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

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

[1] Sanford School of Public Policy and Duke Global Health Institute, Duke University; Durham, USA
[2] Franklin & Marshall College; Lancaster, USA
[3] Sanford School of Public Policy, Duke University; Durham, USA
[4] Social Impact; Arlington, USA
[5] RTI International; Research Triangle Park, USA
[6] Independent Consultant, Aqaba, Jordan
[7] Aquaya Institute; San Anselmo, USA

[ǂ] Corresponding Author: Email: marc.jeuland@duke.edu ; Tel: +1-919-613-4695; Address: Sanford School of Public Policy; Box 90239; Durham, NC, 27708; USA

September 2022

## 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.

**Keywords:** Piped water and sewerage; Jordan; water efficiency; wastewater reuse; water resource systems; development; quasi-experimental methods

## 1. Introduction

As global population and consumption of water rise, concerns that humankind is entering a new age of global water scarcity are increasingly widespread [*Liu et al.*, 2017]. To some, this rising water scarcity is worrisome because water is uniquely essential for myriad purposes – for drinking and critical domestic uses, as an input to food and industrial production processes, and for general human and ecological well-being [*Hanemann*, 2006]. Some predict that water's essentialness will inevitably lead to a zero sum game and loss of livelihoods for specific users, widespread social destabilization, and environmental damage. Such warnings are perhaps most commonly heard in countries or regions where water scarcity is becoming a binding constraint to growth – the Middle East, Western United States, parts of Australia, and in river basins with intense water competition (e.g., the upstream Ganges, Nile, Mekong, or Tigris-Euphrates). Indeed, much of the globe already experiences acute economic water scarcity, due to a lack of high quality infrastructure and an inability of institutions to consistently provide the resource to end users [*Rijsberman*, 2006]. However, both institutional and infrastructure solutions, when designed and operated effectively, can dramatically improve water management, and thereby ease tensions.

At the same time, the effects of both water infrastructure and management interventions may vary, and need to be understood within their particular contexts. Learning which interventions work, and under what conditions, is vitally important, both for the very practical work of improving the performance of subsequent interventions within the specific context being targeted, and for applying broader lessons about drivers and impediments of key mechanisms of change to other contexts. Still, the methods to learn about impacts and mechanisms in the water sector – particularly the science of impact evaluation (IE) as applied to water and sanitation projects – remain imperfect. In short, the fundamental challenge facing researchers working on such IEs is to isolate causal effects of infrastructure from many other contemporaneous changes that affect water supply and sanitation services, water resources systems, and human welfare and well-being. In addition, as we discuss in this paper, the typical "gold standard" methodology for determining causal impacts, the randomized controlled trial (RCT) [*Bothwell et al.*, 2016; *Duflo et al.*, 2007], frequently faces validity, relevance, and practical challenges for the case of water infrastructure evaluation, and is arguably poorly suited to such applications. The considerations we highlight echo and draw on various critiques of the pre-eminence ascribed to RCTs in IE literature more generally [*Deaton*, 2009; 2019; *Ravallion*, 2018].

This paper offers a perspective on the design and role of IEs for large water projects. We begin by discussing prior relevant literature, noting the growing body of rigorous work aimed at documenting the effects water and sanitation investments. The review reveals that prior evaluations typically provide answers to narrow questions and therefore face a challenge in providing a complete perspective on the effects of water infrastructure. This shortcoming is amplified as the scale of investment increases. In other words, the ability of IE to estimate impacts on well-being becomes more limited as the scale of the project under evaluation increases. This observation naturally leads to characterization and description of a number of key challenges that impede more holistic evaluation, including especially the tradeoffs or evaluation burdens inherent in such an investigation. Among these, we highlight the central IE challenge – the problem of rigorous causal attribution, but also focus on a set of other issues that deserve particular attention in the context of water infrastructure. Specifically, we emphasize the imperative of: engaging with project planners to conduct detailed *ex ante* mapping of an intervention's theory of change; planning for spillovers and systems level changes, as well as the design contamination risks; considering distributional impacts (who wins and loses); monitoring non-monetary impacts including, e.g., quality of life changes; and finally, communicating clearly what aspects can and cannot be considered by a pragmatic and cost-effective IE approach.

To more effectively demonstrate these ideas, and shed additional light on the tradeoffs and particular problems that can emerge in such IEs, we discuss an effort to apply these principles to a US$275 million infrastructure investment in urban areas of the Zarqa Governorate, Jordan. The locations targeted by this investment are lower-income areas in one of the most water scarce countries in the world. Through this application, we add to a surprisingly thin empirical literature on the *ex post* economic benefits of large water infrastructure [*Cox et al.*, 1971; *Hanemann*, 2006] and provide new evidence on the economic burden of unreliable water supplies in the Jordanian context. This evidence is timely because one of the Jordanian government's major current objectives, supported by numerous policy reforms and changes in the water sector but few causal IEs, is to improve urban water security [*Royal Commission for Water*, 2009]. Various reforms are occurring in a complicated political economy context that constrains the use of price instruments, owing to widespread popular opposition to higher water bills [*Klassert et al.*, 2018].

The remainder of the paper is organized as follows. In Section 2, we discuss prior literature on the impacts of large infrastructure, with a particular focus on the water and sanitation sector and especially urban piped water and sewer improvement. Section 3 describes the general challenges that confront evaluators of such projects, and Section 4 presents the example urban water infrastructure application that clarifies the nature of many of these problems. Section 5 presents an integrated view of the results from that application, and Section 6 discusses these findings and offers general reflections on the value of IE methods for assessing the impacts of water infrastructure.

## 2. Background: Prior literature on the impacts of infrastructure projects

We begin this section by summarizing prior reviews and perspectives on IEs of large infrastructure, before turning more specifically to those in the water and sanitation sector. We draw on several notable reviews or surveys of infrastructure evaluations in developing countries, which is our contextual focus given the important potential link between such investments and economic development outcomes [*Brakarz and Jaitman*, 2013; *Estache*, 2010; *Raitzer et al.*, 2019; *Sawada*, 2015]. Estache [2010] highlights the relative paucity of large infrastructure IEs (in water, transport, and electricity) relative to those of interventions in other sectors, especially health and education, despite a high policy demand for rigorous evidence. He largely attributes this dearth of evidence to methodological challenges that push academic researchers to forgo evaluation efforts: problems of non-random assignment (which threaten causal inference); the fact that their benefits take a long time to materialize; and the challenge of dealing with complex feedbacks linking economic development and infrastructure. The review work also highlights that impacts are highly variable across intervention types, technology, social and institutional contexts, and the information or knowledge that target beneficiaries possess, and emphasizes that good infrastructure IEs must explicitly address spillovers (i.e., effects on populations not considered to be directly “treated” by the interventions).

These challenges notwithstanding, existing work focusing on road construction and extension of the electrical grid in developing country contexts emphasizes the importance of such infrastructure to economic growth. The evidence is perhaps strongest for roads and connectivity [*Aggarwal*, 2018; *Dercon et al.*, 2009; *Ghani et al.*, 2016; *Jedwab and Storeygard*, 2022], though several studies note substantial heterogeneity in outcomes. Analyzing the effects of road construction in 39 African countries, Jedwab and Storeygard [2022], for example, find that connected cities grow faster than unconnected ones, but that the elasticity of this growth is greater for smaller and more remote cities, and weaker in politically favored or agriculturally suitable areas. The political economy result is particularly important in light of the preferential siting of infrastructure in favored locations [*Blimpo et al.*, 2013]. Growth also appears to be primarily driven by rural to urban migration. In rural settings, however, the complementary conditions needed for positive impacts on village economies (mainly income and wealth gains) may remain elusive even when market connectivity is enhanced [*Asher and Novosad*, 2020; *Mu and Van de Walle*, 2011]. Other work highlights the importance not just of the connectivity provided by roads, but in their quality [*Casaburi et al.*, 2013]. Evidence of the positive benefits of rural electrification, by contrast, is somewhat more mixed. While several studies in middle-income countries have shown positive impacts for some types of outcomes [*Dinkelman*, 2011; *Lipscomb et al.*, 2013], reviews synthesizing the broader literature suggests rather muted impacts, especially in comparison to investment costs, and again points to the importance of heterogeneity [*Bos et al.*, 2018; *Jeuland et al.*, 2021b; *Lee et al.*, 2020; *Peters and Sievert*, 2016].

Turning to water infrastructure specifically, despite the obvious importance of such investments to confront economic water scarcity [*Molden*, 2013], there are surprisingly few evaluation studies outside of rural settings. Indeed, most related literature analyzes the impacts of non-network water, sanitation and hygiene (WASH) solutions, as commonly implemented in rural areas. Moreover, even in the water sector, many infrastructure IEs focus on child health and school attendance as the main primary outcomes, with time savings, social/gender inclusion and political participation, privacy, and various other welfare indicators garnering less attention (income, consumption, coping costs) [*Estache*, 2010]. Common methods deployed in such studies include rigorous field-based IE designs [*Duflo et al.*, 2015; *Hammer and Spears*, 2016; *Lokshin and Yemtsov*, 2003; *Meeks*, 2017; *Pattanayak et al.*, 2010], regression methods applied to cross-sectional or panel data [*Esrey*, 1996; *Pickering and Davis*, 2012; *Whittington et al.*, 1990], and reviews and meta-analyses that synthesize evidence. The latter are almost exclusively focused on health [*Waddington et al.*, 2009; *Wolf et al.*, 2022]. Still in rural areas, a minority of studies examine the effects of piped water [*Brown et al.*, 2013], where the comparison group is typically comprised of households without access to network services. Finally, much of this work speaks to distributional effects in their emphasis on particular sub-populations of beneficiaries, especially children (for health improvements) (e.g., Hammer and Spears [2016]), and women (concerning time savings) (e.g., Pickering and Davis [2012]; Meeks [2017]).

For investments in piped water and sanitation, existing evidence primarily comes from urban or peri-urban settings, and mostly covers the extension of services to unconnected populations, and benefits produced over the near term. Galiani et al. [2009] found that increased water access in shantytowns can lower household coping costs, leading to increased savings of money and time; a set of somewhat different studies examine the impacts of general slum upgrading that included piped water connections among other improvements [*Soares and Soares*, 2005]. Positive but non-statistically significant effects on income and labor allocation were also found in a difference-in-differences analysis of secondary data covering both urban and rural Vietnam [*Nguyen Viet and Vu*, 2013]. Time savings – along with reduced intra-household conflict over water – were also an important outcome of experimentally-induced increases in piped water connections in urban Morocco [*Devoto et al.*, 2011].

Health gains have also been identified in several settings and over different time periods by examining the introduction of piped water and sanitation service access and various measures of illness and mortality over time [*Alsan and Goldin*, 2015; *Galiani et al.*, 2005; *Gamper-Rabindran et al.*, 2008; *Jalan and Ravallion*, 2003]. Health improvements were not detected in the aforementioned study from urban Morocco, however, perhaps because transmission of diarrheal disease in target communities was low to begin with, or because of network water quality problems [*Devoto et al.*, 2011]. To be sure, beneficiaries often worry about the quality of piped water: respondents to a survey in Nigeria were willing to pay a large premium for water from vendors, despite major structural improvements of the water system [*Whittington et al.*, 1991].

Efforts to improve existing piped systems' quality or technology, rather than new connections, have been less frequently studied. In this domain, the phasing in of better water treatment –

specifically chlorination – was found to sharply reduce mortality in the US in the early 20th Century [*Cutler and Miller*, 2004]. Also related to quality and reliability, however, is whether improvements can be sustained over time while keeping costs manageable given the development context of a particular place. Here the experience of urban Yemen is instructive; researchers found that access to piped water supply actually *worsened* health outcomes, and attributed this to rationing and a buildup of pollution in the network [*Klasen et al.*, 2012]. Such issues may be a particular challenge in developing countries where network water is rationed, and where low water tariffs have been linked to shortages and reduced long-term utility performance [*Bucknall et al.*, 2007; *Foster and Briceño-Garmendia*, 2009; *Jeuland*, 2012]. This may lead to perpetuating an infrastructure quality trap [*Burt et al.*, 2018; *Ercumen et al.*, 2015; *McRae*, 2015].

Compared to piped water, there is much less evidence related to urban sewerage, and most of that existing evidence focuses exclusively on health. Two key references are Waddington et al.'s [2009] and Wolf et al.'s [2022] systematic reviews. These studies identified six IE estimates in total in this category of interventions, of which five found no significant impacts on health [*Galdo and Briceño*, 2011; *Klasen et al.*, 2012; *Kolahi et al.*, 2009; *Pradhan and Rawlings*, 2002], versus the one that did [*Moraes et al.*, 2003], though pooled estimates were somewhat more positive [*Wolf et al.*, 2022]. Historical studies from OECD countries are more definitive in relating diffusion of sewers to declining mortality, but are also somewhat limited in number [*Alsan and Goldin*, 2019; *Kesztenbaum and Rosenthal*, 2017].

Importantly, all of the micro evidence on the impacts of water and sanitation infrastructure discussed above concerns impacts on households, though older literature from developing Asia and OECD countries has investigated linkages between water infrastructure and national or regional economic income [*Cicchetti et al.*, 1975; *Uchimura and Gao*, 1993]. Additional descriptive (but not counterfactual-based) evidence related to the impacts of piped water and sewer networks also supports the idea that urban water supply investments may reduce firms' input costs [*Schwartz and Johnson*, 1992]. The idea here is that beneficiary firms will potentially respond to reductions in costs with expanded production and employment, investment, and profit, or by reducing output prices to the benefit of consumers. These linkages are perhaps especially important where scale economies for piped services are large, water is a major input to production, and current alternative sources are inadequate [*Schwartz and Johnson*, 1992]. Like households, however, demand for piped water supply improvements may be limited by concerns over quality [*Davis et al.*, 2001].

All in all, IE literature on the effects of regional or urban water and sanitation infrastructures remains thin. Considering the relative richness of evaluation studies of stand-alone systems in rural areas, this lack of evidence cannot reasonably be attributed to the lack of importance of such evidence. Rather, it seems much more likely that rigorous counterfactual studies of urban and network improvements are difficult to implement. At least as importantly, this literature is highly fragmented: researchers in the studies reviewed above typically look at very specific or narrow sets of outcomes. This is despite the fact that network water sector interventions logically affect a range of outcomes, as documented above, that range from household health and well-being to productivity and net income improvements, and extend to beneficiary businesses and even the utilities providing such services.

We thus conclude this review with a brief summary of important challenges facing water infrastructure IEs. First, due to the nature of the scale of such interventions, which often consist of large, multi-pronged and overlapping activities, impacts can be difficult to attribute to specific investments. Relatedly, observed differences in outcomes for beneficiaries may be the stem from a combination of factors occurring through multiple channels. Some changes may be unrelated to the intervention itself, but may nonetheless mediate its outcomes in ways that are crucial to understand for sound policy-making. Second, the potential outcomes of urban water interventions are many and diverse both in type and in magnitude, which creates practical challenges related to measurement and the statistical power of the evaluation. Third, urban water interventions may cover all residents in an area (making it difficult to find a suitable comparison group), or alternatively target populations and locations that are very different from those who are untargeted [*Lokshin and Yemtsov*, 2003]. Fourth, infrastructure development alone may not deliver quality and reliability, particularly in the long term, if systems are poorly managed and operated [*Zérah*, 1998]. We reflect more critically on these challenges and their implications for IEs in the subsequent sections.

## 3. The central challenge facing water infrastructure IEs, and alternative methods for addressing it

To discuss the challenges identified above more formally, this section begins with a description of the central problem for researchers working to evaluate water and sanitation infrastructure (or really any development) interventions. We then provide an overview of the most commonly implemented approaches, offering reflections on their relative strengths and weaknesses. We close the section with a call for more mixed-methods IE designs that reveal a more complete set of consequences from such projects, before presenting a real-world application of that idea.

The basic problem facing any causal IE is to estimate the difference between what happened as a result of an intervention with what would have occurred in its absence. We present a simple framework to illustrate. Consider the outcome of interest $Y$ and a dichotomous indicator for exposure to an infrastructure intervention $I$, which takes a value of 1 if the intervention occurs, and is zero otherwise. The level of the outcome given "treatment" (or $I = 1$) is defined as $Y^{I}$, and $Y^{0}$ if $I = 0$. The impact caused by the infrastructure is then just the average treatment effect on the treated (ATT):

$$ATT = E(Y^1 - Y^0 | F = 1). \quad (1)$$

In this formulation, the major challenge is finding an unbiased method for approximating the counterfactual outcome $Y^0$ had the unit not been treated, which by definition cannot be observed directly. In a naïve design that simply compares observations that are targeted to receive the improvement against observations that do not, we observe:

$$E(Y^1 | F = 1) \text{ and } E(Y^0 | F = 0). \quad (2)$$

Taking the difference between these terms we observe that:

$$E(Y^1 | F = 1) - E(Y^0 | F = 0) = [E(Y^1 | F = 1) - E(Y^0 | F = 1)] + [E(Y^0 | F = 1) - E(Y^0 | F = 0)] = E(Y^1 - Y^0 | F = 1) + [E(Y^0 | F = 1) - E(Y^0 | F = 0)]. \quad (3)$$

This approximation based on treated and non-treated groups deviates from the ATT from equation (1) by the two final terms in equation 3, which represent the differences in outcomes across the comparison units in the absence of the treatment, and which helps clarify the evaluator's problem with selection bias. In nearly all cases, to achieve high impact, water and sanitation infrastructure is designed and delivered precisely to the groups that are hypothesized to benefit the most from such projects; this will lead to upward bias in the treated-untreated estimate of the ATT because $E(Y^0|F=1) > E(Y^0|F=0)$. On the other hand, if the infrastructure is targeted at groups that somehow benefit less from such investments (perhaps due to their lower income or access to other resources needed for development), the ATT estimate using Equation 3 may be biased downwards. Thus, one of the major challenges plaguing observational studies of the impacts of improved water and sanitation services on outcomes is that beneficiaries who receive improved services tend to be systematically different from those who do not (in terms of socio-economic status (SES), risk-altering behaviors, unobserved preferences for health, or myriad other ways), rendering comparisons of those with and without access suspect.

Rigorous IE methods are meant to minimize this risk of bias, by constructing more comparable groups of treated and untreated observations. In the simplest, most straightforward experimental case, we can assert that $[E(Y^0|F=1) = E(Y^0|F=0)]$, at least in a statistical sense. But random assignment is rarely feasible or desirable for large infrastructures, so a variety of quasi-experimental methods are more commonly utilized (Table 1). The idea is to leverage situations that lead to plausibly exogenous ("as if randomized") variation in exposure to infrastructure, or to apply statistical methods to isolate impacts by creating more comparable groups. A thorough review of how various approaches mitigate selection bias is beyond the scope of this article, but prominent references describing each are provided as a guide in Table 1. There we also specify and comment on three criteria, discussed also in what follows, that help clarify the appropriateness of these methods; comments pertinent to water infrastructure IEs are especially emphasized.

The first such criterion pertains to the validity of the design for proper causal inference, which refers to "internal validity", or the degree of confidence that measured effects are in fact due to the investment in question. Besides the obvious issue of selective targeting discussed above, infrastructure projects such as network water and sanitation give rise to several other common IE challenges. For one, even the most rigorous experimental or quasi-experimental designs may suffer from confounding by (unobserved) differences that are unknown to the evaluator. Such differences can arise due to unbalanced randomization (especially in small sample RCTs) or from a lack of full accounting for such differences (in quasi-experiments), when they correlate with targeting criteria. In difference in differences (DiD) designs, the major concerns are unobserved time-varying unobservables, which can give rise to non-parallel trends in the treatment and comparison groups. In matching designs, the conditional independence assumption requires that the variables that affect treatment assignment and treatment-specific outcomes are fully observable, such that any dependence between them is removed by conditioning on these factors [*Rosenbaum and Rubin*, 1985].

As complex and large interventions, infrastructure projects may also generate substantial spillovers, whereby untreated beneficiaries, who provide the counterfactual in an IE, are indirectly affected, typically as a result of behavioral responses. For example, a piped water improvement may induce some people to move – due to changes in asset values or individuals' desire to capture its benefits. This may reduce resource pressure in comparison areas, or lead utility personnel to adjust system operations in a way that affects those in unimproved areas. Spillovers can also play out through general equilibrium effects or distributional channels, for example, if informal water and sanitation service providers like water tankers respond to improvements and reduced service demand by lowering their prices, thereby generating benefits for untreated (comparison) consumers. An additional issue is selective attrition, whereby beneficiaries observed after an intervention may be more likely to provide data than those unserved (perhaps because beneficiaries are more willing to complete follow-up surveys than those in the comparison group). This is a problem when those providing data are not a random subsample of the larger population they were intended to represent, if, for example, only the most cooperative people participate.

The second criterion in Table 1 refers to the relevance of IE evidence that is produced. Here, evaluations generally place the greatest emphasis on "external validity", or on the generalizability of results to other groups and settings. Besides generalizability, it is important to emphasize relevance to answering decision makers' most crucial questions, which may be constrained by a particular method or set of methods. Related to these issues, Table 1 discusses: a) whether the IE pertains to an artefactual (largely researcher-constructed) or real-world (as implemented) situation; b) the population to which the evaluation evidence applies (i.e., whether it is all of those treated or some unique sub-sample); c) the likelihood that the produced evidence will be convincing to domain and IE experts; d) whether results are highly conditioned by assumptions that cannot be easily substantiated; and e) whether the evidence is likely to be precise from a statistical perspective. The answers to these questions are often related, with tradeoffs among them.

More specifically, real-world infrastructure interventions seldom give rise to situations that can be assessed with the methods deemed most convincing by IE experts (item *a* in the previous list). Conducting an RCT of piped water extension efforts, for example, is typically infeasible: It requires unprecedented coordination of costly investments across a large range of eligible locations, from which a treatment group is randomly selected. In addition, implementers often view randomization as arbitrary and poorly targeted, even in situations that give rise to opportunities for such a design. On the other hand, natural experiments, instrumental variables, and *ex post* regression are commonly applied to fully real world situations. The choice among these approaches demands careful thinking about the confounding and selection threats that face real world infrastructure evaluations. The corollary is that causal evidence derived from the latter methods is more often in doubt, relative to the "gold standard" RCT (issue *c*), whose results within a given experimental population are not conditioned by assumptions embedded within the deployed analytical methods (issue *d*). Nonetheless, critiques of RCTs often highlight the difficulties that arise in transferring their results across contexts, given the (usually) limited attention paid to contextual or institutional assumptions embedded in an experiment [*Peters et al.*, 2018]. The reality is that the vast majority of RCT evaluations are researcher-driven and artefactual, typically designed to test a narrow set of causal relationships.

The issues of the population for which the treatment effect is measured (issue *b*) and statistical precision (issue *e*) also have considerable importance. With the former, it is important to distinguish between estimates of intention to treat (ITT) impacts, which are representative of the entire targeted population and account for partial compliance or uptake, and treatment on the treated (ToT) impacts, which pertain only to those who take up the intervention, or to a subset of the treated. ToT impacts have more limited relevance when there is strong selection into treatment, or when the selection processes are unclear (e.g., as when instrumental variables methods are used) and give rise to a very specific local average treatment effect (LATE). ITT impacts, on the other hand, will be less transferrable if the processes that determine compliance and uptake rates do not generalize across settings. At the risk of oversimplifying an extensive discussion in the literature, we can assert that some methods produce more population-representative estimates (e.g., DiD, RCT, natural experiment), while others allow only treatment effects measurement in a very specific sub-population (e.g., instrumental variables (IV), regression discontinuity (RD)), with others (e.g., matching) falling somewhere in between.

Finally, the question of statistical precision for measuring impacts depends primarily on appropriate sampling rather than on analytical methods. But the truth is that water and sanitation infrastructure has the potential to influence a host of indicators through various channels. As such, relying on traditional single-method IEs will typically fail to capture important phenomena. For example, water and sewer interventions may benefit households in myriad ways (reduced coping costs, increased water consumption, improved health, greater productivity, etc.) when water scarcity is relieved [*Waddington et al.*, 2009; *Zwane and Kremer*, 2007], or benefits may flow mainly to businesses using water as an input to production [*Schwartz and Johnson*, 1992] or to a utility, if the latter collects more revenue from water sales [*Jeuland et al.*, 2020a]. Tracking impacts on these various groups may require very different data collection efforts, or even different methods, and may require combining natural and artefactual experiments [*Sawada*, 2015]. Moreover, many network interventions (such as piped water and sewer investments) have multiple components, and may therefore require complex designs that combine several research strategies.

The third criterion in Table 1 refers to practical and logistical considerations that emerge from the implementation of specific IE methods. The issues covered include a) costs; b) risks and adaptability to mitigate contamination arising from the spread of the treatment to comparison areas; c) need for coordination with project planners; d) interpretability and transparency of results; e) data needs from the pre-project period; and f) applicability or flexibility for covering complex interventions or theories of change. In general, the most rigorous methods – those that maximize internal validity – tend to be more costly, require greater coordination with planners, have greater pre-project data requirements (except for the RCT), and face the most severe contamination threats. This helps to explain why many studies deploying such methods are artefactual, researcher-driven ones, and why they are so rare in infrastructure evaluations. Moreover, there is frequently a tradeoff between transparency and flexibility or adaptability. This is demonstrated by comparing flexible IV, modeling and matching methods, which rely more heavily on analysts' judgement, to less flexible RCT, RD, and natural experiment approaches.

## 4. Design application: The Millennium Challenge Corporation Jordan Compact

With this general discussion in mind, we next turn to an evaluation application that illustrates the tensions and choices more concretely. This example consists of an integrated and holistic infrastructure improvement for urban water and sewer services in Zarqa, Jordan, a populous area in one of the most water-poor countries in the world [*Haddadin*, 2006; *Schyns et al.*, 2015].

### *4.1. Context*

In 2017, the year in which the investment we describe below was fully online, Jordan had per capita annual annual renewable water resources of 96 $m^3$, nearly 3.5 times below the average of other arid Middle Eastern countries, and more than 50 times lower than the world average [*FAO*, 2019]. Recent population growth, urbanization, an influx of refugees from Iraq and Syria, and extremely high water losses, have intensified this strain [*Hashemite Kingdom of Jordan*, 2016]. These growth trends and water losses have been particularly high in urban Zarqa, the region targeted by the infrastructure investment we consider. For example, non-revenue water (NRW) has been estimated to exceed 50% over the recent period, well above global and national averages [*Jeuland et al.*, 2020b].

Most of the population in Zarqa Governorate lives in Zarqa City (the second largest city in Jordan; population ~802,000) and Ruseifa (4th largest city; pop ~482,000), both of which lie in the Zarqa River Basin. Inhabitants of these cities have considerably lower income than those in neighboring Amman. Water supply is highly rationed; both households and businesses experience burdensome and routine water shortages and received water for only about 24 hours a week prior to the infrastructure improvements, in 2015 [*Orgill-Meyer et al.*, 2018]. In addition, though more than 99% of households have access to piped water, only about 70% had sewer connections prior to the new investments, and nearly 30% were thought to consume less than the minimum amount of water that the World Health Organization considers necessary (60L/capita-day) for personal hygiene and food safety [*MCC*, 2009].[1] In addition to water scarcity and reliability problems, households perceive the quality of utility water in Zarqa to be poor. Small businesses have lower reliance on the piped water (48%) and sewer (64%) networks than households, reportedly due to high connection fees but also unwillingness to pay tariffs for water consumption that cross-subsidize domestic users [*Jeuland et al.*, 2015]. Though Jordanians pay relatively high water tariffs compared to populations in neighboring countries, the utility in Zarqa (now called Miyahuna-Zarqa) does not fully cover its costs, and there has been long-standing resistance to water tariff increases [*Pitman*, 2004; *Sommaripa*, 2011].

### *4.2. The intervention: The Jordan Compact*

The Millennium Challenge Corporation – Millennium Challenge Account Jordan Compact (MCC-MCA-J JC, hereafter referred to as the JC) was developed to address Zarqa's most important water and sewer network deficiencies.[2] MCC largely funded the water investments and worked with the Government of Jordan (GoJ) throughout a detailed project identification and preparation period.[3] Construction began in 2014, implementation was largely successful, and the infrastructure handover had been completed as planned by the end of 2016. The final JC included three inter-linked projects:

(i) The Water Network Project (WNP) comprised two activities. The first was rehabilitation and restructuring of water supply transmission and distribution infrastructure in Zarqa and Ruseifa, and replacement of domestic water meters. The aim was to improve overall water efficiency through reduction of physical water losses and a transition from periodic distribution under high pressure to more consistent, gravity-fed distribution. The second activity, Water Smart Homes (WSH), aimed to improve household water storage and sanitation through a general outreach campaign, and to deliver infrastructure and technical assistance to the poor.

[1] Those lacking sewer connections usually rely on septic tanks that regularly require evacuation by tanker trucks, and also raise the risk of network contamination through infiltration of depressurized pipes.

[2] MCC is a U.S. foreign assistance agency that aims to fight global poverty by focusing on good policies, country ownership, and results. MCC provides time-limited grants to recipient countries, and administers the MCA. Thus, when a country is awarded a compact, it sets up a local MCA entity to manage and oversee all aspects of project implementation. Monitoring of funds is rigorous and transparent, often through independent fiscal agents.

[3] This preparation phase involved an analysis that was aimed at identifying key challenges to economic development in Jordan (a constraints analysis), identifying a technical solution to some of those challenges, and conducting feasibility and economic analyses of its anticipated impacts. More information on how MCC works can be found at https://www.mcc.gov/.

(ii) The Wastewater Network Project (WWNP) provided the expansion, rehabilitation and reinforcement of the wastewater network in Zarqa Governorate, and aimed to increase the capture of municipal wastewater and improve wastewater system efficiency.

(iii) The As-Samra Expansion Project (AEP) was designed to raise the capacity of the existing wastewater treatment plant serving this region, to allow treatment of additional wastewater volumes resulting from population growth in Amman and Zarqa, and from the aforementioned WNP and WWNP investments.

The stated goals of the JC – to reduce poverty and stimulate economic growth in Zarqa – were ambitious. These goals were to be achieved by increasing urban water supply, because water scarcity was deemed a key constraint inhibiting the area's development. *Ex ante*, planners believed that the investment would result in benefits through three main channels: two water substitution mechanisms, plus household cost savings from reduced dependence on septic tanks [*Jeuland et al.*, 2020b]. The first (primary) water substitution effect relied on capture and reuse of the additional volumes of treated wastewater resulting from greater water consumption and sewage collection in Zarqa, for irrigation in the Jordan Valley (JV), thereby facilitating reallocation of scarce freshwater sources for high value urban uses, while maintaining high value agriculture. Figure 1 provides a visual and qualitative representation of this primary substitution mechanism. The second (secondary) substitution effect posited that households in Zarqa would switch away from high-cost non-network water vendors (i.e., distribution shops and tankers) to greater use of cheaper network water, once urban water supply became more reliable.

### *4.3. Development of the IE design and rationale for the final approach*

While these three main benefit mechanisms of the JC are simple and intuitive, complex infrastructure projects can entail a slew of varying short and long-term effects that demand more careful examination. To consider the possibilities, the evaluation work began with consultations with a wide range of key stakeholders – representing different GoJ agencies, the implementing unit at the MCA-J, and the MCC. This engagement revealed that different parties had somewhat divergent perspectives on the project's most critical aspects, and that many potential sub-mechanisms and assumptions underpinned or supplemented the three main channels described above. For most real world IEs, participatory elicitation of a program theory of change and the context in which it operates, as conducted for this case, is of preeminent importance prior to design and data collection [*White*, 2011]. A literature review of prior related international experiences helped to identify additional potential channels of impact and clarify planners' assumptions, and resulted in a more comprehensive project logic (Figure 2).[4]

After development of this more complete project logic, the next step was to develop an IE that was practical, rigorous, and relevant to assessing the most important aspects of the investment. The two primary considerations in this design work were to specify the appropriate scope (e.g., populations and locations requiring study) and the nature of approximation of the non-intervention counterfactual. Given the integrated nature of the program's economic logic and its effects on multiple sectors as discussed above, the design endeavored to incorporate a diverse set of affected parties and geographies (Table 2 summarizes these populations; Figure 3 shows the geographic scope of data collection, Figure 4 clarifies the timeline of data collection events relative to infrastructure construction, and the supplementary materials provide additional details). Challenges pertaining to generalizability, statistical power, proper accounting for spillovers, flexibility to tackle changes in program implementation, and adequate contextualization and control for non-intervention confounders all helped to inform the final mix of evaluation components, methods, and data or measurement types. These choices were also subject to limitations on the overall evaluation budget.

In brief, the balancing of these aspects ultimately led to a design with three core components that combined several of the methods discussed in Table 1: i) DiD assessment of the infrastructure improvements' effects on urban populations (households and small enterprises, accounting for expected declines in meter accuracy that occur where service is intermittent) in urban Zarqa Governorate, with matching to ensure comparability of those with varying exposure to the JC; ii) counterfactual water balance modeling and DiD analysis of the infrastructure investments' effects on farmers (stemming from hypothesized increases in treated effluent inflows to the

[4] It is worth highlighting several additional channels and key assumptions that the evaluation team deemed to be potentially consequential at the time of design (Jeuland et al. 2020). First, some hypothesized that households might experience changes in health and general well-being arising from shifts in sourcing and in consumption levels or other behaviors targeted in the outreach campaigns organized around the JC. Second, it was deemed possible that the water utility would capture many of the benefits of the investment, with direct implications for cost recovery and utility performance, and indirect implications for public debt and longer-term quality of service delivered to households. Third, there was a sense that firm-level water decisions might change, and perhaps especially those of water shops and tankers, with effects on the distribution of economic outcomes. Finally, the switch in irrigation water sourcing was posited to have implications for farmers growing salinity-sensitive crops (e.g., citrus) in the Jordan Valley.

Jordan Valley); and iii) tracking of Zarqa utility performance over time, relative to that of two other similarly corporatized urban utilities in Jordan, in Amman and Aqaba. Additional details on the sampling and comparability, as well as approximate locations, of the IE's counterfactual populations are provided in the supplementary materials to this article (in Appendix 1). Needless to say, the use of such diverse combined methods scarcely appears in infrastructure evaluations present in the literature, and represents an important contribution of this application.

Several of the data collection activities shown in Table 2 were not central to the evaluation design but were included to deepen insights emerging from these 3 core IE components. Specifically, an endline cross-sectional survey of water vendors was included to both confirm and clarify distributional impacts related to the secondary substitution mechanism described in Section 4.2 (i.e., the posited shift away from expensive vendor water). Also, owing to the security crisis in neighboring Syria, Jordan received a major inflow of refugees that was contemporaneous with JC implementation; we carried out a refugee survey to understand the additional demand pressure this entailed, whether it was felt evenly across locations with differential exposure to the JC improvements, and to control for those differences, to the extent that they were important. Finally, we employed qualitative key informant interviews to better understand implementation fidelity and contextualize the main quantitative results.

### *4.4. Econometric and modeling analysis of impacts*

The DiD analyses for the three affected populations identified above – households in Zarqa, small enterprises in Zarqa (equation 4), and irrigators (equation 5) located downstream of the As Samra wastewater treatment plant – follow similar econometric specifications, which leverage panel data and net out time-invariant unobserved differences across groups that could otherwise affect or confound interpretation of the impacts of the JC:

$$\mathrm{Y_{ijt}} = \alpha + \gamma_1 \mathrm{T_t} + \delta \mathrm{d_j} + \kappa_t \mathrm{T_t} \cdot \mathrm{d_j} + \beta \mathrm{X_{ijt}} + \nu_i + \delta_{ijt}\,, \tag{4}$$

where $Y_{ijt}$ is the outcome of interest for household/farm/enterprise $i$ in zone $j$ at time $t$ (with $t = 0$ before intervention, and 1 after intervention); d is a vector of dummy variables that are equal to 1 if household/farm/enterprise $i$ is in treatment area $j$ and 0 otherwise; $T_t$ is a vector of dummy variables that are equal to 1 for the period in which the data was collected and 0 otherwise, $X_{ijt}$ is a vector of time-varying control variables that may affect the outcome for unit $i$ in zone $j$ at time

$t$; $v_i$ is a fixed effect for household/farm/enterprise $i$; and $\delta_{ijt}$ is a time-varying error term. The coefficient $\kappa_1$ measures the "treatment effect" or the change in outcome $Y$ for households/farms/enterprises in group $j$ relative to that in the omitted comparison group. This estimate is unbiased so long as the error term $\delta_{ijt}$ is uncorrelated with treatment status. To test robustness, we estimated models with and without individual unit fixed effects, and with and without time-varying controls $X_{ijt}$, though our preferred specifications include both of these. Standard errors were clustered at the level of the sampling cluster, generally the Census block.

The "treatment effect" measured among households and enterprises is derived from distinct sub-samples – for those exposed to different JC interventions and combinations – of treated and comparison observations that were randomly sampled from the zones in these two categories, selected for comparability using matching prior to implementation and baseline data collection. Specifically, zones specified as receiving each of the WNP only, WWNP only, and both WNP and WWNP improvements, were matched to separate samples of comparison zones using Census data and a 1-1 nearest-neighbor Propensity Score Matching (PSM) approach (see Appendix 1 for additional details on this sample construction). Thus, the effects of these 3 different intervention combinations are estimated from 3 separate regressions. Moreover, concern about utility-level spillovers motivated creation of comparison samples from within Zarqa (and subject to utility water supply re-optimization), and from neighboring areas in East Amman (that would not benefit from such spillovers).

The farm comparison is somewhat different, in that separate regressions are estimated to compare outcomes in each of five survey areas – selected on the basis of their varying levels of baseline and Compact-induced exposure to treated wastewater – to those in all other zones. These five areas include four different zones in the Jordan Valley (JV) as well as an additional zone located along the Zarqa River (Figure 3). In this regression, we are especially interested in changes in the mid-north of the JV (labelled zone JV2), where treated effluent as well as surface waters blended with treated effluent was introduced to irrigators for the first time during this period [*Morgan et al.*, 2021]. Impacts measured in this regression are not strictly limited to the impact of the JC, but rather represent the collective effects of several simultaneous changes in irrigation water sourcing by farmers.[5]

The outcomes $Y_{ijt}$ that we consider are the main factors that were expected – based on the full program theory of change – to evolve in the short to medium term, due to the JC investments. Among households, we focus on measures of the reliability of water supply, billed water consumption, perceptions of the quality of service (water pressure and safety), service disruptions and sewer backups, expenses for non-network water purchases, sewer connections, and pit-emptying costs. For firms, we consider changes in network water consumption and connections, sewerage, water expenses, and net profitability. Among farmers, we analyze water sourcing, perceptions of water quality, decisions about cropping across seasons, revenues, and profits.

We also note that the estimated parameter $\kappa$ in equation 1 represents an intention-to-treat (ITT) estimate; that is, it measures impacts on households and farmers whether they choose to comply with the intervention or not (Galasso et al. 2004). This is also the most relevant policy parameter, because it accounts for the behavioral responses of all populations exposed to the JC, whether or not they choose to take up specific infrastructures or improvements. For example, some households may not connect to new sewer pipes, some may not consume additional network water, and some farmers may reject blended wastewater in favor of groundwater or other alternatives.

A key threat to clean identification of impacts using any DiD approach is that time-varying unobserved differences across groups may be responsible for differential changes that are observed, rather than the JC investments *per se*. In the household survey, this threat is mitigated by matching households located in areas with improvements to households outside of those areas, which reduced *ex ante* differences within the sample, and by the verification of parallel pre-intervention trends in network water consumption across treated and comparison zones (see Appendix 1 for details). Due to data limitations, we are unable to establish parallel trends among farms in the JV, however, and emphasize that the natural experiment we exploit – driven by

[5] These changes included other complementary infrastructure works, including most notably the establishment of a new connector between the King Talal Dam that stores blended Zarqa River flow and treated wastewater, and the mid-north JV2 zone. They also include general increases throughout the region, in the availability of treated wastewater.

water system managers' replacement of freshwater supply to farmers with treated effluent – had already been underway for some time prior to the JC, per the GoJ policy encouraging that water substitution. In zone 2, however, the natural experiment we exploit does pertain to the new introduction of treated wastewater in irrigation, since this zone had not been connected to treated effluent prior to the JC. This is important for interpreting the trends we observe in the farm sample.

The other substantial element in the analysis of JC impacts is a water balance modeling exercise, parameterized based on comparative tracking of utility performance indicators over time, and from qualitative interviews with key stakeholders, as mentioned in Section 4.3 above. In the first of these, we modify an existing Jordan-wide water allocation model built using the Water Evaluation and Planning (WEAP) System software used by planners in the Ministry of Water and Irrigation (MWI) of the GoJ. Specifically, we create "with JC" and "without JC" scenarios that differ only in the following two parameters: a) the physical loss in water supply to beneficiaries in Zarqa Governorate; and b) the sewerage rate. We then specify these parameters based on the best empirical estimates of how these factors changed after JC implementation. This water balance analysis allows us to track systems-level changes in water supply to various users – irrigators in the Jordan Valley, urban consumers in Zarqa, and urban consumers elsewhere in Jordan, owing to the primary substitution mechanism, that serve to further contextualize and validate our IE estimates and the program theory of change. More details on the structure and assumptions of the water balance analysis can be found in Jeuland et al. [2021a].

## 5. Evaluation implementation and results

The multi-pronged evaluation strategy described in Section 4 provides evidence of several, but not all, of the changes expected from the Compact, and thereby provides only partial support for the originally hypothesized theory of change. Below, we summarize in succession the main results among beneficiaries in Zarqa (households and small enterprises), farmers in the Jordan Valley, and finally discuss the relative utility performance measures. Additional details can be found in Jeuland et al. [2020b].

*Impacts on households and small businesses in Zarqa*

In areas exposed to the Water Network Project, there is strong evidence of improvements in households' reporting of water pressure, increased network water use, and reduced complaints about water shortage, as well as somewhat weaker evidence of increases in the hours of water received from the piped network (Table 3, columns labelled "WNP" and "Both"). The specifications presented in Table 3 control for time and household fixed effects.[6] These positive changes notwithstanding, a key category of anticipated impacts that would increase net household income – reduced spending on expensive alternatives to utility water and on water overall – does not appear to have materialized.[7] There are several potential explanations for the lack of cost savings among households. First, the evaluation may have been underpowered to detect such impacts, since vendor purchases varied widely in the sample (notably, most of the relevant coefficient estimates, ranging from 5 to about 10 JD/month are negative but imprecisely measured, i.e., they are not statistically significant at conventionally reported levels). Perhaps more importantly, though, perceptions of the quality of networked water did not change following the investment. This, coupled with data showing that customers have generally negative perceptions of the quality of networked water relative to vendor water, seems to have manifested in a continuing preference for purchasing treated water containers (generally in 20L jugs) among many households, at least for drinking purposes.[8] Finally, the lack of clear substitution away from treated vendor water could have been due to an insufficient increase in the reliability of water supply to warrant shifting away from non-network alternatives; after all, supply remained intermittent. This highlights the importance of continuous water supply. Consistent with the apparent lack of substantial household substitution, then, there was no evidence of households saving time on water collection or procurement, following the intervention (results not shown).

Considering next the areas exposed to the Wastewater Network Project, there is strong evidence of an increase in households' likelihood of being connected to the piped sewer network and lower use of stand-alone cesspits, but only weak evidence of cost savings on emptying septic

---

[6] Alternative specifications that additionally control for two key time-varying factors – the number of refugees arriving in the sample area (a water demand shock) and a household-specific wealth index – yield very similar results. These alternative results are available upon request from the authors.

[7] This is also confirmed by the lack of evidence in the vendor survey for a decline in these activities over the period of the JC.

[8] Households reported that their stored drinking water – often purchased from shops – was safer on average than utility water. This perception is at odds with our own test results for *e. coli* and *total coliform*, which showed that network water from household taps contained *lower* contamination than stored drinking water, on average.

tanks, and no real evidence of reduced sewer backups, which are nonetheless infrequent (Table 3, columns labelled "WWNP" and "Both"). The cost savings outcome may take longer to manifest since these tanks take time to fill up, or else the evaluation may have been underpowered to detect it (we note that coefficients are negative in magnitude but imprecisely estimated). Finally, and in line with findings that Zarqa residents exposed to the improvement did not reduce water expenditures or shift away from expensive vendor sources, and that cost savings on septic tank emptying were unclear, there is no consistent evidence that the JC led to economic improvements among households treated by the WNP, WWNP, or the combination of these interventions (in terms of income, expenditure, assets). If anything, net income might have decreased slightly, perhaps due to the one-time cost of connecting to the sewer network.

Finally, there is some support for the idea that the investment generated positive spillovers within Zarqa, compared to neighboring areas in Amman supplied by a different water utility. This is apparent in the relatively larger impacts detected in the comparison with controls in Amman, for several key water supply variables: reported water pressure, hours of supply on days with water, and stronger reductions in water shortages. Impact estimates for the water supply improvements based on the Zarqa comparisons, therefore, appear biased downward due to positive spillovers. We would not expect similar spillovers through the wastewater improvements on most measured outcomes, except perhaps for sewer backups, but no impact on that outcome was detected. Consistent with this expectation, there do not appear to be consistently larger impacts relative to control households in Amman, for use of stand-alone wastewater systems, connection to the sewer, or savings on emptying septic tanks.

In contrast to the generally positive impacts observed on households due to the investments, enterprises did not appear to experience a similar increase in water consumption or service reliability. This is likely because rates of connection to the piped network remained low among this population (Table 4), and promoting connections among firms – which tend to be much higher in cost than household connections – was not an explicit objective of the Compact. In any case, it does not appear to be due to lack of statistical power, since changes for nearly all outcomes are small or even negative. Thus, enterprises did not appear to benefit the way households did from this investment.

*Impacts on farmers in the Jordan Valley*

Among farmers in Compact-affected areas, we observe increased supply and use of blended wastewater for irrigation, relative to areas that were outside the Compact-affected areas (Table 5). The areas most impacted by such increased flows are the JV2 and JV3 locations, followed by those in JV1 (which were newly receiving treated effluents from sources near Irbid) and the highland farms located along the Zarqa River. This greater water availability in turn mostly led to increased land area being irrigated, but was considered to reduce water quality in the areas receiving treated wastewater for the first time – JV1 and JV2. Perhaps owing to these water quality impacts (particularly the increased salinity of treated wastewater), the relative increase in water availability did not clearly translate into changes in the overall value of farm output or profits. The one exception to this pattern was in the highlands, where impacts were large and positive, and also seen for assets. The outcomes for highland famers suggests that irrigators located upstream of the Jordan Valley may have captured many of the expected benefits of the enhanced irrigation water availability, which had not been anticipated by Compact planners.

Meanwhile, further downstream in affected areas of the Jordan Valley, relative vegetable production and farm input costs increased as tree output decreased, suggesting a substitution away from saline- (or wastewater-) intolerant tree crops and into less sensitive horticulture, as shown in other work [*Morgan et al.*, 2021]. Finally, it may be reasonable to infer, based on the water balance analysis [*Jeuland et al.*, 2021a] and the fact that land values in these areas remained stable, that the investment increased the value of farm output relative to a no-Compact counterfactual with increasing water scarcity, in which agricultural activity would not have been similarly sustained.

*Evolution in the behavior and performance of the utility*

Comparing performance measures for the Zarqa water utility against other corporatized urban utilities in Jordan (namely in Amman and Aqaba), the JC appears to have improved measures of utility functionality. This elevated performance is reflected in evidence of sharply lower incidence of pipe failures (after an initial uptick in leaks during the infrastructure transition), declining administrative losses as measured by billing and collection efficiency, higher utility revenue collection, and reduced overall non-revenue water (NRW) measured in volume per subscriber per day (Table 6). As noted previously, the WNP component of the project primarily aimed to reduce physical water losses, but may also have reduced administrative losses through meter replacement. Meter testing on a small sample of meters in survey areas conducted prior to replacements indicated 25% under-measurement of consumption on average (results available upon request from the authors), and showed increased billed consumption following replacement in both rehabilitated and non-rehabilitated network areas (likely due to more accurate measurement of consumption). It is well known that the accuracy of mechanical meters deteriorates in intermittent systems, as a function of age, water pressure fluctuations, and the presence of air in the network [*Walter et al.*, 2017].

That said, the evaluation cannot differentiate the effect of the JC investments from that of the contemporaneous utility corporatization reform, which likely also affected indicators of functionality and efficiency. Notably, however, NRW declines resulting from the intervention did underperform relative to JC targets and expectations, especially when measured in percentage terms. This may have been driven by several factors including incomplete isolation of rehabilitated network areas, increases in water supply among portions of the network outside the intervention area (or spillovers from greater water allocation by managers, to better meet demand in unimproved water scarce areas), or illicit water use. Key informant interviews with utility personnel helped to support these explanations; several senior utility operators noted that water "rotations" – periods of servive to different neighborhoods within Zarqa – were adjusted following the Compact to allow the benefits of the improvements to be shared more equally across households in improved and unimproved areas. Another area of potential spillovers was in maintenance; because servicing needs declined in WNP areas, effort could be reallocated to non-Compact areas [*Jeuland et al.*, 2020b].

Finally, rising operating costs meant that short-term improvements in the Operating Cost Recovery Ratio between 2014 and 2016 were not sustained. Higher operating costs were largely driven by increased per-unit volume energy costs and the cost of additional wastewater management, as well as increase of imported water pumped from the far-away Disi aquifer.

## 6. Discussion and conclusions

This paper provides a contemporary discussion of the problem of impact evaluations of large water infrastructure projects. The contemporary perspective is important in light of concerns over the validity and plausibility of previous assessments [*Angrist and Pischke*, 2010], the particular challenges of these long-lived projects, the complex causal chains that determine their impacts [*Polasky et al.*, 2019; *Tallis et al.*, 2019], and the relatively limited evidence on which to base future similar investment decisions. To that end, we described the core challenge facing researchers aiming to establish the causal nature of observed changes in outcomes, and then proposed a set of three key criteria – validity, relevance and practicality – that researchers working in this domain should consider when designing and implementing evaluation research. As there are tradeoffs across these three criteria, scholars must thoughtfully weigh the particular pros and cons, and opportunities and vulnerabilities of different methods, to advance knowledge and achieve greater policy impact.

We demonstrated the approach with discussion of a multi-faceted evaluation implemented to assess the outcomes attributable to the Jordan Compact, a highly integrated project that combined water and wastewater investments, with the goal of reducing water scarcity and enhancing economic opportunity in Zarqa, Jordan. Overall, the evaluation found a positive net economic impact, though impacts in some domains fell short of expectations [*Jeuland et al.*, 2020b]. On the positive side, there was clear evidence of improved service and reduced water loss where the water network was rehabilitated. Investments in sewerage meanwhile increased the number of household connections and reduced the incidence of sewer backups. Finally, the water savings and increased wastewater capture in Zarqa led to increased supply of treated wastewater to irrigators along the Zarqa River and in the Jordan Valley, which is supporting a profit-neutral shift away from cultivation of water-intensive and salinity-sensitive citrus trees, and toward vegetable field crops. This shifting supply, in turn, is freeing up freshwater previously allocated to irrigation for higher-value uses in urban areas.

On the negative side, however, there was no evidence that the JC improvements reduced expenses for high cost non-network water, which was a key expected channel of benefits to households. Such behavioral changes may take longer to manifest, but survey evidence suggests that households maintained their skepticism about the safety of piped network water (Orgill-Meyer et al. 2018). Moreover, small and medium enterprises did not benefit from the investments, likely owing to their continued low rate of connection to the piped water and sewer system. Finally, improvements in water supply reliability to rehabilitated areas were somewhat diminished by utility adjustments that increased supply to areas outside the intervention area. Ultimately, the mixed realized benefits of the investments in Zarqa may therefore perpetuate a low-equilibrium trap that plagues water utilities in developing countries and is difficult to resolve sustainably [*Jeuland*, 2022]. Specifically, because consumers do not trust water utilities, they often resist paying tariffs that allow full cost-recovery, or engage in theft from the water network. Such behaviors compromise the utility's ability to invest in long-term maintenance and reliability.

Overall, the evaluation contributes to a relatively thin literature on the economic benefits of investments in urban water and sewer systems, which represent one of the most important quasi-public goods provided by governments in low- and middle-income countries. Examined through the lens of the economic analysis justifying the JC, the success of the intervention was largely contingent on the effective substitution, for irrigation uses, of recycled wastewater for freshwater supplies. The conditions leading to successful substitution of the type observed in this case have rarely been documented [*Jeuland*, 2012], and this represents an important new contribution of this evaluation to the literature. Indeed, most studies in the Middle East and globally have rather highlighted the relatively limited success of attempts to increase wastewater reuse owing to salinity or other water quality concerns [*Carr et al.*, 2011; *Jeuland*, 2015].

Perhaps more significantly, though, the JC example serves as a useful demonstration of a number of evaluation issues and challenges discussed in this paper, that warrant additional research and policy engagement. First, researchers seeking to carry out policy-relevant evaluations of large infrastructure investments must work harder to engage with project planners to understand these interventions' complete theories of change and to track the most important set of anticipated impacts. The eventual evaluation design in this case, with data analysis focused on a range of stakeholder groups and both pecuniary and non-pecuniary quality of life outcomes, could then be crafted to allow a critical appraisal of planners' most critical assumptions, rather than focusing on a narrow set of questions (e.g., whether the investment reduced diarrheal disease incidence). Second, and relatedly, it allowed for nuanced understanding of the distributional effects of the investment. Specific and documented failures, for example related to water consumers lack of confidence in utility water in this case, can then inform development of future remedies to address them. Thus, Zarqa policy-makers might consider investing to convince users of the safety of network water, given our results showing that this water may be safer than more expensive water purchased from vendors [*Orgill-Meyer et al.*, 2018].

Third, large infrastructure projects nearly always have spillovers and overlapping or systems-level impacts, which good evaluations must try to anticipate. The JC project is a particularly salient example of this, with its highly integrated design. Owing to a fairly sophisticated understanding of the project theory of change, the evaluation therefore worked to combine several complementary data collection and quasi-experimental analytical techniques to provide a comprehensive view of the investment impacts. Importantly, this also motivated construction of two alternative comparison samples (one from unimproved areas in Zarqa, which were highly subject to infrastructure spillovers but highly comparable to treated areas, and one from neighboring areas in Amman Governorate, which were not subject to spillovers but also less comparable). This approach allowed for learning about both spillovers and impacts, which is highly valuable for policy-making.

Fourth, there are often important tradeoffs between theoretical internal validity (with RCTs serving as a gold standard) and risks of contamination of the evaluation, which occurs when areas identified *ex ante* for treatment are not improved, or when evaluation comparison areas end up receiving improvements. With the JC, such threats were borne out, as one planned neighborhood ended up not being rehabilitated due to local politics, and Compact implementers seeking to exhaust the implementation budget also worked to extend the improvements beyond the originally planned areas. Throughout the process, implementer-evaluator coordination helped to minimize the risks of contamination, and clarify their severity, such that the design's integrity was ultimately maintained. Upon reflection, much of the success for this was due to the communication and trust between the two parties, and stemmed from implementers' appreciation for the evaluation's efforts to create a cost-effective and pragmatic evaluation that was respectful of the original project design and objectives.

Finally, a critical limitation of this specific evaluation that is highly relevant in the context of infrastructure projects is its relatively short, four-year time frame (from baseline to endline). Indeed, a focus on short-run and medium-term impacts is a challenge to IE in infrastructure domains that may take many years to realize benefits. Even as there is also substantial uncertainty about the long-term performance of such investments and the continued evolution of beneficiary behavior, however, which should motivate longer-term work, the validity of the treatment-control comparisons become increasingly tenuous as time goes on. Future research on the JC's effects should attempt to verify the persistence of the short-term changes measured here, and the project's distributional consequences. For example, farmers in the Northern Jordan Valley will likely continue to adjust to the shift in water supply over time, and households and businesses may gain confidence in the quality of network water or persist in purchasing more expensive alternatives. Data from Miyahuna-Zarqa could help to obtain a more complete picture of the effects on the utility, and whether short-term changes in NRW continue or are reversed.

## References


Aggarwal, S. (2018), Do rural roads create pathways out of poverty? Evidence from India, *Journal of Development Economics*, *133*, 375-395.

Alsan, M., and C. Goldin (2015), Watersheds in infant mortality: The role of effective water and sewerage infrastructure, 1880 to 1915, National Bureau of Economic Research.

Alsan, M., and C. Goldin (2019), Watersheds in child mortality: The role of effective water and sewerage infrastructure, 1880–1920, *Journal of Political Economy*, *127*(2), 586-638.

Angrist, J. D., and J.-S. Pischke (2010), The credibility revolution in empirical economics: How better research design is taking the con out of econometrics, *Journal of economic perspectives*, *24*(2), 3-30.

Asher, S., and P. Novosad (2020), Rural roads and local economic development, *American economic review*, *110*(3), 797-823.

Blimpo, M. P., R. Harding, and L. Wantchekon (2013), Public investment in rural infrastructure: Some political economy considerations, *Journal of African Economies*, *22*(suppl_2), ii57-ii83.

Bos, K., D. Chaplin, and A. Mamun (2018), Benefits and challenges of expanding grid electricity in Africa: A review of rigorous evidence on household impacts in developing countries, *Energy for sustainable development*, *44*, 64-77.

Bothwell, L. E., J. A. Greene, S. H. Podolsky, and D. S. Jones (2016), Assessing the gold standard—lessons from the history of RCTs, *N Engl J Med*, *374*(22), 2175-2181.

Brakarz, J., and L. Jaitman (2013), Evaluation of slum upgrading programs: Literature review and methodological approaches, *Inter-American Development Bank*.

Brown, J., V. T. Hien, L. McMahan, M. W. Jenkins, L. Thie, K. Liang, E. Printy, and M. D. Sobsey (2013), Relative benefits of on-plot water supply over other 'improved'sources in rural Vietnam, *Tropical medicine & international health*, *18*(1), 65-74.

Bucknall, J., A. Kremer, T. Allan, J. Berkoff, N. Abu-Ata, M. Jarosewich-Holder, U. Deichmann, S. Dasgupta, R. Bouhamidi, and V. Ipe (2007), *Making the most of scarcity: accountability for better water management results in the Middle East and North Africa*, World Bank Publications.

Burt, Z., A. Ercümen, N. Billava, and I. Ray (2018), From intermittent to continuous service: costs, benefits, equity and sustainability of water system reforms in Hubli-Dharwad, India, *World Development*, *109*, 121-133.

Carr, G., R. B. Potter, and S. Nortcliff (2011), Water reuse for irrigation in Jordan: Perceptions of water quality among farmers, *Agricultural Water Management*, *98*(5), 847-854.

Casaburi, L., R. Glennerster, and T. Suri (2013), Rural roads and intermediated trade: Regression discontinuity evidence from Sierra Leone, *Available at SSRN 2161643*.

Cicchetti, C. J., V. K. Smith, and J. Carson (1975), An economic analysis of water resource investments and regional economic growth, *Water Resources Research*, *11*(1), 1-6.

Cox, P. T., C. W. Grover, and B. Siskin (1971), Effect of water resource investment on economic growth, *Water Resources Research*, *7*(1), 32-38.

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

Davis, J., A. Kang, J. Vincent, and D. Whittington (2001), How important is improved water infrastructure to microenterprises? Evidence from Uganda, *World Development*, *29*(10), 1753-1767.

Deaton, A. (2009), Instruments of Development: Randomisation in the Tropics, and the Search for the Elusive Keys to Economic Development, paper presented at Proceedings of the British Academy.

Deaton, A. (2019), Randomization in the tropics revisited: a theme and eleven variations, in *Randomized controlled trials in the field of development: A critical perspective.* , edited by F. Bédécarrats and I. Guérin, Oxford University Press.

Dercon, S., D. O. Gilligan, J. Hoddinott, and T. Woldehanna (2009), The impact of agricultural extension and roads on poverty and consumption growth in fifteen Ethiopian villages, *American Journal of Agricultural Economics*, *91*(4), 1007-1021.

Devoto, F., E. Duflo, P. Dupas, W. Pariente, and V. Pons (2011), Happiness on tap: Piped water adoption in urban Morocco, National Bureau of Economic Research.

Dinkelman, T. (2011), The effects of rural electrification on employment: New evidence from South Africa, *American Economic Review*, *101*(7), 3078-3108.

Duflo, E., R. Glennerster, and M. Kremer (2007), Using randomization in development economics research: A toolkit, *Handbook of development economics*, *4*, 3895-3962.

Duflo, E., M. Greenstone, R. Guiteras, and T. Clasen (2015), Toilets can work: Short and medium run health impacts of addressing complementarities and externalities in water and sanitation, National Bureau of Economic Research.

Ercumen, A., B. F. Arnold, E. Kumpel, Z. Burt, I. Ray, K. Nelson, and J. M. Colford Jr (2015), Upgrading a piped water supply from intermittent to continuous delivery and association with waterborne illness: a matched cohort study in urban India, *PLoS medicine*, *12*(10), e1001892.

Esrey, S. A. (1996), Water, waste, and well-being: a multicountry study, *American journal of epidemiology*, *143*(6), 608-623.

Estache, A. (2010), A survey of impact evaluations of infrastructure projects, programs and policies, *European Centre for Advanced Research in Economics (ECARES) Working Paper*, *5*, 2010.

FAO (2019), Aquastat Database. Available at: http://www.fao.org/nr/water/aquastat/data/query/index.html, edited by U. N. Food and Agriculture Organization, Rome. .

Foster, V., and C. M. Briceño-Garmendia (2009), *Africa's infrastructure: a time for transformation*, The World Bank.

Galdo, V., and B. Briceño (2011), Evaluating the impact on child mortality of a water supply and sewerage expansion in Quito: is water enough?, Inter-American Development Bank.

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

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

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

Ghani, E., A. G. Goswami, and W. R. Kerr (2016), Highway to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing, *The Economic Journal*, *126*(591), 317-357.

Haddadin, M. J. (2006), *Water resources in Jordan: evolving policies for development, the environment, and conflict resolution*, Resources for the Future.

Hammer, J., and D. Spears (2016), Village sanitation and child health: Effects and external validity in a randomized field experiment in rural India, *Journal of health economics*, *48*, 135-148.

Hanemann, W. (2006), The economic conception of water, in *Water Crisis: Myth or Reality?*, edited, p. 61 pp.

Hashemite Kingdom of Jordan (2016), National Water Strategy of Jordan, 2016 – 2025, Ministry of Water and Irrigation, Amman.

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

Jedwab, R., and A. Storeygard (2022), The average and heterogeneous effects of transportation investments: Evidence from Sub-Saharan Africa 1960–2010, *Journal of the European Economic Association*, *20*(1), 1-38.

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

Jeuland, M. (2015), Challenges to wastewater reuse in the Middle East and North Africa, *Middle East Development Journal*, *7*(1), 1-25.

Jeuland, M. (2022), Systems Thinking for More Holistic Analysis of Low-and Middle-Income Country Water Utility Problems and Solutions, *Water Economics and Policy*, 2271002.

Jeuland, M., T. Dauwalter, and O. Hopkins (2020a), The economics of institutional changes in the water sector: Methods, evidence, and a call for systems thinking, edited by D. U. W. Paper, Durham, USA.

Jeuland, M., M. Moffa, and A. Alfarra (2021a), Water savings from urban infrastructure improvement and wastewater reuse: evidence from Jordan, *International Journal of Water Resources Development*, 1-20.

Jeuland, M., J. Orgill, S. Alikhan, and N. Cutler (2015), Impact Evaluation of the MCA Jordan Compact Baseline Results of Component 1 – Water Network and Wastewater Network Projects, Social Impact, Inc., Washington, DC.

Jeuland, M., M. Pucilowski, D. Hudner, A. Smith, T. Thompson, C. Seybolt, J. Orgill-Meyer, S. Morgan, and A. Wyatt (2020b), Impact Evaluation of the MCA Jordan Compact: Endline Report (Draft), Millennium Challenge Corporation, Washington, DC.

Jeuland, M., T. R. Fetter, Y. Li, S. K. Pattanayak, F. Usmani, R. A. Bluffstone, C. Chávez, H. Girardeau, S. Hassen, and P. Jagger (2021b), Is energy the golden thread? A systematic review of the impacts of modern and traditional energy use in low-and middle-income countries, *Renewable and Sustainable Energy Reviews*, *135*, 110406.

Kesztenbaum, L., and J.-L. Rosenthal (2017), Sewers' diffusion and the decline of mortality: The case of Paris, 1880–1914, *Journal of Urban Economics*, *98*, 174-186.

Klasen, S., T. Lechtenfeld, K. Meier, and J. Rieckmann (2012), Benefits trickling away: the health impact of extending access to piped water and sanitation in urban Yemen, *Journal of Development Effectiveness*, *4*(4), 537-565.

Klassert, C., E. Gawel, K. Sigel, and B. Klauer (2018), Sustainable transformation of urban water infrastructure in Amman, Jordan–meeting residential water demand in the face of deficient public supply and alternative private water markets, in *Urban Transformations*, edited, pp. 93-115, Springer.

Kolahi, A.-A., A. Rastegarpour, and M.-R. Sohrabi (2009), The impact of an urban sewerage system on childhood diarrhoea in Tehran, Iran: a concurrent control field trial, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *103*(5), 500-505.

Koundouri, P., C. Nauges, and V. Tzouvelekas (2006), Technology adoption under production uncertainty: theory and application to irrigation technology, *American Journal of Agricultural Economics*, *88*(3), 657-670.

Lee, K., E. Miguel, and C. Wolfram (2020), Does household electrification supercharge economic development?, *Journal of Economic Perspectives*, *34*(1), 122-144.

Lipscomb, M., A. M. Mobarak, and T. Barham (2013), Development effects of electrification: Evidence from the topographic placement of hydropower plants in Brazil, *American Economic Journal: Applied Economics*, *5*(2), 200-231.

Liu, J., H. Yang, S. N. Gosling, M. Kummu, M. Flörke, S. Pfister, N. Hanasaki, Y. Wada, X. Zhang, and C. Zheng (2017), Water scarcity assessments in the past, present, and future, *Earth's future*, *5*(6), 545-559.

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

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

McRae, S. (2015), Infrastructure quality and the subsidy trap, *American Economic Review*, *105*(1), 35-66.

Meeks, R. C. (2017), Water works the economic impact of water infrastructure, *Journal of Human Resources*, *52*(4), 1119-1153.

Molden, D. (2013), *Water for food water for life: A comprehensive assessment of water management in agriculture*, Routledge.

Moraes, L., J. A. Cancio, S. Cairncross, and S. Huttly (2003), Impact of drainage and sewerage on diarrhoea in poor urban areas in Salvador, Brazil, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *97*(2), 153-158.

Morgan, S., J. Baker, J. Orgill-Meyer, and M. Jeuland (2021), Valuing water quality with adaptation: Evidence from a natural experiment in Jordan, *Water Economics and Policy*.

Mu, R., and D. Van de Walle (2011), Rural roads and local market development in Vietnam, *The Journal of Development Studies*, *47*(5), 709-734.

Nguyen Viet, C., and T. Vu (2013), The impact of piped water on household welfare: evidence from Vietnam, *Journal of Environmental Planning and Management*, *56*(9), 1332-1358.

Orgill-Meyer, J., M. Jeuland, J. Albert, and N. Cutler (2018), Comparing contingent valuation and averting expenditure estimates of the costs of irregular water supply, *Ecological Economics*, *146*, 250-264.

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

Peters, J., and M. Sievert (2016), Impacts of rural electrification revisited–the African context, *Journal of Development Effectiveness*, *8*(3), 327-345.

Peters, J., J. Langbein, and G. Roberts (2018), Generalization in the tropics–development policy, randomized controlled trials, and external validity, *The World Bank Research Observer*, *33*(1), 34-64.

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

Pitman, G. (2004), Jordan: An evaluation of bank assistance for water development and management. A country assistance evaluation, *World Bank, Washington, DC*.

Polasky, S., C. L. Kling, S. A. Levin, S. R. Carpenter, G. C. Daily, P. R. Ehrlich, G. M. Heal, and J. Lubchenco (2019), Role of economics in analyzing the environment and sustainable development, *Proceedings of the National Academy of Sciences*, *116*(12), 5233-5238.

Pradhan, M., and L. B. Rawlings (2002), The impact and targeting of social infrastructure investments: Lessons from the Nicaraguan Social Fund, *The World Bank Economic Review*, *16*(2), 275-295.

Raitzer, D. A., N. Blöndal, and J. Sibal (2019), *Impact evaluation of energy interventions: A review of the evidence*, Asian Development Bank.

Ravallion, M. (2018), Should the randomistas (continue to) rule, *Center for Global Development working paper*, *492*.

Rijsberman, F. R. (2006), Water scarcity: fact or fiction?, *Agricultural water management*, *80*(1-3), 5-22.

Rosenbaum, P. R., and D. B. Rubin (1985), Constructing a control group using multivariate matched sampling methods that incorporate the propensity score, *The American Statistician*, *39*(1), 33-38.

Royal Commission for Water (2009), Water for Life: Jordan's Water Strategy 2008-2022, Amman, Jordan.

Sawada, Y. (2015), The impacts of infrastructure in development: A selective survey.

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

Schyns, J., A. Hamaideh, A. Hoekstra, M. Mekonnen, and M. Schyns (2015), Mitigating the risk of extreme water scarcity and dependency: the case of Jordan, *Water*, *7*(10), 5705-5730.

Soares, F., and Y. Soares (2005), The socio-economic impact of favela-bairro: what do the data say?, *Inter-American Development Bank, Washington DC*.

Sommaripa, L. (2011), Jordan Fiscal Reform Project II: Water Public Expenditure Perspectives Working Paper, *United States Agency International Development, Washington, DC*.

Tallis, H., K. Kreis, L. Olander, C. Ringler, D. Ameyaw, M. E. Borsuk, D. Fletschner, E. Game, D. O. Gilligan, and M. Jeuland (2019), Aligning evidence generation and use across health, development, and environment, *Current Opinion in Environmental Sustainability*, *39*, 81-93.

Uchimura, K., and H. Gao (1993), The importance of infrastructure on economic development, *Latin America and Caribbean Regional Office, World Bank, Washington DC*.

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

Walter, D., M. Mastaller, and P. Klingel (2017), Accuracy of single-jet water meters during filling of the pipe network in intermittent water supply, *Urban Water Journal*, *14*(10), 991-998.

White, H. (2011), Achieving high-quality impact evaluation design through mixed methods: the case of infrastructure, *Journal of development effectiveness*, *3*(1), 131-144.

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

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

Wolf, J., S. Hubbard, M. Brauer, A. Ambelu, B. F. Arnold, R. Bain, V. Bauza, J. Brown, B. A. Caruso, and T. Clasen (2022), Effectiveness of interventions to improve drinking water, sanitation, and handwashing with soap on risk of diarrhoeal disease in children in low-income and middle-income settings: a systematic review and meta-analysis, *The Lancet*, *400*(10345), 48-59.

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

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

## 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 | | | | |
| 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<br>• *Spillovers* (violation of the stable unit treatment value assumption, or SUTVA), whereby some units benefit as a result of other units' uptake.<br>• Vulnerable to *selective attrition* | • Typically *artefactual*, w/ limited evaluation questions<br>• Treatment effect can be *representative*<br>• "Gold standard" for causal researchers<br>• Results are *not conditioned* by assumptions<br>• *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<br>• *Contamination risk*: Moderate, as pressure to help "untreated" units increases over time<br>• *Coordination*: Mainly pertains to maintaining integrity of randomization<br>• *Interpretation*: Intuitive and highly transparent<br>• *Pre-intervention data needs*: Low to none<br>• *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<br>• *Spillovers* (i.e., violation of SUTVA) outside of treatment area | • Evidence arises directly from the *real world*<br>• Treatment effect is *representative* but contingent on natural experiment conditions<br>• Generally accepted by researchers<br>• Results are *not conditioned* by assumptions<br>• *Statistical power*: Difficult to | • *Cost*: Low to moderate, depending on data collection needs<br>• *Contamination risk*: Low<br>• *Coordination*: Moderate; mainly in combining with other methods (DiD) to strengthen validity<br>• *Interpretation*: Intuitive but not always transparent<br>• *Pre-intervention data needs*: Low to none<br>• *Flexibility to adapt*: Impossible |

| | | | | |
|---|---|---|---|---|
| | | | anticipate ex ante | • *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<br>• *Spillovers* (i.e., violation of SUTVA)<br>• Vulnerable to *selective attrition* | • Evidence arises directly from the *real world*<br>• Treatment effect is usually *representative* (unless combined w/other methods)<br>• Generally accepted by researchers, subject to showing parallel trends<br>• Results are *not conditioned* by assumptions<br>• *Statistical power* is a design feature | • *Cost*: Moderate to high, depending on data collection needs<br>• *Contamination risk*: Moderate to high<br>• *Coordination*: Moderate; mainly in combining with other methods (matching) to strengthen validity<br>• *Interpretation*: Intuitive and transparent<br>• *Pre-intervention data needs*: Moderate to high (parallel trends)<br>• *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<br>• *Spillovers* (i.e., violation of SUTVA) | • Evidence arises directly from the *real world*<br>• Treatment effect only applies to units with suitable comparisons (common support region)<br>• Researchers are often skeptical that the CIA has been met<br>• Results are *conditioned* by assumptions of the matching algorithm<br>• *Statistical power* is a design feature | • *Cost*: Moderate to high, depending on data collection needs<br>• *Contamination risk*: High<br>• *Coordination*: Moderate; mainly in combining with other methods (DiD) to strengthen validity<br>• *Interpretation*: Intuitive, but matching may lack transparency<br>• *Pre-intervention data needs*: Moderate (matching)<br>• *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<br>• *Spillovers* (i.e., violation of SUTVA) | • Evidence arises directly from the *real world*<br>• Treatment effect (LATE) is not representative, and not always for the most relevant population<br>• Researchers are often skeptical about exclusion restriction<br>• Results are *conditioned* by exogeneity assumptions<br>• *Statistical power* is often | • *Cost*: Low to moderate, depending on data collection needs<br>• *Contamination risk*: Not applicable<br>• *Coordination*: Low<br>• *Interpretation*: Unintuitive, lacks transparency<br>• *Pre-intervention data needs*: Low<br>• *Flexibility to adapt*: High<br>• *Other*: Suitable IV may not exist |

| | | | | |
|---|---|---|---|---|
| | | | reduced by 2-stage estimation | |
| 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<br>• *Spillovers* (i.e., violation of SUTVA)<br>• Vulnerable to *selective attrition* | • Evidence arises directly from the *real world*<br>• Treatment effect is limited to units very near the discontinuity<br>• Generally accepted by researchers<br>• Results are *conditioned* on proximity to eligibility cutoff<br>• Statistical power may be limited | • *Cost*: Low to moderate, depending on data collection needs<br>• *Contamination risk*: Moderate, depending on rigor with which eligibility is assessed<br>• *Coordination*: Low<br>• *Interpretation*: Intuitive, but transparency may be lacking due to definition of the RD bandwidth<br>• *Pre-intervention data needs*: Low<br>• *Flexibility to adapt*: Low |
| <u>Other</u> | | | | |
| *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<br>• *Confounding* by unobservables<br>• *Spillovers* (i.e., violation of SUTVA) | • Evidence arises directly from the *real world*<br>• Treatment effect is usually representative<br>• Causal researchers are typically highly skeptical of results<br>• Results are *conditioned* on controls<br>• *Statistical power*: Difficult to anticipate ex ante | • *Cost*: Low to moderate, depending on data collection needs<br>• *Contamination risk*: Not applicable<br>• *Coordination*: Low<br>• *Interpretation*: Intuitive, but transparency may be lacking (contingent on choice of controls)<br>• *Pre-intervention data needs*: None<br>• *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<br>• Treatment effect is usually representative, but may not align with policy-maker priorities and needs<br>• Not widely used by causal social science researchers, who are wary of over-calibration<br>• Results are *conditioned* on model assumptions<br>• *Statistical power*: Not applicable | • *Cost*: Low<br>• *Contamination risk*: Not applicable<br>• *Coordination*: Low<br>• *Interpretation*: Not intuitive and not always transparent (requires interdisciplinary expertise)<br>• *Pre-intervention data needs*: Moderate to high, depending on calibration needs<br>• *Flexibility to adapt*: High<br>• *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<br>• Random sampling within sample geocodes<br>• Replacements selected from sample geocodes | WNP only – WNP only control<br>WWNP only – WWNP only control<br>WNP+WWNP – WNP+WWNP control<br><br>Distinct control groups from:<br>• Zarqa<br>• Amman | Representative of sample geocodes at baseline, based on comparisons to Census and other sources | 1. 3359<br>2. 3416<br>3. 3596<br>4. 3662 |
| Enterprise | 2-wave Panel | Zarqa and Amman (from DoS + household referrals) | • Same geocodes as household sample<br>• Random selection within sample geocodes<br>• Referrals for informal enterprises<br>• 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<br>Informal enterprises likely under-represented (due to low referral rates) | 1. 345<br>2. 418 |
| Farm | 3-wave Panel | Jordan Valley and highlands (from DoS) | • Survey zones selected based on expected differences in exposure to treated wastewater<br>• Random selection in sample zones<br>• Replacements selected within zones | Five locations:<br>• Highlands u/s KTD (↑ river flow)<br>• JV1 North (↑ Non-Compact WW)<br>• JV2 Mid-North (↑ Compact WW)<br>• JV3 North-Central (↑ Compact WW)<br>• JV4 South-Central (little change in WW) | Representative of sample zones | 1. 551<br>2. 539<br>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<br>• Random sampling by treatment status<br>• Referrals for unregistered refugees | Treatment status:<br>• WNP only<br>• WWNP only<br>• Both WNP and WWNP<br>• Controls in Zarqa<br>• Control Amman | Representative of registered population in sample areas<br>Unregistered population likely under-represented (due to low referral rates) | 1617 |
| Water vendor | Single cross-section | Shops: Ministry of Health list + canvassing<br>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)<br>• Random sample of meters within selected zones | Compact and non-Compact zones | Not representative | 1. 37<br>2. 223 |
| Water loss testing | Single cross-section | Canvassing of land plots in selected areas | • "Well isolated" zones selected (as suggested by utility)<br>• Comparison of meter registered data to bulk meter inflow<br>• 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<br>• 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 | **+0.63**** (0.30) | n.a. | -0.29 (0.34) | +0.32 (0.27) | n.a. | -0.58 (0.36) |
| 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) |
| 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 $\kappa_t$ 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 | -0.05 (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) |
| 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 $\kappa_t$ 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) | -0.14 (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) |
| Farm land value (JD) | **+62124***** (18012) | +25717 (19918) | **-61272***** (19275) | -6484 (21277) | **-42200*** (22611) |

Notes: All estimates are difference-in-differences estimates for coefficient $\kappa_t$ 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 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Data Source |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Reduced NRW | Indicator: Total losses | | | | | | | | | | | | | |
| | Total losses; network (%) | Aqaba | 20.6 | 23.9 | 21.1 | 22.2 | 25.9 | 25.8 | 27.8 | 27.7 | 24.8 | 25.4 | n.d. | MWI, PMU, Utilities |
| | | Amman | 40.1 | 38.3 | 37.6 | 32.3 | 41.2 | 40.5 | 47.3 | 46.2 | 46.5 | 45.7 | n.d. | |
| | | Zarqa | 56.1 | 54.9 | 54.9 | 56.9 | 55.0 | 56.2 | 65.3 | 63.3 | 60.6 | 58.9 | 58.3 | |
| | Total losses; network (L / Subscriber / Day) | Aqaba | 445 | 484 | 434 | 421 | 488 | 461 | 504 | 475 | 407 | 386 | n.d | MWI, PMU, Utilities |
| | | Amman | 340 | 308 | 302 | 243 | 308 | 321 | 446 | 404 | 403 | 398 | n.d | |
| | | Zarqa | 583 | 563 | 555 | 571 | 502 | 550 | 788 | 732 | 663 | 636 | 621 | |
| | Indicator: Pipe breaks/bursts per km of mainlines | | | | | | | | | | | | | |
| | Main bursts/ 100km | Aqaba | 139 | 100 | 71 | 76 | 72 | 73 | 90 | 70 | 59 | n.d. | n.d. | MWI, PMU, Utilities |
| | | Amman | 109 | 89 | 62 | 57 | 45 | 61 | 73 | 64 | 68 | 63 | n.d. | |
| | | Zarqa | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. | 175 | 144 | 96 | 88 | 91 | |
| | Service leaks / 1000 connections | Aqaba | 122 | 121 | 82 | 192 | 153 | 119 | 116 | 124 | 98 | n.d. | n.d. | MWI, PMU, Utilities |
| | | Amman | 221 | 200 | 172 | 158 | 118 | 179 | 175 | 160 | 140 | 137 | n.d. | |
| | | Zarqa | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. | 158 | 239 | 155 | 161 | 184 | |
| Increased revenue to utility | Utility revenue (2015 JD/$m^3$ sold) | Aqaba | 1.08 | 0.91 | 0.90 | 0.93 | 1.02 | 1.04 | 1.03 | 1.08 | 1.07 | 1.04 | n.d | MWI, PMU, Utilities |
| | | Amman | 1.40 | 1.17 | 1.19 | 1.19 | 1.35 | 1.40 | 1.49 | 1.29 | 1.34 | 1.28 | n.d | |
| | | Zarqa | 0.90 | 0.70 | 0.61 | 0.62 | 0.71 | 0.60 | 0.53 | 0.81 | 0.97 | 0.88 | 0.87 | |
| Increased cost recovery by utility | Billing efficiency (%) | Aqaba | 91.5 | 91.5 | 91.5 | 91.5 | 91.5 | 91.5 | 91.5 | 91.3 | 92.4 | 91.8 | n.d | MWI, PMU, Utilities |
| | | Amman | 97.0 | 97.0 | 97.0 | 97.0 | 68.0 | 99.3 | 99.3 | 99.2 | 99.5 | 99.0 | n.d | |
| | | Zarqa | n.d | n.d | n.d | n.d | 90.0 | 90.0 | 90.0 | 80.8 | 90.4 | 98.5 | 98.9 | |
| | Collection efficiency (%) | Aqaba | 99.4 | 101.0 | 97.9 | 92.7 | 95.3 | 94.5 | 92.9 | 97.9 | 99.2 | 96.0 | n.d. | MWI, PMU, Utilities |
| | | Amman | 96.0 | 96.0 | 96.0 | 96.0 | 96.0 | 100.1 | 97.4 | 95.8 | 95.1 | 97.8 | n.d. | |
| | | Zarqa | n.d. | n.d | n.d | 64.6 | 108.4 | 72.9 | 85.1 | 92.1 | 103.1 | 96.0 | 91.9 | |
| | Operating Cost Recovery Ratio (OCRR) | Aqaba | 1.43 | 1.26 | 1.32 | 1.34 | 1.34 | 1.36 | 1.28 | 1.36 | 1.32 | 1.24 | n.d | MWI, PMU, Utilities |
| | | Amman | 1.07 | 1.09 | 1.10 | 1.07 | 1.03 | 1.10 | 1.00 | 1.09 | 1.07 | 1.15 | n.d | |
| | | Zarqa | 0.87 | 0.85 | 0.70 | 0.74 | 0.84 | 0.70 | 0.59 | 0.88 | 0.83 | 0.72 | 0.68 | |

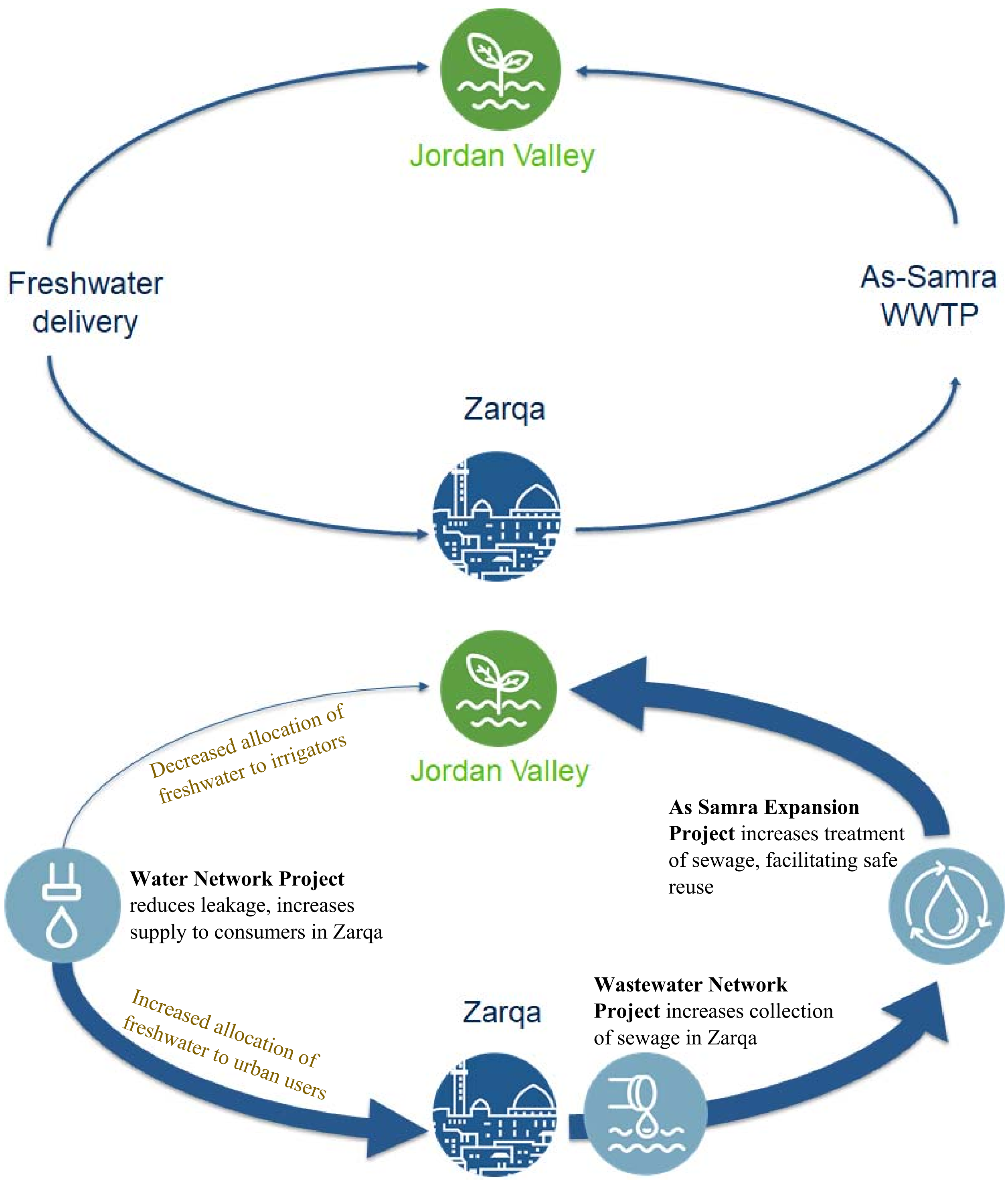


**Figure 1.** Qualitative depiction of (Top) pre- and (Bottom) ex pected post- Jordan Compact situations

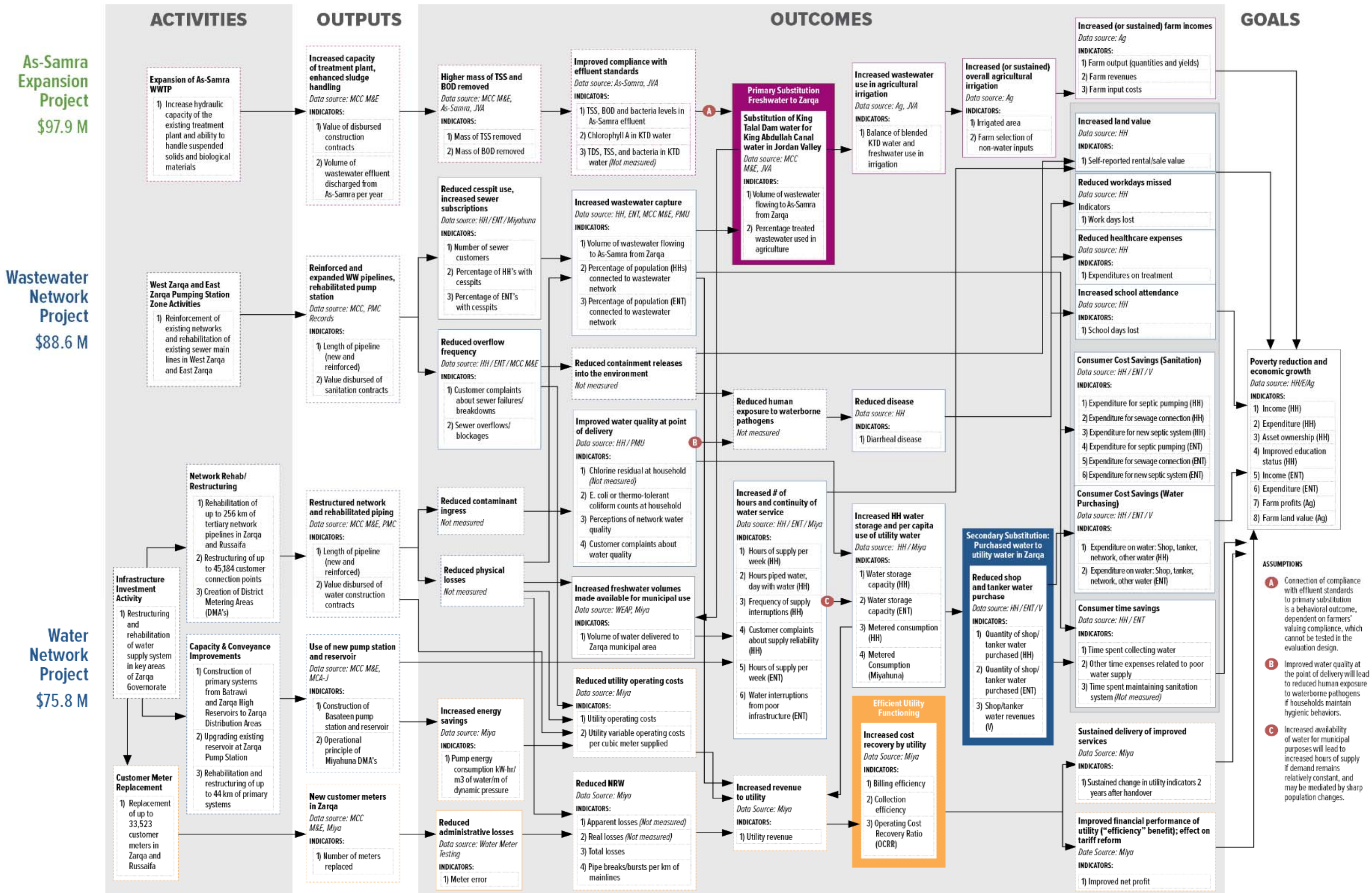


**Figure 2.** Full program theory of change, as elicited through participatory stakeholder consultations

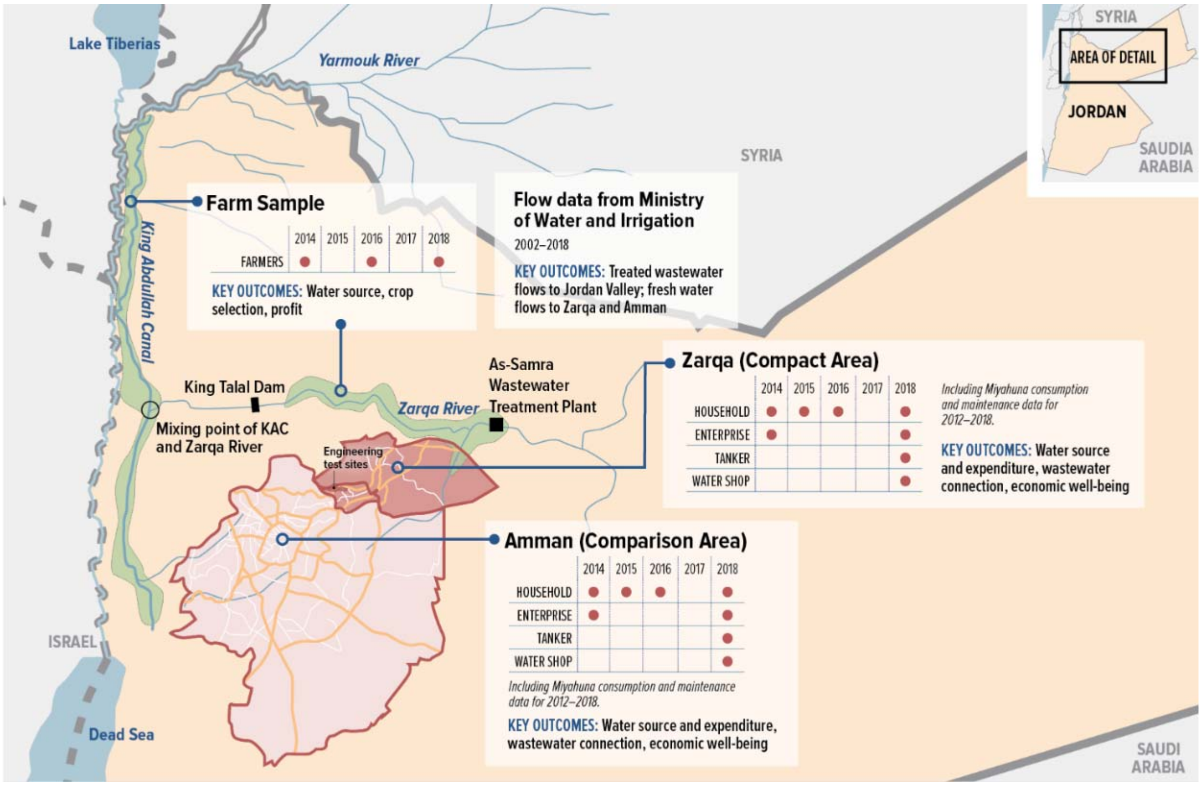


**Figure 3.** Locations and timing of data collection activities to support the evaluation

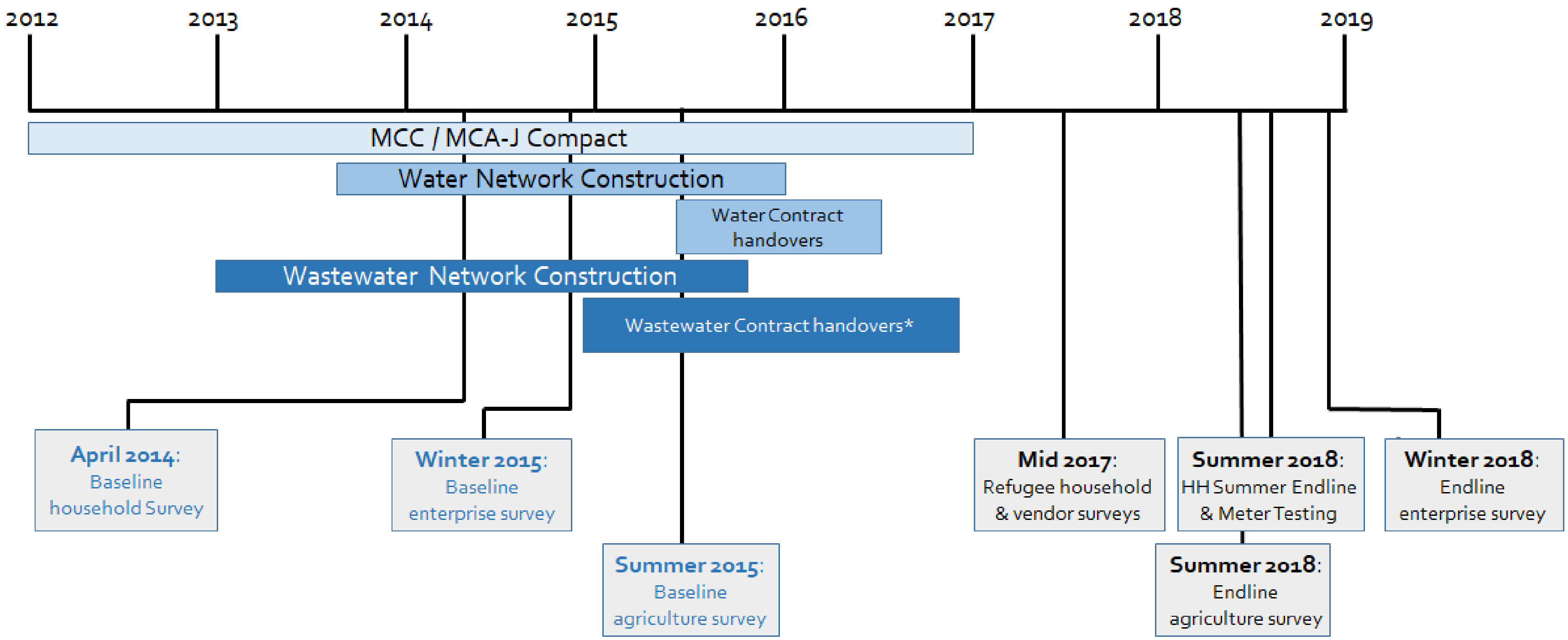


**Figure 4.** Timing of data collection relative to infrastructure intervention [Note that baseline surveys were conducted prior to any infrastructure operations, except for a few small wastewater Contract handovers preceding the baseline agriculture survey]