to bulk meter inflow
- Random sub-sample of meters evaluated to adjust for meter error

& Meter error testing sub-sample stratified by meter replacement status & Not representative; only relevant to “well isolated” zones & 1797
Key informant interviews & Single cross-section & Listing of key JC stakeholders &

- Contact to all listed stakeholders
- Replacements included as suggested by stakeholders

& None & Representative of institutions, but likely not all perspectives & 22

Table 3. Summary of main impacts on household behaviors and outcomes

Outcome	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample				
	(1) WNP	(2) WWNP	(3) Both	(4) WNP	(5) WWNP	(6) Both
Water supply						
Reported water pressure rating[1]	-0.38*** (0.15)	n.a.	-0.49*** (0.14)	-0.63*** (0.12)	n.a.	-0.84*** (0.14)
Reported perception of network water quality	n.a.			n.a.	(0.36)	
Assessed water quality (E. coli count)[2]	(0.053)	n.a.	(0)	n.a.	n.a.	n.a.
Hours piped water, for days w/water		n.a.	+0.37 (0.64)	+1.15** (0.46)	n.a.	+1.83*** (0.47)
Reported water shortage, past month	-0.10* (0.05)	n.a.		-0.12*** (0.04)	n.a.	

Outcome	DiD impact of intervention – relative to non-intervention areas in Zarqa, by sub-sample	DiD impact of intervention – relative to non-intervention areas in Amman, by sub-sample				
Network water use – Utility sample[3]	**+2.9***** (0.66)	n.a.	**+2.9*** (1.5)	+0.52 (0.59)	n.a.	**+1.9*** (1.1)
Network water use – Survey sample		n.a.	+6.1 (3.9)	+5.1 (3.5)	n.a.	+3.7 (4.3)
Expenditure on water from vendors (JD/month)	(5.0)	n.a.	(5.5)	(4.4)	n.a.	**-9.7*** (5.8)
Expenditure on water, all sources (JD/month)	(5.8)	n.a.	(5.7)	(5.1)	n.a.	(6.3)
Wastewater management						
Use of stand-alone cesspits	n.a.	**-0.13***** (0.04)	**-0.07*** (0.04)	n.a.	**-0.14***** (0.05)	**-0.11*** (0.04)
Sewer connection	n.a.	**+0.14***** (0.05)	**+0.17***** (0.05)	n.a.	**+0.09*** (0.05)	**+0.12**** (0.05)

Outcome	DiD impact of intervention – relative to non-intervention areas in Zarqa, by sub-sample	DiD impact of intervention – relative to non-intervention areas in Amman, by sub-sample				
Expense for septic tank evacuation	n.a.			n.a.		
Sewer backup prevalence	n.a.	(0.01)	(0.01)	n.a.	**-0.02** (0.01)**	(0.01)
<u>Overall welfare</u>						
Expenditure (JD/month)	(32.6)	+30.9 (39.5)	+15.6 (40.8)	+5.3 (32.8)	**+76.8*** (39.0)	(47.1)
Net income	(29.2)	(27.1)	(33.7)	**-66.6*** (38.7)	(35.8)	**-131***** (46.6)
Assets	+0.03 (0.03)	+0.04 (0.04)	+0.04 (0.04)	+0.04 (0.03)	+0.02 (0.04)	(0.04)
Sample size for comparison	,914	,443	,389	,359	,559	,418

Notes: All estimates are difference-in-differences estimates for coefficient $_t$ for period t=1 (after the intervention). Standard errors are shown in parentheses. Statistical significance is denoted as follows: *** p<0.01; ** p<0.05; * p<0.1. The specification controls for time and household fixed effects, and includes additional time-varying controls for the number of refugees arriving in the sample area (a demand shock) and a household-specific wealth index yield very similar results (for alternative results omitting the controls and fixed effects, see Appendix 3). The subsample comparisons are as follows: WNP – Water Network Project treatment zones and matched control zones; WWNP – Wastewater Network Project treatment zones and matched control zones; Both – Water and Wastewater Network Project treatment zones and matched control zones.

[1] Measured on a four point scale (1 = excellent; 4 = poor)

[2] Water samples were only collected and analyzed in Zarqa

[3] The regressions for this outcome do not control for the time-varying factors because we use the full utility database, rather than restricting to the survey sample.

Table 4. Summary of main impacts on small enterprise behaviors and outcomes

Outcome	**DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample**	**DiD impact of intervention – relative to non-intervention areas in Amman, by subsample**				
	(1) WNP	(2) WWNP	(3) Both	(4) WNP	(5) WWNP	(6) Both
Water supply						
Piped water is primary source	(0.10)	n.a.	(0.10)	(0.11)	n.a.	+0.09 (0.12)
Hours piped water, for days w/water	(2.5)	n.a.	(2.5)	**-8.1***** (3.0)	n.a.	(3.0)
Water consumption (m^3/month)	+16.9 (16.8)	n.a.	+24.7 (19.2)	+12.5 (16.5)	n.a.	+25.9 (21.1)

Outcome	DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample	DiD impact of intervention – relative to non-intervention areas in Amman, by subsample				
Reported water interruption	+0.03 (0.11)	n.a.	(0.12)	+0.03 (0.14)	n.a.	+0.05 (0.15)
Expenditure on water from vendors (JD/month)	+0.21 (0.46)	n.a.	(0.49)	(0.54)	n.a.	**-1.50**** (0.61)
Expenditure on water, all sources (arcsin, JD/month)	+0.22 (0.28)	n.a.	(0.31)	**-0.57*** (0.33)	n.a.	(0.34)
Wastewater management						
Use of some wastewater system	n.a.	(0.08)	+0.02 (0.08)	n.a.	(0.10)	+0.05 (0.08)
Sewer connection	n.a.	**-0.18**** (0.09)	(0.09)	n.a.	(0.10)	+0.02 (0.09)

Outcome	**DiD impact of intervention – relative to non-intervention areas in Zarqa, by sub-sample**	**DiD impact of intervention – relative to non-intervention areas in Amman, by sub-sample**				
Cost of wastewater management (arcsin, JD/month)	n.a.	(0.32)	**-0.46*** (0.26)	n.a.	+0.05 (0.41)	(0.37)
<u>Overall welfare</u>						
Expenditure (arcsin, JD/month)	**-0.46***** (0.15)	(0.15)	**+0.36*** (0.18)	**-0.34**** (0.17)		**-0.39*** (0.21)
Asset value (arcsin, JD)	+0.25 (0.33)	+0.09 (0.36)	**-0.77***** (0.28)	+0.24 (0.30)	(0.31)	(0.31)
Land value (arcsin, JD)	+0.57 (0.35)	+0.28 (0.45)	(0.37)	+0.60 (0.43)	+0.41 (0.53)	(0.34)
Sample size for comparison						

Notes: All estimates are difference-in-differences estimates for coefficient ${}_t$ for period t=1 (after the intervention). Standard errors are shown in parentheses. Statistical significance is denoted as follows: *** p<0.01; ** p<0.05; * p<0.1. The specification controls for time and enterprise fixed effects, as well as the following other time-varying factors: reported complaints about sewer overflows; respondent years with enterprise, number of total employees, reported obstacles to growth, and frequency of water interruptions. Alternative specifications without controls yield very similar results (see Appendix C). The subsample comparisons are as follows: WNP – Water Network Project treatment zones and matched control zones; WWNP – Wastewater Network Project treatment zones and matched control zones; Both – Water and Wastewater Network Project treatment zones and matched control zones.

Table 5. Summary of main impacts on farm behaviors and outcomes

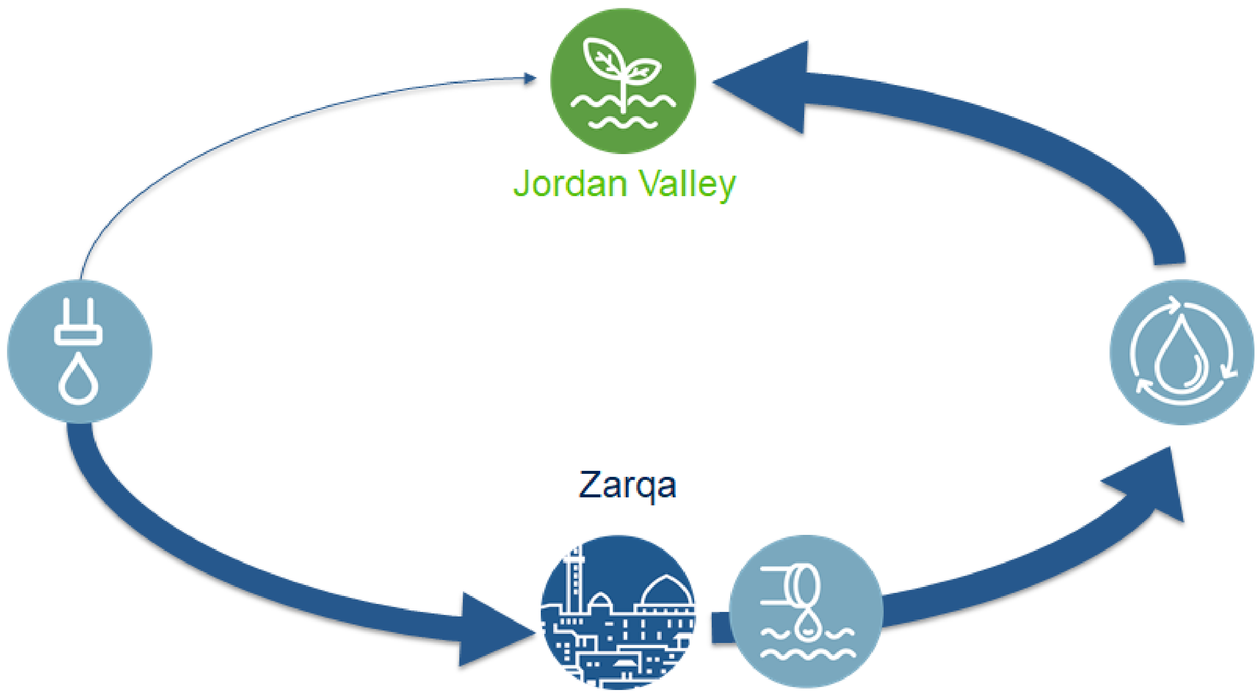
Outcome	**DiD impact of intervention – relative to non-intervention areas in Zarqa, by subsample**				
	(1) JV1	(2) JV2 (Treatment)	(3) JV3 (Treatment)	(4) JV4	(5) Highlands (Treatment)
Wastewater use in irrigation	**+0.10**** (0.04)	**+0.14***** (0.04)	**+0.15***** (0.04)	**-0.47**** (0.04)	**+0.08*** (0.04)
Perceived water quality[1]	**-0.81***** (0.3)	**-0.53*** (0.31)	(0.31)	**+1.40***** (0.31)	(0.34)
Irrigated area (dunum)	**+6.0*** (3.4)	**+8.4**** (3.5)	**-9.0**** (3.6)	**-19.9***** (3.5)	**+12.5***** (3.8)
Farm revenue (JD/yr)	**+91823*** (48757)	(51173)	(51336)	**-91772*** (51908)	**+186422***** (55180)
Farm profit (JD/yr)					
Farm land value (JD)					

Notes: All estimates are difference-in-differences estimates for coefficient $_{t}$ for period t=1 (after the intervention). Standard errors are shown in parentheses. Statistical significance is denoted as follows: *** p<0.01; ** p<0.05; * p<0.1. The specification controls for time and farm fixed effects. The subsamples are as follows: JV1 is furthest north in the Jordan Valley, and represents a set of farms that were mostly unaffected by the Compact since their water supply is independent of the Zarqa system; JV2 and JV3 represent areas where flows of recycled wastewater newly arrived (JV2) and increased substantially (JV3); JV4 represents an area that already had substantial flows of recycled water prior to the investment; Highlands farms, finally, are located along the Zarqa River and also received access to more steady water supply.

[1] Measured on a ten point scale (1 = poor; 10 = excellent)

Table 6. Summary of utility performance indicators, relative to other urban utilities in Jordan

Result	**Indicator**	**Utility**	**2008**	**20**
Reduced NRW	Indicator: Total losses			
	Total losses; network (%)	Aqaba	20.6	23
		Amman	40.1	38
		Zarqa	56.1	54
	Total losses; network (L / Subscriber / Day)	Aqaba	445	48
		Amman	340	30
		Zarqa	583	56
	Indicator: Pipe breaks/bursts per km of mainlines			
	Main bursts/ 100km	Aqaba	139	10
		Amman	109	89
		Zarqa	n.d.	n.
	Service leaks / 1000 connections	Aqaba	122	12
		Amman	221	20
		Zarqa	n.d.	n.
Increased revenue to utility	Utility revenue (2015 JD/m^3 sold)	Aqaba	1.08	0.
		Amman	1.40	1.
		Zarqa	0.90	0.
Increased cost recovery by utility	Billing efficiency (%)	Aqaba	91.5	91
		Amman	97.0	97
		Zarqa	n.d	n
	Collection efficiency (%)	Aqaba	99.4	10
		Amman	96.0	96
		Zarqa	n.d.	n.
	Operating Cost Recovery Ratio (OCRR)	Aqaba	1.43	1.
		Amman	1.07	1.
		Zarqa	0.87	0.



1. Qualitative depiction of (Top) pre- and (Bottom) ex pected post- Jordan Compact situations

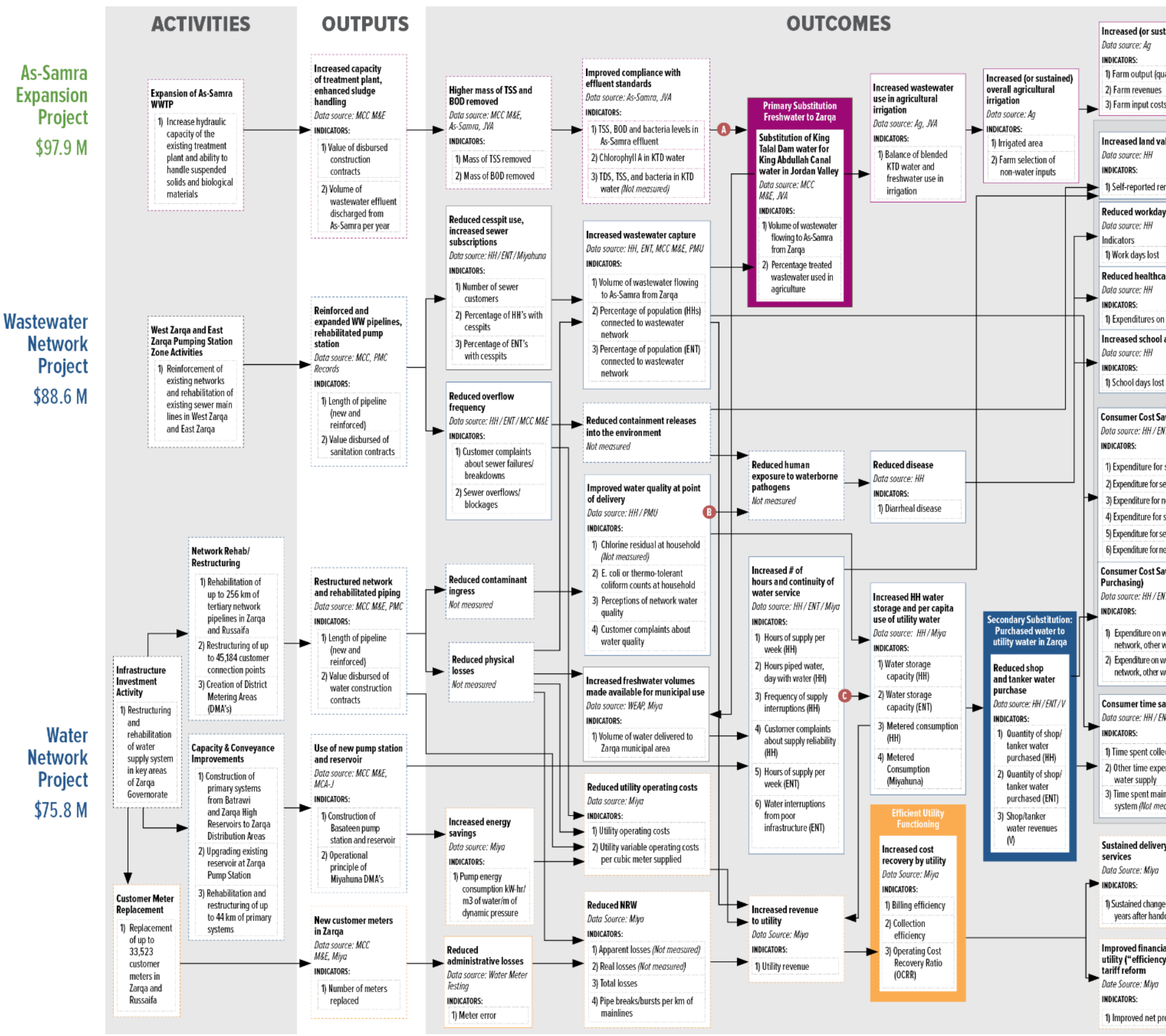


Figure 2. Full program theory of change, as elicited through participatory stakeholder consultations

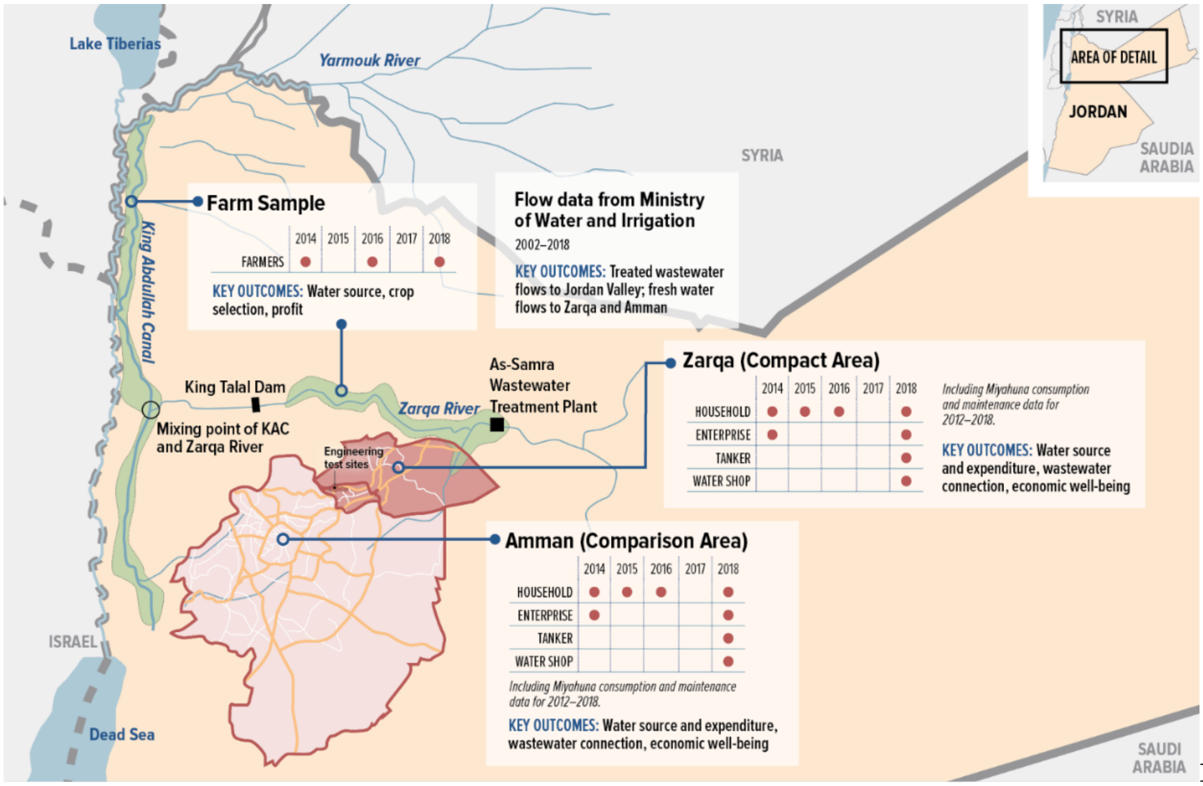


Figure 3. Locations and timing of data collection activities to support the evaluation

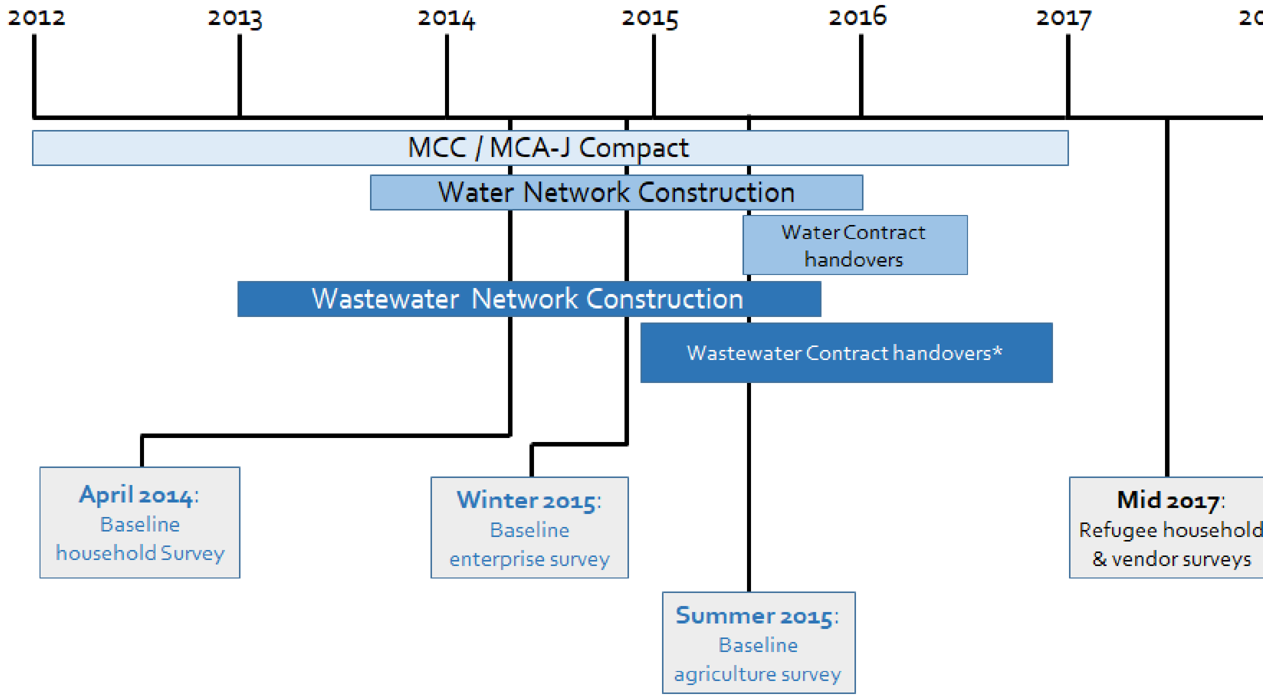


Figure 4. Timing of data collection relative to infrastructure intervention [Note that baseline surveys were conducted prior to any infrastructure operations, except for a few small wastewater Contract handovers preceding the baseline agriculture survey]

Appendix A: Sampling and description of study populations, and comparability of treatment and counterfactual groups

A. Household and small enterprises

The evaluation design had to consider that areas selected for intervention within Zarqa (potential "treatment" areas) were likely different from non-intervention (potential comparison or "control") areas. To deal with this issue, we implement a strategy of matching on pre-intervention characteristics plus difference-in-differences (DiD) analysis (Jeuland et al. 2015). The ex ante matching allowed us to select Census block-level geographic zones that were to be differentially treated by the JC but that appeared observationally similar prior to the intervention. DiD meanwhile enhances the validity of measures of impacts by netting out any unobserved and time-invariant pre-existing differences across the units in the matched areas. The key identifying assumption of this design is that key time-varying differences are also balanced across these zones. We are able to validate this parallel trends assumption by comparing pre-intervention trends in at least one key variable that pertains to water sourcing, that is, water consumption from the utility network.

Specifically, we created three different sets of matched units using data from the most recent prior Census conducted in 2004. The first set matches areas "treated" with the WNP alone with unaffected "control" units; the second matches WWNP-only areas with similarly unaffected controls; and the third matches areas treated with both the WNP and WWNP improvements to controls. In addition, because we were concerned about the potential for utility-wide spillovers within Zarqa, control samples were created in both Zarqa Governorate and in East Amman, the urban zone nearest to Zarqa that was served by a different utility. A map of the final sample zones selected for surveying in Zarqa is shown in Figure A1. Pre- and post-matching sample zone characteristics are summarized in Tables A1 and A2, and parallel trends in pre-intervention water consumption in Table A3; these show that the matching approach was successful in removing differences in Census characteristics across zones, and confirm the plausibility of the main identifying assumption of the panel DiD design.

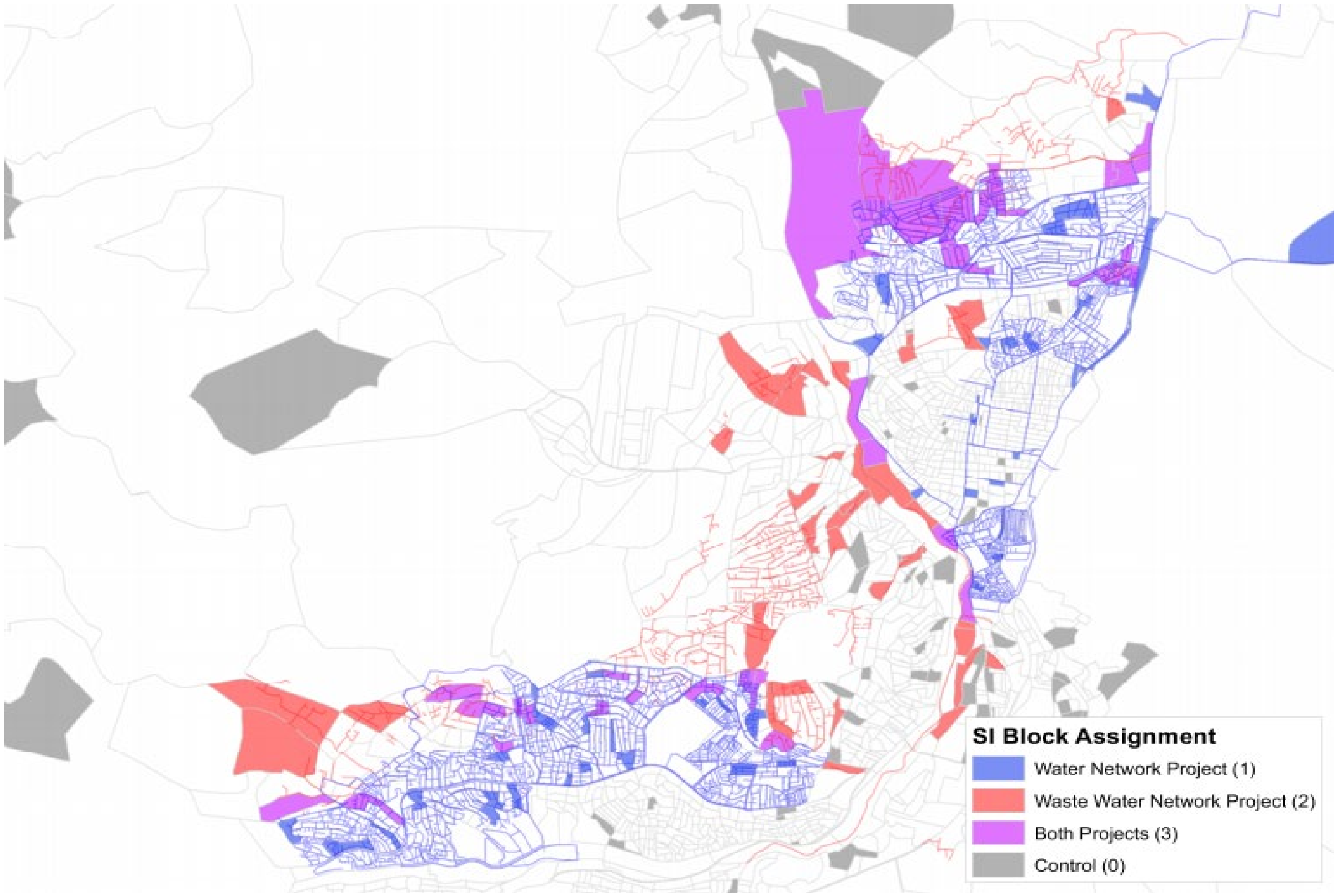


Figure A1. Map of treatment and control blocks and infrastructure works, with rehabilitated water pipes in blue and new wastewater networks shown in red (Note that some areas are off the map and therefore not shown, e.g. all controls in Amman, and some in Zarqa).

Table A1. Pre-matching descriptive statistics for Census blocks

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

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

Table A2. Post-matching descriptive statistics for selected Census blocks

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

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

Table A3. Pre-intervention trends in water consumption (all in m3) for matched sample areas

	COMPARISON WITH ZARQA CONTROLS					
	Area A Both		Area B WWNP only		Area C WNP only	
Variable	Δ	(se)	Δ	(se)	Δ	(se)
Quarter 1 consumption	1.11	(0.99)	0.48	(0.92)	0.55	(0.80)
Quarter 2 consumption	-0.53	(1.18)	-0.85	(1.21)	-0.67	(1.22)
Quarter 3 consumption	-1.11	(1.00)	-0.68	(1.12)	-2.02*	(1.16)
Quarter 4 consumption	0.66	(1.37)	-0.01	(1.68)	-0.19	(1.29)
Annual consumption	0.13	(2.21)	-1.07	(3.23)	-2.33	(2.64)
	COMPARISON WITH AMMAN CONTROLS					
	Area A Both		Area B WWNP only		Area C WNP only	
Variable	Δ	(se)	Δ	(se)	Δ	(se)
Quarter 1 consumption	0.43	(1.16)	0.62	(0.92)	0.91	(1.63)
Quarter 2 consumption	0.36	(0.88)	2.02*	(1.06)	1.89*	(1.03)
Quarter 3 consumption	-2.1**	(0.95)	0.00	(0.94)	0.60	(0.99)
Quarter 4 consumption	1.71	(1.23)	1.43	(1.14)	1.96	(1.24)
Annual consumption	0.41	(2.64)	4.07	(2.88)	5.36*	(3.18)

On the basis of the results of the matching procedure for Component 1 and pre-intervention power calculations that are detailed elsewhere (Albert et al. 2014), 325 Census clusters in Amman and Zarqa were retained for the household surveys, with a target of 3575 households (or 11 per cluster) overall. Households were then randomly selected from a sampling frame that had been updated by Jordan's Department of Statistics shortly before the baseline survey. Due to challenges in locating households and refusals, the final sample size for the baseline survey fell somewhat short of the initial target, an issue that was addressed by adding households in subsequent surveys (see Table A4). For the enterprise survey, meanwhile, enumerators enrolled a total of 425 enterprises from a sample frame provided by Jordan's Department of Statistics (76 in Amman and 349 in Zarqa) (Table 4.2). Questionnaires were recorded as complete for 356 of these enterprises; 2.0% of enterprises could not be located, while 12.3% of enterprises were closed, and 5.9% of enterprises refused to complete the survey.

Household sample. The average household size in the sample was 4.9 people at baseline in 2014, with 0.4 children per household under the age of 5, and 3.0 adults over the age of 18 years (Table A5). Most households are Jordanian (93%); 85% have a male head of household. Education levels in the sample are generally high: 91% of respondent had no trouble reading a written newspaper article, and enumerators judged that 87% had no trouble understanding the survey. The average years of education among all household adults was reported to be 10.6. Seventy-three percent of respondents own their residences, 55% of which are apartments or flats. On average, households report spending slightly more (450 JD/month) than they earn (426 JD/month). Finally, ownership of household durable goods and vehicles varies: nearly all (98%) own a washing machine, but slightly fewer than half of households own a computer (45%) and at least one vehicle (45%), respectively. Twenty percent own at least one air conditioner. The Zarqa sample has slightly lower income and expenditures (409 JD/month and 429 JD/month, respectively), but most variables are balanced across the Zarqa and Amman sub-samples.

Table A4. Sample Summary

	Wave 1	Wave 2	Wave 3	Wave 4
Completed sample size	3,362	3,399	2,780	3,600
WNP treatment households	495	493	415	529
WWNP treatment households	391	397	306	392
Both treatment households	381	382	309	407
Zarqa control households	797	808	671	764
Amman control households	1,079	1,096	748	1,096
Attrition from baseline	N/A	20%	54%	34%
New sample added	3,362	721	727	855

Turning to water and sanitation-related variables, household connection rates to utility water and sewer in our sample were 97% and 79% respectively in 2014. A large percentage (39%) of households share water meters with others. For the 75.9% of households who produced a bill or otherwise provided self-reported estimates of network water consumption, the average monthly amount used was 7.7 m3 (see Figure A2 left panel for the distribution of these monthly amounts in the sample). Meanwhile, households expressed relatively low confidence in the quality of water obtained from the utility network, rating the safety of that water to be 5.1 on a 10 point scale ranging from not safe at all to completely safe. Households however considered their own drinking water (which may be treated or obtained from other sources) to be 8.2. An average of 0.15 members per household reported having a case of diarrhea in the prior 2 weeks (representing an incidence of about 2.9%), and 0.048 reported suffering from some other water-related illness (1.0% incidence). On a 5-point scale, households rated their sanitation situation to be 2.83 (closer to acceptable (3) than good (2)).

Table A5. Household survey descriptive statistics – Demographic, socio-economic, and health status variables

	Overall sample			Zarqa sample		
Variable	**N**	**Mean**	**(SD)**	**N**	**Mean**	**(SD)**
Demographic variables						
Household Size	3359	4.91	(2.05)	2259	4.91	(2.03)
# of Children < 5 yrs	3359	0.42	(0.70)	2259	0.41	(0.69)
# of adults >=18 yrs	3359	3.02	(1.41)	2259	3.02	(1.40)
Female head of household	3359	0.15	(0.36)	2259	0.14	(0.35)
Age of head of household (in yrs)	3322	50.1	(14.0)	2247	49.9	(13.9)
Number of disabled HH members	3359	0.070	(0.30)	2259	0.064	(0.28)
Jordanian	3359	0.93	(0.25)	2259	0.93	(0.26)
Resident of Zarqa	3359	0.67	(0.47)	n.a.	n.a.	
Socio-economic status						
Respondent is literate	3359	0.91	(0.29)	2259	0.91	(0.29)
Average years of adult education	3359	10.6	(3.45)	2259	10.6	(3.33)
Respondent understood survey well (enumerator rating)	3359	0.87	(0.33)	2259	0.87	(0.34)
Homeowner	3359	0.73	(0.44)	2259	0.74	(0.44)
Area of home (m^2)	3359	247	(2843)	1909	250	(2934)
Home is an apartment/flat	3359	0.55	(0.50)	2259	0.56	(0.50)
# rooms	3358	4.24	(3.57)	2258	4.19	(4.27)
# bathrooms	3358	1.53	(0.65)	2258	1.49	(0.62)
Total expenditure (JD/month)	3272	450	(341)	2191	429	(297)
Total income (JD/month)	3214	426	(351)	2152	409	(303)
NAF recipient	3351	0.027	(0.16)	2253	0.026	(0.16)
Own washer	3359	0.98	(0.14)	2259	0.98	(0.13)
Own computer	3359	0.45	(0.50)	2259	0.44	(0.50)
Own air conditioner	3359	0.20	(0.40)	2259	0.22	(0.41)
Own vehicle	3359	0.45	(0.50)	2259	0.43	(0.450)
Have a home business	3359	0.039	(0.19)	2259	0.037	(0.19)
Took out a loan in the past year	3359	0.20	(0.40)	2259	0.21	(0.41)
Enumerator rating of wealth	3359	2.78	(0.86)	2259	2.75	(0.84)
Health perceptions						
Remember hearing water/sanitation message	3359	0.48	(0.50)	2259	0.42	(0.49)
Believe diarrhea can be prevented	3319	0.71	(0.46)	2240	0.70	(0.46)
Perceived safety of utility water (0=not at all, 10=completely)	3359	5.12	(3.22)	2259	4.88	(3.26)
Perceived safety of home drinking water (0=not at all, 10=completely)	3359	8.18	(2.00)	2259	8.11	(2.05)
Anthropometrics and health measures						
# of HH members w/diarrhea, past 2 wks.	3359	0.15	(0.53)	2259	0.15	(0.56)
# of HH members w/other water-related illness	3348	0.048	(0.21)	2253	0.050	(0.22)

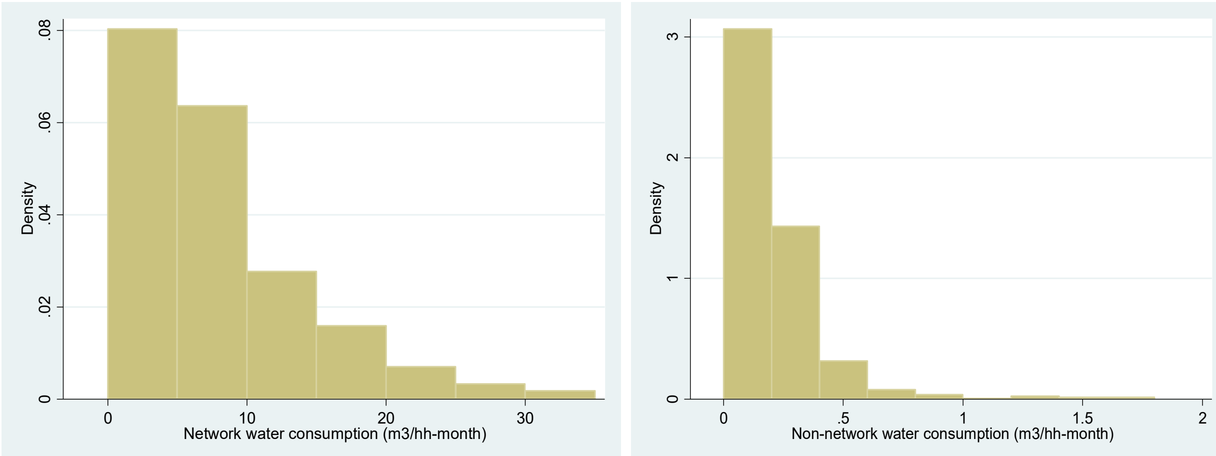


Figure A2. Distribution of monthly (left) network water amounts used and (right) non-network water amounts used *(Note the differences in the scale for the x-axis)*

Households used a variety of non-network water sources at baseline (44% use at least one other source), and shop water was by far the most frequent alternative, with 38% of households purchasing it. The quantity of water taken from non-network sources (0.38 m3/month) was far lower than the average network consumption (see Figure A2 right panel). While 34% of households change their water consumption patterns in the alternative (winter) season, very few households (<1%) change sources. Twenty-three percent of households reported experiencing water shortages; a somewhat higher proportion (29%) experienced shortages in Zarqa. Households reported receiving water 9.2 days per month on average (and 8.3 days per month in Zarqa). Water and hygiene behaviors were also variable in the sample. Thirty-five percent of households reported treating water in house at the time of the survey, and most of these (34%) were able to show enumerators a sample of treated water during the visit. Most of these households (90%) consumed treated water less than one day after treatment. A large majority of households (88%) also had soap at the time of the visits.

There was some variation in the quality of the household samples that were tested. These samples were collected from storage containers; their quality is thus likely affected by storage and handling within the home. We found little evidence of E. coli contamination problems: It was below detection (1 colony-forming unit per 100 mL sample) for all tap water samples collected from household storage tanks. Three of 91 samples that had been sourced from water shops and subsequently stored at home showed modest E. coli contamination (7, 28, and 54 CFU/100mL, respectively). Unsurprisingly, there was somewhat higher prevalence of total coliforms in stored water: these were detected in ~10% of the tap water samples, though counts were generally low: All samples except one had less than 100 CFU/100mL. Finally, likely due to the fact that such water is stored for longer periods of time, we detected total coliforms in over 70% of shop water samples, with 29% of samples having more than 100 CFU/100mL and 11% more than 1000 CFU/100mL. Collectively, these results suggest that dangerous levels of microbial (*E.coli*) contamination are not widespread in this sample, but that storing water for long periods may elevate risks.

Using the extensive data collected in the baseline survey, we were able to estimate water- and wastewater-related expenditures, and compute coping costs in different categories (e.g., time spent collecting water from alternative sources, or in-house water treatment costs). The calculation of various components of these coping costs are described in more detail in Orgill-Meyer et al. (2018) and summarized in Figure A3 and Table A6.

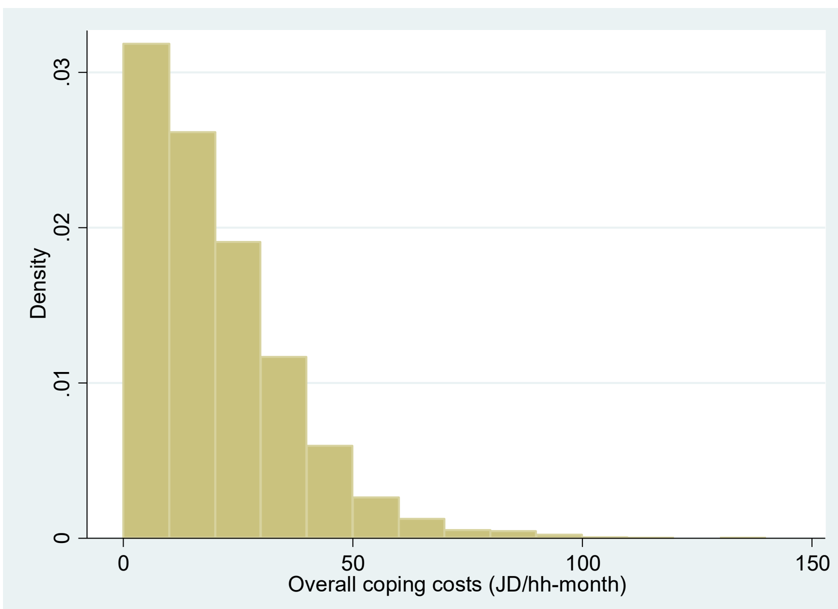


Figure A3. Distribution of total water and wastewater-related coping costs among survey households

As shown, households spend about 6.2 JD/month on network water and sewer; the amount is only slightly lower (5.8) in Zarqa. The largest category of water-related coping costs is for purchase of non-network water (roughly 6.5 JD/month), which is striking given the much lower amounts of non-network water that are consumed; the possibility of reducing this significant expenditure is one of the key motivators for the Compact. Households spend another 5 JD/month on water collection, treatment, storage, and on repairs to household infrastructure, such that overall water-related coping costs are 11.5 JD/month (and 11.8 JD/month in Zarqa). Wastewater-related coping costs are primarily in pit emptying (3.2 JD/month) and toilet infrastructure (6.5 JD/month). The wastewater-related coping costs in the full sample and in Zarqa are similar (9.9 JD/month vs. 9.6 JD/month).

Table A6. Water and wastewater-related expenses, and coping costs (all in JD/month)

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

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

Enterprise sample. Table A7 provides an overview of firm composition, obstacles, and growth. The average enterprise in our sample has 5 employees, most of whom are male and full-time workers. The average full-time skilled worker for the businesses in our sample is 30 years old. The majority of business owners are male, and the average business owner in our sample is 46 years old with 15 years of experience. Respondents in our sample cited inflation and price instability, electricity costs, and water supply costs as the largest obstacles to growth for their enterprises. Very few enterprises in our sample reported making any investments in their business or taking out a loan in the past year.

Table A7. Enterprise survey descriptive statistics – employee data, business owner characteristics, obstacles to growth, assets, and costs

	Full sample			Zarqa sample		
Variable	**N**	**Mean**	**(St. Dev)**	**N**	**Mean**	**(St. Dev.)**

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

Table A8 contains detailed descriptive statistics of water and wastewater usage and costs for enterprises. We observe that firms use a combination of piped connections, water tankers, and water shops for their business needs. Figure A4 shows the distribution of water sources that enterprises reported as their "primary water source." We observe a fairly even distribution of private (company) piped connections, tankers and water shops as enterprise primary water sources. Some enterprises share connections with others or with households. Enterprises spend an average of 57 JD/month for their total water usage, and only 10% of enterprises treat any water, either for drinking or production purposes.

Table A8. Enterprise water and wastewater use practices and characteristics

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

While most enterprises do not report having to shut down due to water shortages, enterprises with piped connections report receiving an average of 7.9 days of piped water during a summer month and only 42.8 hours in a normal week. Figure A5 displays the primary reasons that enterprises do not have piped connections. Nearly 50% of enterprises report that high connection (not tariff) costs prevent them from using piped water, and almost 90% of unconnected enterprises list this as the main obstacle to their use of network water.

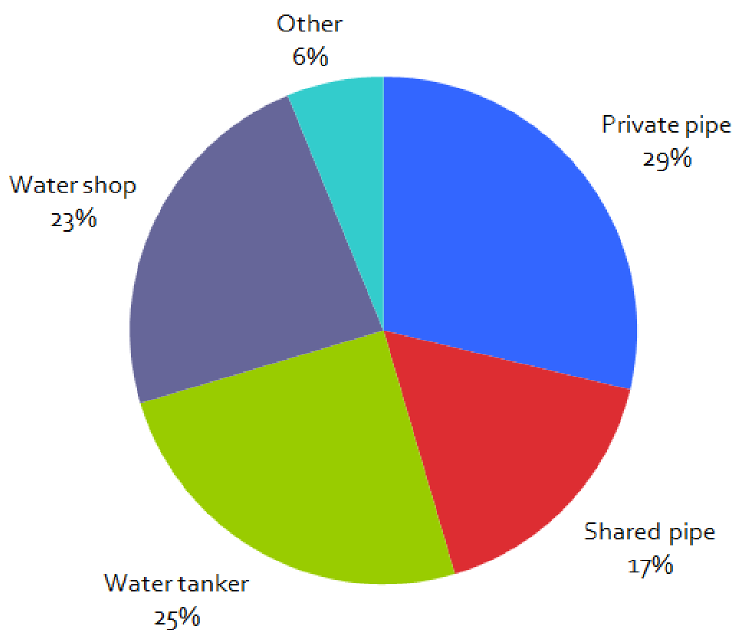


Figure A4. Distribution of main water sources used by surveyed enterprises (n=341)

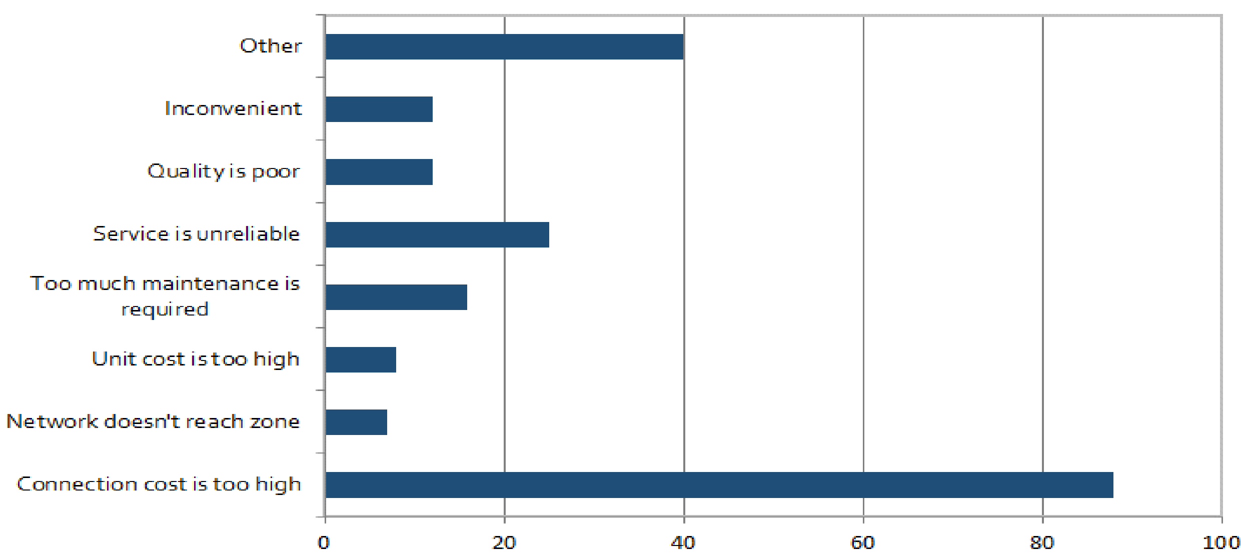


Figure A5. Reasons why enterprises do not have piped water (n=178)

B. Farms

With this sample, we exploit the natural experiment of temporal and spatial variation in the source and timing of water supply to farmers. This variation arises as a dual consequence of natural hydrological variability and a gradual tightening of water supplies in Jordan due to urban population growth. The natural experiment is supposed to accelerate following the JC investments, which facilitate a transition into greater reuse of treated wastewater and away from freshwater in the JV. It is important however to emphasize that the changes observed in the JV are not wholly due to the JC, however. Other investments – in expanded wastewater management capacity and treatment, and in conveyance infrastructure – also play a role in this transition.

For the farm analysis, we apply this DiD analysis to the longitudinal farm-level data collected from 4 different areas in the JV as well as an additional area located along the Zarqa River (Table A9 and Figure A6). These areas are characterized along a continuum that ranges from primary use of freshwater from the Yarmouk River and Lake Tiberius, in Northern zones, to primary use of blended water (treated wastewater mixed into the Zarqa River), in the Middle to Southern areas.

A fifth area along the Zarqa River in the highlands is also included since agricultural expansion may occur there as a consequence of augmented upstream production of treated wastewater. We contextualize these farm-level changes using systems-based water balance modeling that helps isolate the relative contribution of the Compact investments to the changing water supply. A systems model is needed to account for other dynamics such as hydrological variability, expansion of alternative water supplies, and other changes that lead to increased reuse of treated wastewater, for example, from other cities supplying treated wastewater to the JV. An illustrative schematic depicting the sample areas and how they relate to the water balance is shown in Figure A6.

For the agricultural surveys in Component 2, 110 farms were randomly selected from lists of farms operating in each of the five sample areas (550 in all). Accordingly, the samples may not be representative, of households and enterprises in Zarqa and Amman, on the one hand, and of farmers in the Jordan Valley and highlands, on the other.

Table A9. Overview of the Natural Experiment Design

Zone	Column 1: Before Expansion	Column 2: After Expansion (anticipated)	Column 3: After Expansion (measured)	Interpretation

JV1: North	No blended water supply (all freshwater)	No blended water supply	New supply from non-JC sources	Validation and contextualization of findings
JV2: North-Mid	No blended water supply (all freshwater)	New blended water supply	New blended water supply	Treatment
JV3: Mid-North	Mostly blended water supply	Slightly increased blended water supply	Increased blended water supply	Treatment
JV4: Mid-South	Full blended water supply	No change in blended water supply	No change in blended water supply	Control
Highlands	Substantial supply	Increased blended water supply	Increased blended water supply	Treatment

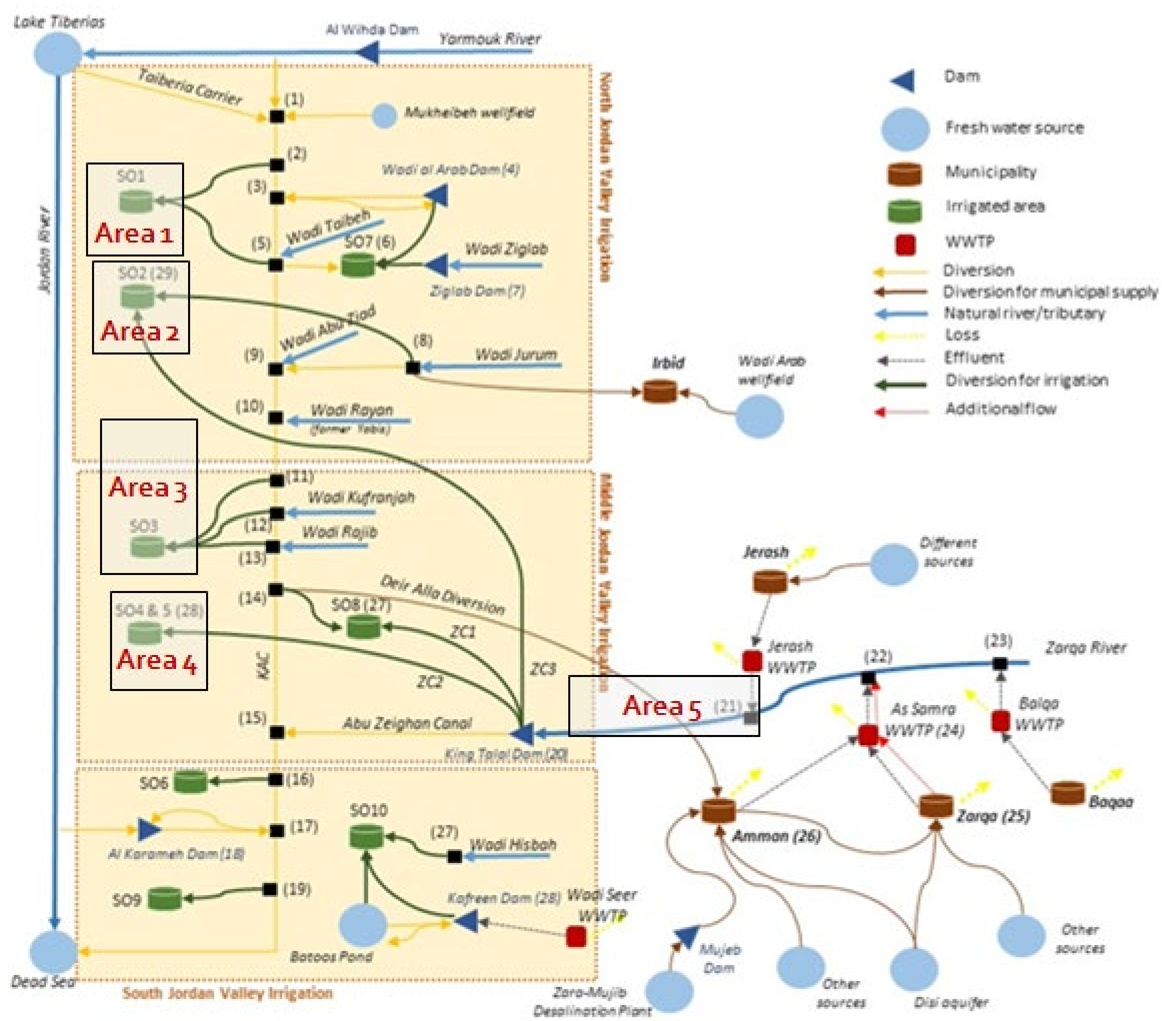


Figure A6. Schematic representation of the irrigation water system; showing the locations included in the agriculture survey

Farm sample. The average age of sample farmers is 52.1 (Table A10). Most (95%) are Jordanian (94%), married, and average household size is about 5. Half of these farmers own their farm; the average farm size is 38 dunums, 32 of which are cultivated. The average farm value was reported to be worth 166,700 JD. Most farms did not have buildings (residence, storage facilities, or livestock buildings) on their land. Eighty percent of farmers reported growing crops in the winter in the prior year, and 40% grew crops in the summer. Reported average monthly incomes and consumption were 1297 JD and 892 JD, respectively. The average farm had about 22,400 JD in assets, with 7,190 JD of this asset wealth being in water-related assets.

Table A10. Farmer survey descriptive statistics (Demographics, farm and socioeconomic characteristics)

Variable	N	Mean	SD
Demographic variables			
Age	548	52.1	12.7
Gender	550	1.00	0.06
Married	550	0.95	0.21
Children in hh	550	1.69	2.02
Adults in hh	550	3.52	2.99
Literate	550	0.93	0.25
Farmer training	550	0.03	0.18
Jordanian	550	0.94	0.23
Basic Farm information			
Area of land	550	37.6	34.7
Area cultivated	550	32.3	27.1
Market value of land (thousands of JD)	469	166.7	125.3
Own farm	550	0.50	0.50
Manage farm for others	550	0.06	0.24
# of buildings on farm	550	0.84	0.92
Grow crops in winter	550	0.80	0.40
Grow crops in summer	550	0.40	0.49
Socioeconomics			
Total HH consumption/month (JD)	550	892	750
Total HH income/month (JD)	550	1297	1577
NAF recipient	550	0.01	0.09
Took loan in past year	550	0.03	0.16
Saved in past year	550	0.15	0.36
Total value of assets (‘000 JD)	550	22.4	29.3
Value of water-related assets (‘000 JD)	550	7.19	7.32
Own tractor	550	0.14	0.34
Own transport vehicle	550	0.32	0.47
Own cooling equipment	550	0.01	0.10
Own plough	550	0.11	0.31
Enumerator opinion of wealth (1=very poor; 5=rich)	550	3.13	0.90

On average, farmers cultivating during the winter season (n=441) spent about 13,400 JD in agricultural inputs in 2015, with most of those costs being spent on vegetables (9,630 JD) and trees (3,250 JD) (Table A11). Field crops and flowers had considerably lower input costs, partly because of the smaller area and lower number of farmers growing such crops. In the summer, input costs among farmers growing crops (n=220) were slightly higher for trees (3,230 JD) than for vegetables (2,840 JD). The majority of farms (0.64) had at least one permanent worker, and the average number of permanent workers per farm was 1.55. Farms, on average, also have roughly 4 unpaid employees.

Consistent with the lower costs observed in summer, the average area of land used for farming was larger in winter than summer, though in both seasons, trees took the largest area (13.8 and 14.5 dunum, in winter and summer, respectively), followed by vegetables (12.2, and 6.97), field crops (4.56 and 3.06), and lastly, flowers (0.18 and 0.09). In winter, vegetables produced the highest output of 76 tons, followed by, trees (28.5 tons), and field crops (26.5 tons). In summer, the order was different with trees producing the greatest output (34.0 tons, followed by vegetables (16.1 tons), and then field crops (7.85 tons). On average, farms generated revenues of 23600 JD in the winter and 8860 JD in the summer.

Farmers did not report feeling very constrained by water-related issues on their farm. Their biggest reported constraint was with the irrigation water amount (ranked on average of 3.05 on a scale of 1 being excellent and 5 being very poor). Farmers ranked canal position, irrigation water quality, drainage and soil fertility as 2.87, 2.81, 2.69, and 2.29, respectively. For both the prior winter and last summer, farmers reported receiving more water (17,600 m3 and 6,850 m3, respectively) than they had planned on (13,900 m3 and 5,050 m3), on average (Table A12). On average, water quality was viewed as a modest constraint (ranked 2.89 in winter and 2.78 in summer on a scale of 1 being excellent and 5 being very poor), and about half of the farmers reported having water quality problems. Few farmers found the level of water they received to be sufficient (0.36), but very few (0.07) adjusted their crop mix because of water shortages. A quarter of farms use water pumps, and on average, a farm spends 318 JD/year on these pumps and 450 JD/year on network water repairs. Only 3% of farmers used groundwater. The total payment for water is 479 JD/year.

Table A11. Farming Inputs, Outputs and Constraints

Variable	N	Mean	SD	N	Mean	SD
Inputs‡		*Winter*			*Summer*	
Vegetables: Total ('000 JD)	441	9.63	15.0	221	2.84	8.17
Field crops: Total ('000 JD)	441	0.38	1.52	221	0.30	1.53
Trees: Total ('000 JD)	441	3.25	7.00	221	3.23	9.11
Flowers: Total ('000 JD)	441	0.20	2.63	221	0.22	1.31

Overall: Total ('000 JD)	441	13.4	14.6	221	6.59	11.4
Have permanent workers	550	0.64	0.48			
Number of permanent workers	550	1.55	2.42			
Total pay for permanent workers ('000 JD/year)	550	4.12	5.62			
Number of workers without pay	550	4.01	15.0			
Outputs/revenues‡		*Winter*			*Summer*	
Vegetables: Area (dunum)	441	12.2	14.9	221	6.97	11.2
Field crops: Area (dunum)	441	4.56	14.8	221	3.06	11.5
Trees: Area (dunum)	441	13.8	29.6	221	14.5	21.1
Flowers: Area (dunum)	441	0.18	1.94	221	0.09	0.95
Overall: Area (dunum)	441	30.8	16.7	221	24.6	13.3
Vegetables: Output (tons)	441	76.0	139	221	16.1	44.5
Field crops: Output (tons)	441	26.5	341	221	7.85	37.0
Trees: Output (tons)	441	28.5	102	221	34.0	53.8
Flowers: Output (tons)	441	0.00	0.048	221	0.00	0.00
Total revenue (calculated, in '000 JD)	441	23.6	52.3	221	8.86	27.8
Other farming‡		*Winter*			*Summer*	
Last year yield=normal	441	0.80	0.40	222	0.83	0.38
Land area left fallow last year	441	4.92	14.3	222	7.85	14.8
Livestock		*Both Seasons*				
Have livestock	550	0.04	0.20			
# of animals	550	2.21	17.6			
Purchases of animals last year (JD)	550	10.5	138.6			
Sales of animals last year (JD)	550	38.9	400.1			
Cost of inputs for animals last year (JD)	550	255	2147			
Constraints (1=excellent; 5=very poor)		*Both Seasons*				
Soil fertility	550	2.29	0.79			
Irrigation water amount	550	3.05	1.06			
Irrigation water quality	550	2.81	1.09			
Canal position	550	2.87	1.02			
Drainage	550	2.69	0.74			
Input costs	550	3.52	0.91			
Access to capital	550	3.00	0.84			
Access to markets	550	2.92	0.96			
Volatility in crop prices	550	3.71	0.92			
Taxes and licensing fees	550	3.21	0.96			

‡ Inputs and revenues are only reported for farms with activity in that season, hence the smaller sample sizes for these statistics.

Table A12. Farm irrigation

Variable	N	Mean	SD	N	Mean	SD
		Winter			*Summer*	
Planned water (m^3)	441	13893	20503	221	5052	7482
Actual water (m^3)	441	17559	25854	221	6846	9647
Average water quality (1=excellent, 5=very poor)	441	2.95	0.95	221	2.58	0.85
Water amount last season=normal	441	0.83	0.37	222	0.85	0.36
		Both seasons				

Water sufficiency	550	0.36	0.48
Have spray irrigation	550	0.05	0.21
Have drip irrigation	550	0.76	0.43
Have water storage tank	550	0.25	0.43
Use pumps	550	0.25	0.43
Spending on pumps	550	318	795
Spending on network	550	450	633
Use groundwater	550	0.03	0.18
Changed crop mix because of water shortage	550	0.07	0.26
Have water quality problems	550	0.49	0.50
Water quality rating	550	5.38	2.03
Pay for water	550	0.83	0.38
Amount paid for water each year	550	479	1249
Farmers in area try to avoid payment	462	0.19	0.39
Water user association exists	550	0.20	0.40
Pay WUA (yes/no)	550	0.04	0.20
Effectiveness of WUA	110	2.55	0.88
Member of WUA	550	0.05	0.23

As expected, farm characteristics and activities varied significantly across sample zones. Some of the most important differences are discussed here. While the total average farm area was smallest in the North Jordan Valley, more of the land in that zone was used, and there was a gradient in intensity of farming from North to South. Reported land values, meanwhile, were highest in the middle Jordan Valley and in the highlands. Farming is much more labor intensive in the South and Middle Jordan Valley; this reflects the greater production of vegetables and flowers in those regions. Trees, meanwhile, are hardly grown in the southern zones, probably due to salinity problems with reused water. Total revenues are highest in the South and Central Jordan Valley, where agriculture is more focused on vegetables and field crops. Meanwhile, water quantity and canal position are most severe constraints among farmers in the North Valley (compared to water quality and soil fertility), while, water quality and quantity are most severe in the Middle and South. Pumps and spending on water and water-related assets are highest in the South, and only farmers in the South and South Central report changing their crop mix due to water quantity problems.

Finally, baseline willingness-to-pay for a more regular supply of blended water, which was the subject of a contingent valuation exercise conducted in the baseline survey, was uniformly low in the North Jordan Valley, and relatively higher in the South Jordan Valley. This evidence is consistent with the fact that the latter zone has had experience using such water for irrigation purposes, contrary to the former.

Appendix B: Description of surveys

For each of the surveys, enumerators from Jordan's Department of Statistics were trained to collect data using electronic survey instruments that were thoroughly pre-tested prior to deployment. The household survey included 13 modules and took approximately 40 minutes to complete. It was developed to collect information on household demographics; water sourcing (including network, tanker and shop water), storage, and use behaviors; preferences and satisfaction with water and sewer service; water quality measured at the tap and in in-house storage containers for a sub-sample (E. coli and thermo-tolerant coliform counts); coping and health costs related to intermittent water supply and poor water quality; and expenditures, income, and other socio-economic characteristics.

The enterprise surveys focused on enterprise characteristics, production inputs and outputs, costs and revenues, and assess constraints with regards to using water as an input to production. In addition, for assessing impacts on Zarqa's important informal sector, we relied on the informal production activities carried out by households selected into our sample. The instrument for the enterprise survey had 10 modules and took approximately 1 hour to complete.

The agriculture surveys, which took roughly 1 hour to complete, recorded information on basic farm characteristics (soils, canal location, etc.); inputs and costs; outputs and revenues; advantages and constraints; assets and equipment; animal husbandry; irrigation water situation and management; willingness to pay (WTP) for more dependable irrigation water supply (of blended, not freshwater); and socio-economic status and characteristics of the farmer.

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

Instruments used in the follow-on surveys with households, enterprises, and farmers were slightly shorter than those deployed at baseline.