

Evaluation Replication Worksheet

Mailing Inserts to Increase Paperless Billing Uptake - Scottsdale, AZ

Use this worksheet to develop your city's Trial Protocol – the document that will guide your evaluation.

Since this is an evaluation replication guide, we've gotten you started by filling in some of the elements of your approach that stay the same. The rest is up to you!

Before we dive in - a few reminders:

- A big part of designing and running an evaluation is **careful planning** - we've designed this guide following our "Trial Protocol" template. A protocol can help you to document the decisions and intentions around your intervention. It's a forward-looking plan that will help you down the road - and also make it easy for people to replicate your work in the future (like we are doing now).
- We provide some guidance within the document for each section but we have designed this guide in mind for someone who's comfortable handling data and knows how to understand a power calculation. If that's not you yet - that's ok! We'd recommend you partner with someone in your city to complete this document. Or, you can try [courses](#) on the What Works Cities Academy to get up to speed or join a BIT Sprint for "Evaluation Foundations" to learn some of the key concepts.

To use this document: Please download a copy and fill in the "Your Approach" column

Part 1 - Scoping Your Evaluation

Step & Guidance	EXAMPLE – Scottsdale, AZ	Your Approach [Fill in this column]
Define your problem statement What is your high-level goal?	<i>Sending physical utility bills is costly. Getting more residents to sign up for paperless billing would lead to considerable savings for the City of Scottsdale.</i>	Sending physical utility bills is costly. If more residents sign up for paperless billing the City will save a considerable amount of money.
Describe your solution How will you try to reach your goal?	<i>We included a behaviorally designed insert in the envelope containing the monthly utility bill; the insert aimed to encourage Scottsdalians to sign up for paperless billing.</i>	We will add an insert to the utility bills encouraging residents to sign up for paperless billing.
Describe your comparison What will you be comparing your approach to? What does your comparison or control group receive? Is the comparison business-as-usual, or an alternate version of a new process? If not business-as-usual, why?	<i>We randomly assigned Scottsdalians who had their monthly utility bill due to receive either a business as usual envelope containing their utility bills (control group) or an envelope that contains the utility bills as well as an insert (treatment).</i>	We will randomly assign residents who have their monthly utility bill due to receive either a business as usual envelope containing their utility bills (control group) or an envelope that contains the utility bills as well as an insert (treatment).
Define your outcome & indicator How will you measure the success of your approach? What is the target outcome and what is the indicator? - What is the unit of your outcome measure? - How does it relate to your unit of assignment? - How will you link your outcome measure to treatment assignment?	Outcome: More residents will sign up for paperless billing, saving a lot of money for the city. Indicator: The indicator is the proportion of residents that signed up for paperless billing within 20 days of receiving their utility bill. (removing duplicates for the same account number and controlling for the type of letter people received). This is a binary outcome measure recorded at the individual level. It will be extrapolated by comparing the number of accounts that receive paperless bills before and after the intervention, as reported on the dataset provided by Scottsdale Business Services.	Target outcome: Indicator:
Describe your research question	<i>Does attaching a behaviorally designed insert to the paper utility bills increase the number of residents signing up for paperless billing?</i>	Research Question:
Identify your population Describe who you are trying to target with this intervention	Target Population: All Scottsdale residents who are still receiving paper utility bills (regardless of whether they then pay these bills online or not).	Target Population:
Identify your sample Describe who you will test your intervention on (and if / how that is different than the population) - Who will you actually be able to reach with your communication? - How will you identify and reach your sample?	Trial Sample: All Scottsdale residents that receive 'regular' or 'no-return envelope' physical utility bills (as classified in the Scottsdale utility billing database system) within the trial period and who have their monthly utility bill due during the duration of the trial.	Trial Sample:

- Does this sample represent your population of interest?
- Do you anticipate any selection bias in who is represented in your sample?

DATA QUALITY CHECKPOINT

Before going too much further - it's important to check the quality of your data

Before you design your evaluation, you need to understand **what data is available** and **any potential issues** you may have with the data.

Reviewing your data in advance helps you determine how you design your evaluation. For example, you might find out that you have less available data than we thought, which could make your planned evaluation more difficult and may compromise your ability to measure the outcomes you seek to evaluate. Or you may discover your outcome measures are structured differently in your data than you had thought because it may require you to make different decisions about your evaluation design or explore other opportunities to generate data to help you with your evaluation.

Before you move forward, check your dataset to determine the following:

- ✓ You have access to the data set
- ✓ Your data set contains the information on our outcome variable you want to measure
- ✓ You are aware of the ways the data could be inaccurate or unreliable (e.g. self-reported, incomplete, etc.)
- ✓ If new data is being collected or if data is being collected manually, consider could it go wrong?
- ✓ You've reviewed an output of this data to determine the accuracy and how difficult it is to produce (if possible)
- ✓ You are handling any personal identifying information and consent processes in an accordance with legal guidelines and ethical best practices

Part 2 - Designing your evaluation

Step / Description / Resources	EXAMPLE – Scottsdale, AZ	Your Approach
Decide randomization strategy		
Unit: <i>What is being randomized (e.g. individual, household, school, etc.)?</i>	Unit: Batches of 500 utility bills	Unit:
Method: <i>How will you conduct the randomization? (E.g., through a random number generator, lottery, coin toss, randomized paper sequence, etc.)</i>	Procedure: We randomized assignment by manually attaching the insert to the bills in either the odd or even-numbered batches each week, in an alternating fashion.	Procedure:
Verification: <i>How will you make sure that every participant is assigned to one, and only one, treatment group? If there is a risk that participants receive both treatments, will you be able to track this and control for it in analysis?</i>	Verification: Every day, for the duration of the trial, the Scottsdale printers will share with BIT the list of batches being sent on that day, indicating which are assigned to treatment and control based on the agreed-on randomization rule.	Verification:

Blinding & Masking: <i>Will participants know their own treatment assignment? Will their treatment status be known to others involved in the trial or intervention? (For example, will frontline staff know the difference between people in the control group versus treatment group?)</i>	Blinding and masking: We made sure that participants would not know their treatment assignment. Treatment group members would assume that everyone in their situation receives an insert along with their utility bills. Control group members would not be aware of the intervention or trial.	Blinding and masking:
Spillovers: <i>What are the ways in which someone's treatment status might affect the outcome of someone in a different treatment group? Is it possible that a previous intervention might influence their behavior in this trial? If so, can these be minimized by creating distance between participants?</i>	Spillovers: It is possible that a treatment group member could have told a control group member about the insert or may have shared the additional information around the signup process. Therefore, we may have underestimated the positive impact of our intervention because we did not capture control group participants who signed up for paperless billing due to the insert being sent to someone in their network.	Spillovers:
Calculate your sample size and power requirements		
Baseline: <i>What is the current average for your indicator? Or, if you don't have historical data, what data do you have available that might give you an indication of what your current average might be? This should be your best guess for the expected outcome for your control group in your trial. If you don't have a precise number, run the power calculations for a range of different baselines based on your assumptions.</i> Standard deviation: <i>If your outcome indicator is a continuous measure, what is the baseline standard deviation?</i>	Baseline: Before the trial began Scottsdale Business Services shared a data set comprising all accounts due to receive a paper bill in the coming month. This served as our baseline data for the proportion of people who already adopted paperless billing.	Baseline:
Power calculation: Use the power calculator to complete your power calculations using the numbers you listed above. • Per group sample size: • Minimum detectable effect size: • Significance level: 0.05 • Power level: 0.80	Number of participants available: The trial included 56,328 bills for people who had their utility bills due during the month of the intervention. It excluded new customer groups and people who were based outside the country. After adjusting for attrition and eliminating duplicate account numbers we had a total sample size of 55,106 bills. Power calculation: Based on a conservative estimate informed by data from the prior months, we expected 41,000 bills to be included in the trial. We clustered them at the batch level (one batch contains 500 bills or less) considering that there are 18 workdays throughout the duration of the trial and that the City prints out 6 batches per day on average. In total we calculated 108 clusters. We assume a low ICCR of 0.05, a significance level of 0.05 and a power level of 0.80.	Number of participants available: Power calculation: Per group sample size: Minimum detectable effect size:

	The average N per cluster is $41,000/108 \approx 379$. Our Design Effect is therefore $1+(379-1)*0.05 = 19.93$. Where our nominal N is 41,000, our true N is therefore $41,000/19.93 = 2,057$. Assuming further, a baseline proportion of 0.10, based on past utility bill data, our MDES is 2.10 percentage points.	
Target effect size: How large of an effect size do you think is reasonable to expect from this intervention (based on prior evidence if available)? Alternatively, what would be a meaningful effect size (based on break-even point if applicable), and why?	Target effect size: This trial design did not specify a target effect size. Note: If referencing academic literature or other relevant work to estimate your effect size, you can follow this template: "Some of the most relevant examples are (...). Based on those, we expected an average treatment effect of (...) for a [duration] trial."	
Attrition: What is your best guess of the number/percentage of participants that will leave your sample between randomization and outcome data collection. Adjust your sample to ensure it's large enough for analysis at the end of the trial.	Attrition: We did not expect a lot of attrition, the only thing that has decreased our sample size has been the occurrence of duplicate account numbers.	

FEASIBILITY CHECKPOINT:

Is the minimum detectable effect size reasonable? If not, can you change the sample size by running the trial for longer or changing the intervention to make it potentially more impactful? If no changes can be made, do you still want to go ahead with the trial as an implementation pilot?

Consider experimental threats & risks - What things can you do to make sure people receive the intervention as it was intended? - Is there a way you can double-check that your randomization has been implemented correctly and the participants are receiving the treatment they were assigned? - Are there things you can do to make sure that staff is prepared and ready to implement the trial (e.g., training on data collection, a pilot period to adjust to new workflows, extra time to adjust to a new call volume)?	Risk: Human error due to the complicated nature of the randomization rule could lead to the randomization not being implemented correctly or to participants receiving the wrong treatment. Likelihood: Low Impact: High Mitigation Approach: The team consulted with BIT to clarify the randomization procedures, and document them in writing. An email was circulated with the randomization process details to relevant stakeholders. Risk: Letters are not opened, which will hinder the effectiveness of the trial. Likelihood: Medium Impact: Low/none Mitigation Approach: None. There is no reason to expect that this issue will	Risk: Likelihood: Impact: Mitigation Approach:
---	--	---

	differ across the treatment and control groups, so any effect will be spread and will not skew the results Risk: Significant uptake to paperless billing may put strain on the e-billing site • Likelihood: Low - the e-billing site is equipped to handle significant traffic already considering the number of residents who already use the system. Also, since the intervention is going out through a mailer, we do not anticipate significant numbers of simultaneous attempts to access the system that might occur with digital outreach • Impact: High - if the system experiences downtime, residents may revert to paper bills as more reliable • Mitigation Approach: Check with technology vendor or IT services to ensure that the e-billing system can handle additional traffic, and that tech support is on call to address any issues during the trial.	
Consider ethical risks • <i>How might participating in the trial harm people in your sample or others?</i> • <i>How will you monitor the trial to ensure you can detect the harm early and change your implementation if needed?</i>	Risk: We did not identify ethical risks to participants or others in the trial design. For replication of this trial: • Consider whether your electronic billing system has associated fees that might disadvantage trial participants • Consider the reliability and ease-of-use of the e-billing system. Downtime, site maintenance, or a difficult to navigate user interface may disadvantage trial participants that sign up for paperless billing. • Consider whether there are penalties for late bill payments that might exacerbate the difficulties listed above for trial participants.	Your answer:
Consider race & equity implications • <i>If not already included above, how will you check to see if your intervention had differential effects by race? Disaggregating your data</i>	Note: We did not identify potential equity concerns with the intervention. However, we did not conduct subgroup analysis to check for any differential impact within demographic subgroups.	Your answer:

by race is often a good first step here, but consider descriptive analyses that might help explain any trends you are seeing. In cases where you do not have race demographics, can you use proxy variables (e.g. Census tract information matched to zip codes?)

For replication of this trial:

The intervention could have the following equity concerns, which could cause it to benefit some populations more than others:

- **Internet access:** if some populations do not have reliable internet, paperless billing could not be a viable option or could result in inability to access and pay bills on time
- **Device access:** If the e-billing site is not optimized for mobile access, it could result in inability to access or pay bills online for groups that have limited desktop computer access.

Plan for Data Analysis

Understand and Specify Your Variable for Analysis

There are two parts to planning for your analysis. First, you need to think through your different variables and make sure to document how you plan to use them for your trial analysis. This step helps you to ensure you have the data you need and sets you up for success for your analysis plan. Additionally, it's good to specify what you will do with bad data or data that doesn't match.

Here are some questions to guide your data checks:

- Where are you getting the data you need to complete the trial analysis (e.g., treatment assignment, outcome indicator data, other participant characteristics)?
- Are there any data security procedures that need to be followed?
- How will you assemble the data? How will your variables be constructed (e.g., units, interpretation of values, etc.)?
- How will you check your data for accuracy (e.g., to make sure that any data merges were done correctly, or that missing values have been identified and dealt with as needed)?

Specify Your Analysis Plan

Next, you need to specify your analysis plan. You should have a "hypothesis" you are testing - that your intervention will work -- but specifying ahead of time will allow you to say what statistical test & analysis you will use to determine if it worked. See below for an example of Scottsdale's analysis plan. If you have questions on how to choose what test to run, check out [resources here](#).

Questions to guide your analysis plan:

- *For your Primary outcome:*
 - **What statistical test will you run for your outcome indicator?** For example, will you run a t-test to see if the mean outcome for your treatment group is statistically different than the mean outcome for your control group, a regression to control for other factors, or a more complex analysis?
- *For your Secondary analysis:*
 - **Are there any other statistical tests you would like to run** (e.g., other outcome indicators, or looking at sub-groups)?
 - **Will you want to do any cost-benefit analysis?**

Sample Data

Variable name	Type	Source	Measurement
AccountNumber	Unique ID	Scottsdale Business Services	Many
Treatment	Treatment assignment	From randomization	0 - control 1 - treatment
Duplicates	Remove from analysis	Scottsdale Business Services	0 - unique value 1 - duplicate
Letter Type	Exploratory	Scottsdale Business Services	0 - Normal Envelope

	outcome		1 - No Return Envelope
SignedUp	Primary outcome	Scottsdale Business Services	0 - did not sign up 1 - did sign up

Sample Analysis Plan

We analyzed the results using an ordinary least squares regression (OLS), with standard errors clustered at the batch level and robust to heteroskedasticity. It will take the following form:

$$Y_{ib} = \alpha_{ib} + \beta_1 X_{ib} + \beta_2 X_{ib} + \epsilon_{ib}$$

Where Y_{ib} is our main outcome, a binary variable equal to 1 if individual i in batch b switches to paperless billing, and 0 otherwise;

α_{ib} is our constant term, that equals the value of the outcome where all following regressors are set to 0;

θ_{ib} is the treatment indicator, a binary variable equal to 1 if individual i in batch b is assigned to receive an insert, and 0 otherwise. β_1 can therefore be interpreted as the average treatment effect of the intervention;

X_{ib} is a letter type fixed effect, taking the possible values of 'regular' or 'no-return envelope' utility bills;

ϵ_{ib} is an error term robust to heteroskedasticity (using Eicker-Huber-White standard errors), clustered at the batch level.

Your Approach:

Please describe how you will analyze your data as well as what steps you will take to ensure the data is clean and ready for analysis. Note: You can use a different analysis than the model specified in the example above, based on your available data.

Part 3 - Implementing your evaluation

Plan for Implementation

Create a project implementation plan and timeline

At this point, an evaluation starts to feel just like any other project -- with a few extra checks! You will need to carefully plan for implementation and ensure randomization is able to occur and data is able to be collected at the correct intervals. Be sure to build in time for getting any necessary approvals and for double-checking your work. We've included a sample timeline here - but we find that implementation works best when it follows a city's normal project management process.

Sample Implementation Plan

Task	Owner	Deadline
Agree of final version of the insert	Scottsdale (UBD) and BIT	Thursday 23 March

Share the final version of the insert to be printed in bulk ahead of the trial	Scottsdale (UBD) and BIT	Monday 27 March
Confirm via email the final randomization rule	Scottsdale (UBD) and BIT	Monday 27 March
Share the baseline dataset	Scottsdale Business Services	Wednesday 5 April
Launch of the Trial	Scottsdale (UBD)	Wednesday 5 April
Confirm each week the alternation of the randomization	Scottsdale (UBD) and BIT	• Wed 5 April • Mon 10 April • Mon 17 April • Mon 24 April
Share the final dataset	Scottsdale Business Services	Wednesday 3 May
Data analysis completed	Scottsdale (UBD)	Wednesday 10 May
Final Report drafted	Scottsdale and BIT	Monday 15 May

Your Approach:

Please fill in a chart of your implementation plan for this project - feel free to copy in from above!

Monitor your evaluation

While your evaluation is in the field, be sure to check in and ensure everything is happening to plan. This can take the form of automated data checks, regularly scheduled check-ins with your partners or observations of the process. It's better to catch anything before the end of your evaluation so you can adapt as needed!

Part 4 - Analyzing your results

Analyzing your data

You've already specified your analysis plan before you launched the trial - so now is the time to use it!

If possible, build in some time and budget to have someone who is not involved with the project to review your work. It can help you eliminate any blind spots or highlight any assumptions you've made about the data.

Once you've analyzed your results you can ask yourself a few questions to help you reflect your results:

- Descriptive analysis:
 - What is the average and standard deviation for your full sample?
 - What is it for each of your treatment groups?
 - Are there any characteristics of the sample to describe (demographics, location, time, etc.)?
- Primary outcome:
 - Statistically significant: Yes/No
 - Effect size: What is the effect of your treatment, on average? (e.g., an increase of 2 percentage points or a decrease of \$100, on average)
 - Distribution: What is the range of outcomes – was the confidence interval quite large, or narrow?

- Interpretation: What does this difference mean practically? Is it large enough to make a meaningful difference? If there was a wide range of outcomes, is it acceptable to implement something with that range of results?

Congrats on running your evaluation!

Now that you have your results, you can decide how you want to use them. If your result was positive, you may consider scaling your solution. To see how Scottsdale scaled up their paperless billing insert campaign, head back to the Replication Guide.