J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

Happiness on Tap: Piped Water Adoption in Urban Morocco

Why Randomize?

Photo: Aude Guerrucci | J-PAL/IPA

This case study is based on: Devoto, Florencia, Esther Duflo, Pascaline Dupas, William Parienté, and Vincent Pons. 2012. "Happiness on Tap: Piped Water Adoption in Urban Morocco." American Economic Journal: Economic Policy, 4(4), 68-99. doi:10.1257/pol.4.4.68

J-PAL thanks the authors for allowing us to use their paper and for sharing their data.

Learning Objectives

- Identify the strengths, limitations, and underlying assumptions of various quantitative methods commonly used to estimate the impact
- Provide a deeper understanding of potential sources of bias in impact evaluation

Subjects Covered

Causality, counterfactual, impact, comparison groups, selection bias, omitted variables, randomization.

Key Vocabulary	
Comparison group	A group that is as similar as possible to the treatment group in order to be able to learn about the counterfactual. In an experimental design, the comparison group (also called the control group) is a group from the same population as the treatment group that, by random assignment, is not intended to receive the intervention.
Counterfactual	What would have happened to the participants of an intervention had they not received the intervention. The counterfactual can never be observed; it can only be inferred from a comparison group.
Estimate	In statistics, a "best guess" about an unknown value in a population (such as the effect of a program on an outcome) according to a rule (known as the "estimator") and the values observed in a sample drawn from that population.
Impact	The impact of the intervention is the effect of the treatment. The impact is estimated by measuring the differences in outcomes between the treatment group and the comparison group.
Omitted variable bias	Statistical bias that occurs when relevant (and often unobservable) variables/characteristics are left out of the analysis. When these variables are correlated with both the primary outcome and a variable of interest (e.g., participation in an intervention), their omission can lead to incorrectly attributing the measured impact solely to the program. For example, omitting socioeconomic status, which is correlated with test scores, could lead to overestimating the impact of a tutoring intervention on a group of high-income students.
Treatment group	The group that receives the intervention.
Selection bias	Bias that occurs when the individuals who receive the program are systematically different from those who do not. For example, consider an elective, after-school tutoring program. Is it effective at raising children's exam scores? Comparing scores for those who participate and those who don't will produce a biased estimate of the effect of the tutoring program if these groups differ across characteristics that correlate with test scores. For example, those who choose to participate may be more motivated, and may have scored better than non-participants even without the tutoring program. Randomization minimizes selection bias because it breaks the link between characteristics of the individual and their treatment status. Selection bias can occur in other ways in a randomized evaluation. For example: - Participants can choose to take up a treatment or refuse it - Participants can choose to leave the study (i.e., attrit/attrition)

	Method	Description	What *assumptions* are required, and how demanding are the assumptions?	Required data
Randomized Evaluation	Randomized Evaluation/ Randomized Control Trial	Measure the differences in outcomes between randomly assigned program participants and non-participants after the program took effect.	*The outcome variable is only affected by program participation itself, not by assignment to participate in the program or by participation in the randomized evaluation.* Examples of such confounding effects could be information effects, spillovers, or experimenter effects. As with other methods, the sample size needs to be large enough so that the two groups are statistically comparable; the difference being that the sample size is chosen as part of the research design.	Outcome data for randomly assigned participants and non-participants (the treatment and comparison groups).
Basic Non-Experimental Comparison Methods	Pre-Post	Measure the differences in outcomes for program participants before the program and after the program took effect.	*There are no other factors (including outside events, a drive to change by the participants themselves, altered economic conditions, etc.) that changed the measured outcome for participants over time besides the program*. In stable, static environments and over short time horizons, the assumption might hold, but it is not possible to verify that. Generally, a difference-in-differences or regressions discontinuity design is preferred (see below).	Data on outcomes of interest for program participants before program start and after the program took effect.
	Simple Difference	Measure the differences in outcomes between program participants and another group who did not participate in the program after the program took effect.	*There are no differences in the outcomes of participants and non-participants except for program participation, and both groups were equally likely to enter the program before it started*. This is a demanding assumption. Non-participants may not fulfill the eligibility criteria, live in a different location, or simply see less value in the program (self-selection). Any such factors may be associated with differences in outcomes independent of program participation. Generally, a difference-in-differences or regression discontinuity design is preferred (see below).	Outcome data for program participants as well as another group of non-participants after the program took effect.
	Difference in Differences	Measure the differences in outcomes for program participants before and after the program *relative* to non-participants.	*Any other factors that may have affected the measured outcome over time are the same for participants and non-participants, so they would have had the same time trajectory absent the program*. Over short time horizons and with reasonably similar groups, this assumption may be plausible. A "placebo test" can also compare the time trends in the two groups before the program took place. However, as with "simple difference," many factors that are associated with program participation may also be associated with outcome changes over time. For example, a person who expects a large improvement in the near future may not join the program (self-selection).	Data on outcomes of interest for program participants as well as another group of non-participants before program start and after the program took effect.

	Method	Description	What assumptions are required, and how demanding are the assumptions?	Required data
More advanced statistical non-experimental methods	Multivariate Regression	The "simple difference" approach can be—and in practice almost always is—carried out using multivariate regression. Doing so allows accounting for other observable factors that might also affect the outcome, often called "control variables" or "covariates." The regression filters out the effects of these covariates and measures differences in outcomes between participants and non-participants while holding the effect of the covariates constant.	*Besides the effects of the control variables, there are no other differences between participants and non-participants that affect the measured outcome.* This means that any unobservable or unmeasured factors that do affect the outcome must be the same for participants and non-participants. In addition, the control variables cannot in any way themselves be affected by the program. While the addition of covariates can alleviate some concerns with taking simple differences, limited available data in practice and unobservable factors mean that the method has similar issues as simple difference (e.g., self-selection).	Outcome data for program participants as well as another group of non-participants, as well as "control variables" for both groups.
	Statistical Matching	Exact matching: participants are matched to non-participants who are identical based on "matching variables" to measure differences in outcomes. Propensity score matching uses the control variables to predict a person's likelihood to participate and uses this predicted likelihood as the matching variable.	Similar to multivariable regression: *there are no differences between participants and non-participants with the same matching variables that affect the measured outcome*. Unobservable differences are the main concern in exact matching. In propensity score matching, two individuals with the same score may be very different even along observable dimensions. Thus, the assumptions that need to hold in order to draw valid conclusions are quite demanding.	Outcome data for program participants as well as another group of non-participants, as well as "matching variables" for both groups.
	Regression Discontinuity Design (RDD)	In an RDD design, eligibility to participate is determined by a cutoff value in some order or ranking, such as income level. Participants on one side of the cutoff are compared to non-participants on the other side, and the eligibility criterion is included as a control variable (see above).	*Any difference between individuals below and above the cutoff (participants and non-participants) vanishes closer and closer to the cutoff point.* A carefully considered regression discontinuity design can be effective. The design uses the "random" element that is introduced when two individuals who are similar to each other according to their ordering end up on different sides of the cutoff point. The design accounts for the continual differences between them using control variables. The assumption that these individuals are similar to each other can be tested with observables in the data. However, the design limits the comparability of participants further away from the cutoff.	Outcome data for program participants and non-participants, as well as the "ordering variable."
	Instrumental Variables	The design uses an "instrumental variable" that is a predictor of program participation. The method then compares individuals according to their predicted participation, rather than actual participation.	*The instrumental variable has no direct effect on the outcome variable. Its only effect is through an individual's participation in the program*. A valid instrumental variable design requires an instrument that has no relationship with the outcome variable. The challenge is that most factors that affect participation in a program for otherwise similar individuals are also in some way directly related to the outcome variable. With more than one instrument, the assumption can be tested.	Outcome data for program participants and non- participants, as well as an "instrumental variable."

Introduction

How do we know if a program or policy had an impact? Typically, we would want to know if changes in the outcomes of participants can be directly attributed to an intervention rather than other factors. Ideally, evaluators would be able to track the outcomes of participants overtime as they participate in a program, measure any changes that occur, and then go back in time and measure the same group's progress without the program in place. This second, hypothetical set of outcomes represents what *would have happened* in the absence of treatment and is called the **counterfactual**.

Because we can never observe the counterfactual, the central challenge of any impact evaluation is to find a valid proxy for the counterfactual. We typically do this by selecting a group of people who resemble participants as much as possible but who did not participate in the intervention. This group is called the **comparison group**. It is important that the comparison group and the participant group are, on average, as similar as possible, so that we can attribute any differences in outcomes to the intervention. We can then estimate the **impact** by calculating the difference in outcomes observed at the end of the intervention between the comparison group and the **treatment group**.

A valid, unbiased impact estimate can only be attained if the comparison group is a good representation of the counterfactual. If the comparison group poorly represents the counterfactual, then the estimated impact will be **biased**, leading us to either over- or underestimate the true effect. The method used to select, construct, or estimate the comparison group is a key decision in the design of any impact evaluation.

This case study presents different methods for estimating the impact of a policy or program and uses the same data to show how different methods may produce different results. To motivate the concepts covered, we draw on an evaluation of an interest-free loan program offered to households in Morocco to cover the cost of private water connections.

Water Access

At the time of this evaluation, about 1.1 billion people worldwide had no access to a drinking source of water within 1 kilometer, and among those that had access, only 42% had a household connection (World Health Organization 2005).[1] Many low-income households around the world spend

[1] While access has since improved, 703 million people worldwide remain without a drinking source of water within a 30-minute walking distance of their homes in 2022 (World Health Organization and United Nations Children's Fund 2023).

a significant amount of time fetching water, which is often the prime responsibility of women and girls. The burden of water collection can be a major source of stress, tension and conflict. In the study sample in Morocco, 65% of households without a water connection report that water is a major source of concern, and 16% and 12% report that they have had water-related conflicts within the family or with neighbors, respectively. Given the benefits for individual and household health and well-being, ensuring access to safe drinking water is a major priority in many countries, and it is one of the UN's Sustainable Development Goals.

Piped Water Adoption in Morocco

While many studies focus on the impact of clean water access on physical health, piped water access can also improve non-health outcomes such as time available for other activities, life satisfaction, and stress. In order to estimate the impact of access to piped water on households' wellbeing, researchers in J-PAL's network collaborated with Amendis, the local affiliate of an international, private utility company responsible for the management and operation of various public services in Tangiers, Morocco.

Amendis launched the "Social Home Connections" program in 2007 to offer interest-free loans to households in Tangiers to buy a connection to the water and sanitation network.[2] Households were required to repay the loan over a period of three, five, or seven years at a monthly rate of 105 Moroccan dirhams (US$15). All households in the sample were eligible for the loan program.

This case study focuses on two main questions. First, did the program increase households' water quantity? Second, did the program increase households' satisfaction? With these questions in mind, the researchers collected data on:

- Households' self-reported water quantity as measured by whether there is enough water for bathing, cleaning, cooking and drinking
- Households' self-reported level of satisfaction with their life[3]
- Background variables such as age, gender, and socio-economic status

Did the piped water access program work? The following (fictitious) new releases and blog excerpts illustrate different methods of evaluating impact to answer this question. (Refer to the previous table for a list of different evaluation methods.).

[2] The price of the connection depended on the work required to install a pipe from the network to homes.

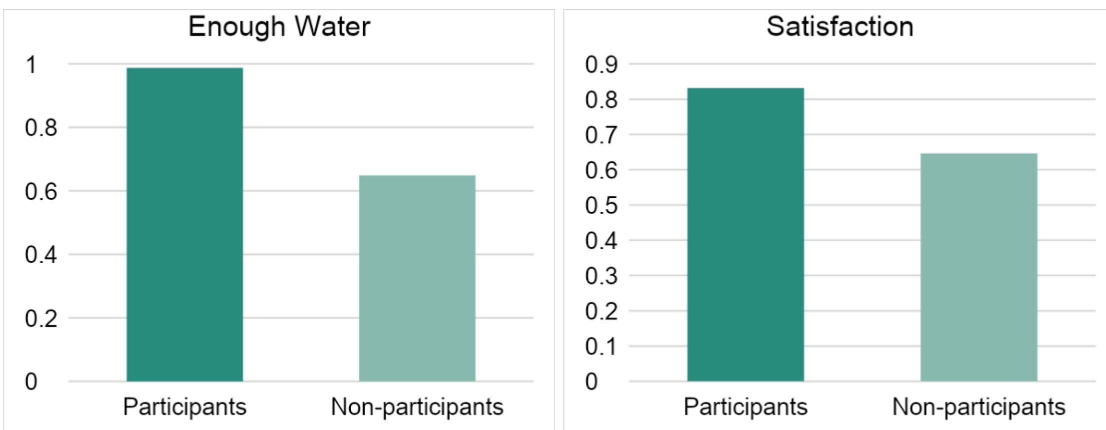
[3] The researchers measure this through the share of households who rate their level of life satisfaction as 5 or above when asked to rate this on a scale of 1 to 10.

Estimating the Impact of Piped Water Access

Method 1

Blog excerpt: Loan program leads to improved well-being and water access

Researchers and program implementers in Morocco partnered to examine the impact of offering interest-free loans to cover the cost of piped water connections to homes. Their data has just been published, and the program had a huge impact! Households who took up the loan and installed piped water were 34 percentage points more likely to have enough water for drinking, cooking, cleaning, and bathing, and 19 percentage points more likely to be satisfied with their lives compared to households that did not install piped water. The local government in Tangiers is making plans to roll-out the program across the city next year



Discussion

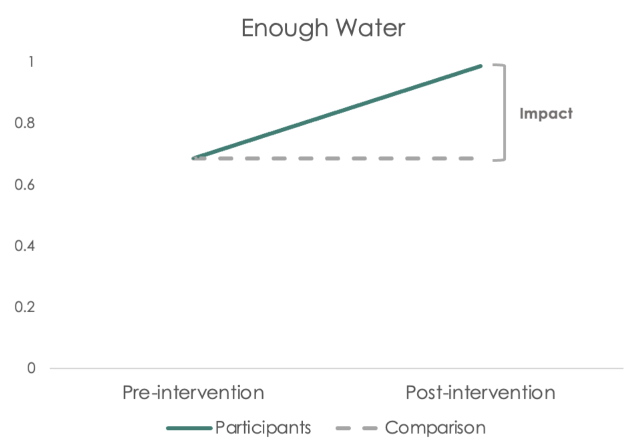
1. What type of evaluation method does this blog excerpt imply? What is the comparison group used to mimic the counterfactual in this study?

2. What assumptions do we have to make to believe this estimate? What might threaten these assumptions?

METHOD 2

Blog excerpt: Life is good thanks to the piped water adoption program

A new report confirms previous results that piped water access significantly increases households' water quantity and life satisfaction in Morocco. The report uses longitudinal data to compare the water quantity and life satisfaction of households that took an interest-free loan to purchase a piped water connection before and after taking part in the program. The report finds that these households were 30 percentage points more likely to have enough water for household consumption and 18 percentage points more likely to be satisfied with their life after the program as compared to before.



DISCUSSION

1. What type of evaluation method does this blog excerpt imply? What is the comparison group used to mimic the counterfactual in this study?

2. What assumptions do we have to make to believe this estimate? What might threaten these assumptions?

Method 3

News release: Increases in life satisfaction from piped water adoption program may be more modest than previously thought

A recent program in Tangiers, Morocco implemented a door-to-door awareness campaign to encourage households to sign up for interest-free loans to cover the cost of piped water connections to their homes. A few blogs reported positive impacts of piped water adoption on households' water quantity and life satisfaction, but a new study finds an even greater effect on water quantity.

The new study compares changes in outcomes over time among households that took up the loan to purchase a piped water connection to the changes in outcomes over time among households that did not adopt piped water (comparing the difference in changes across the two groups). The study shows that adoption of piped water increased the likelihood of having enough water by 39 percentage points and the likelihood of life satisfaction by 10 percentage points, as illustrated in the figures below.

The design simultaneously accounts for both selection bias (by comparing participants to themselves over time) and time trends (by accounting for over time changes among non-participants). Therefore, these new results are more reliable than the results promoted by previous blogs.

Discussion

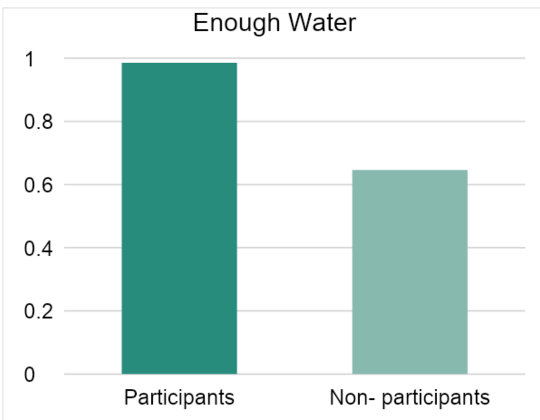
1. What type of evaluation method does this news release imply? What is the comparison group in this study used to mimic the counterfactual?

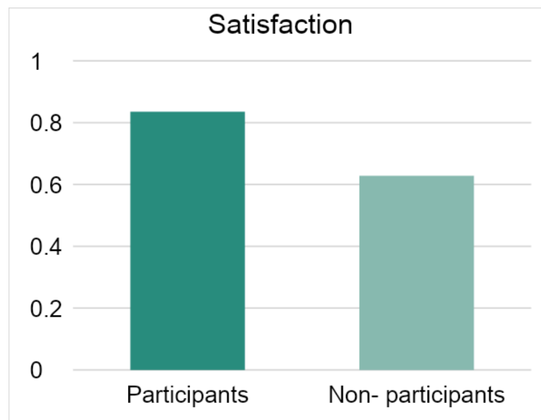
2. What assumptions do we have to make to believe this estimate? What might threaten these assumptions?

Method 4

Report: Piped water adoption improves life satisfaction in Morocco

A new study of a piped water adoption program in Morocco finds a greater effect of piped water on life satisfaction compared to previous results. The study compared the outcomes of households that participated in the program to a sample of matched households that did not participate but were otherwise similar in terms of their age, gender, income, and pre-intervention levels of water quantity and life satisfaction. As shown in the figures below, the study found that the program increased participating households' likelihood of having enough water by 34 percentage points and their likelihood of being satisfied with their life by 21 percentage points, as compared to non-participating households.





The table below takes a closer look at differences across households, comparing those who elected to participate in the program to those who did not. It shows that participants and non-participants differed significantly on the basis of gender, age, and income. Perhaps most importantly, they varied on the basis of their pre-intervention levels of the study's key outcome variable–satisfaction.

Pre-intervention characteristics by participation status, before matching

	Participants	Non-participants	Difference	N
Age (head of household)	48.68	51.20	-2.52 **	799
Gender (head of household)	0.82	0.73	0.09 ***	833
Average monthly income over last year	4.69	4.09	0.60 ***	753
Satisfaction (pre-intervention)	0.64	0.56	0.08 **	808
Enough water (pre-intervention)	0.69	0.73	-0.04	829

Notes: ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. Sample sizes vary across outcomes due to missing data.

After matching non-participants to participants based on similar observable characteristics, the differences between participants and non-participants decreased in magnitude and all the differences became statistically insignificant (as shown in the table below). This suggests that non-participants are more comparable to participants in the matched sample and thus more likely to represent a valid estimate of the counterfactual. (However, this comes at the cost of a reduced sample size, going from 845 households in the full sample to 669 in the matched sample as households with no comparable match were dropped.)

While the matched sample is well-balanced on these observable variables measured in the pre-intervention survey and included in the matching algorithm, there is no guarantee that they will be balanced on unobservable characteristics that were not included. For instance, participants and non-participants may be very different in terms of their water use habits and preferences, which may be difficult-to-measure yet potentially important determinants of whether a household has enough water.

Pre-intervention characteristics by participation status, after matching

	Participants	Non-participants	Difference	N
Age (head of household)	48.05	47.93	0.12	669
Gender (head of household)	0.83	0.86	-0.03	669
Average monthly income over last year	4.73	4.57	0.16	669
Satisfaction (pre-intervention)	0.63	0.61	0.02	669
Enough water (pre-intervention)	0.70	0.69	0.01	669

Discussion

1. What type of evaluation method does this report imply? What is the comparison group in this study used to mimic the counterfactual?

2. What assumptions do we have to make to believe this estimate? What might threaten these assumptions?

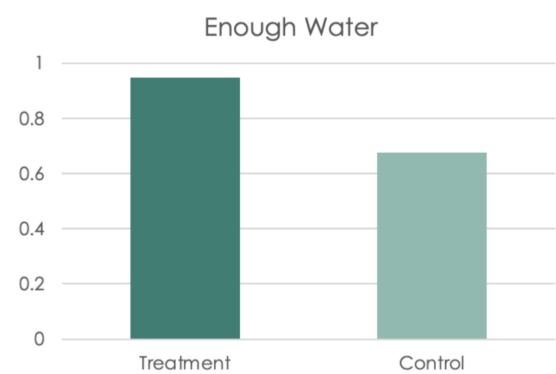
METHOD 5

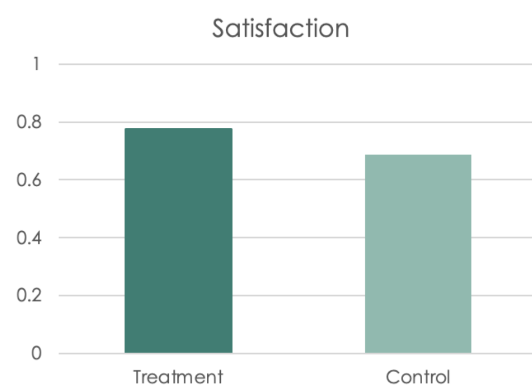
Report: Estimating the impact of an interest-free loan program on households' water access and well-being

Recognizing the disagreement among previous reports on the true impact of the piped water adoption program in Tangiers, Morocco, researchers in J-PAL's network conducted an evaluation to test the impact of the loan program on households' water access and well-being. After enrolling 845 households in the study, all of whom were eligible for the loan program, the researchers randomly assigned 434 households to be part of a door-to door awareness campaign encouraging them to take up the loan program and offering assistance with the application, and 411 to a comparison group.

Of households in the treatment group, 69 percent participated by obtaining the loan and installing water pipes. Of households in the control group, 10 percent also obtained the loan and installed piped water. Even though not everyone in the treatment group participated and some households in the comparison group participated, the difference in program take-up between the two groups allowed the researchers to estimate the effect of the piped water connection for those who complied with their treatment assignment.[4]

The figures below depict these estimates, showing a positive impact from the program on households' likelihood of having enough water (27 percentage points) and a smaller impact on likelihood of life satisfaction (9 percentage points).





[4] In this case study, we focus on the "Local Average Treatment Effect" (LATE) estimate, which is the average treatment effect for compliers, i.e., those who are induced by their treatment assignment to take up the treatment. This means that the estimates presented for the randomized evaluation method reflect the impact of taking a loan to purchase a piped water connection on those who were reached by the door-to-door awareness campaign. The LATE should not be confused with the "Intent to Treat" (ITT) estimate, which compares those assigned to the treatment group to those assigned to the control group, regardless of their actual take up of the program. This will be covered in more depth in the Threats and Analysis lecture.

Discussion

1. What type of evaluation method does this report imply? What is the comparison group in this study used to mimic the counterfactual?

2. What assumptions do we have to make to believe this estimate? What might threaten these assumptions?

Comparing All Five Methods

The table below presents impact estimates of the piped water adoption program using the five different methods discussed in this case study.

Comparison of impact estimates of the piped water adoption program under different methods

Method	Satisfaction	Enough Water
Simple difference	0.19 ***	0.34 ***
Pre-Post	0.18 ***	0.30 ***
Difference in differences	0.10 **	0.39 ***
Matching	0.21 ***	0.34 ***
Randomized evaluation[5]	0.09 *	0.27 ***

Notes: Impact estimates are shown in percentage points.
*, **, and *** indicate significance at the 10, 5 and 1 percent levels.

As you can see, not all methods yield the same result. Hence, the choice of method is crucial. There are many ways to estimate a program's impact and reasons why we might choose one method over another for a given evaluation. Any method relies on the validity of its underlying assumptions and the possible biases or challenges that these assumptions introduce.

[5] In this case study, the impact estimates for the randomized evaluation method are measured by the LATE estimator. The ITT estimate, which measures the impact of *offering* the program, is equal to a 5 percentage point increased likelihood of life satisfaction and a 16 percentage point increased likelihood of having enough water.

Whatever method we use, it is important to think critically about how the method constructs a counterfactual and the assumptions underlying this.

Although all methods show a positive impact, we know that many may be influenced by selection bias in who decides to take up the loan versus not which in this case may overestimate the true impact of the program.

References

J-PAL. "Household Water Connections in Tangier, Morocco." Abdul Latif Jameel Poverty Action Lab. Cambridge, MA.

World Health Organization, United Nations Children's Fund Joint Monitoring Programme for Water Supply and Sanitation. 2005. Water for Life: Making it Happen. Geneva, CH: WHO Press.

World Health Organization and United Nations Children's Fund. 2023. Progress on household drinking water, sanitation and hygiene 2000–2022: special focus on gender. New York, USA.

Reuse and Citations

To request permission to reuse this case study or access the accompanying teachers' guide, please email training@povertyactionlab.org. Please do not reuse without permission. To reference this case study, please cite as:

J-PAL. "Case Study: Piped Water Adoption: Why Randomize." Abdul Latif Jameel Poverty Action Lab. 2023. Cambridge, MA.

To reference the original study, please cite as:

Devoto, Florencia, Esther Duflo, Pascaline Dupas, William Parienté, and Vincent Pons. 2012. "Happiness on Tap: Piped Water Adoption in Urban Morocco." American Economic Journal: Economic Policy, 4(4), 68-99. doi:10.1257/pol.4.4.68