Text Messaging for 211 Services - Tulsa, OK

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 this guide has been designed for those who are comfortable handling data and 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 <u>courses</u> 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 – Tulsa, OK	Your Approach [Fill in this column]
Define your problem statement	In times of crisis, not enough residents at risk of eviction call the 211 helpline which connects callers to support services.	Not enough residents call the helpline which connects callers to support services
What is your high-level goal?		
Describe your solution	We will send an SMS to residents	We will send an SMS to residents
How will you try to reach your goal?		
Describe your comparison	We will randomly assign Tulsans at risk of eviction to either get the treatment SMS	We will randomly assign residents at risk of eviction to either get the treatment SMS
What will you be comparing your approach to?	reminder or not (business-as-usual control group).	reminder or not (business-as-usual control group).
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?		
Define your outcome & indicator	Primary outcome: More residents will call the beloline and get connected to the	Target outcome:
How will you measure the	services.	Indicator:
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?	Primary Indicator: Whether a recipient has called 211 and has been logged into the system by a 211 responder at least once during the trial period. These are binary outcome measures recorded at the individual level based on call records from 211 ReferNET and UniteUs. 211 helpline workers keep track of the contact of their callers, so if someone in the treatment or control group has called the service during the trial period their name and number will have a match on the UniteUS records, otherwise they will show up as new accounts. Exploratory Outcome: Decreasing housing instability so that more people at risk of eviction will access services that can help them avoid eviction. Secondary Indicators: (1) Whether the caller has received assistance from a 211 service provider. (2) Whether the recipient received specific agency or service information for a need categorized as "housing"-related and (3) TWhether the recipient was involved in an Eviction Event (i.e. checking if they were named in an eviction filing in Tulsa County court as reported in OKPolicy Institute eviction data).	
Describe your research question	Does sending an SMS reminder increase the number of at risk residents calling 211?	Research Question:

Identify your population Describe who you are trying to target with this intervention	Target Population: All residents of Tulsa who are having financial troubles and might be facing eviction soon (as determined by a predictive model based on water bill payment history).	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? Does this sample represent your population of interest? Do you anticipate any selection bias in who is represented in your sample? 	Trial Sample: Residents of Tulsa who display predictive characteristics of eviction (according to the water-billing predictive model) in the month prior to the SMS message being sent and who live in single-family occupancy households. This was not an exact match for our target population, but we do expect that it included a large portion of them.	Trial 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 you have and any potential issues you may have with the data.

Reviewing your data in advance helps you determine how you design your evaluation. Oftentimes, when reviewing the data you might learn key information that could change our evaluation design. For example, you might find out that you have less available data than we thought, which could make your planned evaluation more difficult and comprise your ability to measure the outcomes you seek to evaluate. Or you may discover that our outcome measures are structured differently in your data than you originally thought.

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

- You have access to the data set 1
- 1 Your data set contains the information or outcome variable that 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, could it go wrong?
- ✓ ✓ You've reviewed an output of this data (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 – Tulsa, OK	Your Approach [Fill in this column]	
Decide randomization strategy			
Unit: What is being randomized (e.g. individual, household, school, etc.)?	Unit: Water utility customer (sample participants are individually identified by their water billing account number)	Unit:	
<i>Method:</i> How will you conduct randomization? (E.g., through a random number generator, lottery, coin toss, randomized paper sequence, etc.)	Procedure: We conducted a complete randomization using Python.	Procedure:	
Verification: 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?	Verification: Prior to sending the SMS to the treatment group, we assigned a few City phone numbers to the control and treatment groups and tested whether the correct phone numbers received the SMS.	Verification:	
Blinding & Masking: 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?)	 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 SMS. Control group members would not be aware of the intervention or trial. 211 call takers did not know whether a caller had received an SMS message unless the caller explicitly offered that information. 	Blinding and masking:	
Spillovers: 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?	Spillovers: It is possible that a treatment group member could have told a control group member about the 211 hotline. Therefore, we may have underestimated the positive impact of our intervention because we did not capture control group participants who called 211 due to the SMS message being sent to someone in their network.	Spillovers:	
Calculate your sample size and power requirements			
Baseline: 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	Baseline: We do not have a precise baseline estimate for calls placed to a helpline during a crisis after an SMS campaign. We started from a rough estimate of the percentage of Tulsans who called 211 during the COVID crisis. We estimated this percentage to be 0.83% based on the number of 211 calls recorded in April 2020	Baseline:	

Standard deviation: If your outcome indicator is a continuous measure, what is the baseline standard deviation?

range of different baselines based on

your assumptions.

211 calls recorded in April 2020 divided by the cumulative population of the 6 largest counties served by the helpline. We believe that our targeted population may already be contacting 211 at a higher rate than the general population, and that the total population in the 211 jurisdiction

	is higher than that of the 6 largest counties. Based on data from the control group, we estimated the baseline for the involvement in eviction events to be an interval with a lower bound of 3% and an upper bound of 10%.	
 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: 5000 unique accounts. Adjusting for attrition we estimated 5,000 * 70.5% = 3,525 for one month of intervention. For a three month intervention we estimated that the total number of participants would be: 3,525 + (3,525 * 25%) + (3,525 * 25%) = 5,288, taking into account the number of residents who may enter our sample. Power calculation: We use a significance level of 0.05 and a power level of 0.80 Even at the most conservative baseline estimate of 50%, we are comfortable with the MDES of .86 pp for the Primary Outcome. 	Number of participants available: Power calculation: Per group sample size: Minimum detectable effect size:
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: There are very few trials in the academic literature that could help us estimate the percent of people who will call 211 after receiving an SMS during a time of crisis. Some of the most relevant examples come from BIT's previous work in cities. Based on those trials, we expected an average treatment effect of 1-5 percentage points. For a 3 month trial, we expected to be able to detect a 0.86 percentage point effect size.	
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: Attrition could be caused by SMS bouncebacks due to landlines or disconnected phones. We followed an intent-to-treat analysis and did not remove bounce backs. However, we adjusted for attrition. A previous SMS trial in Tulsa had found that among 1,367 SMS messages sent to residents with code violations, 29.48% were undelivered, so we used that as our attrition coefficient.	

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?	Risk: Randomization is not implemented correctly or participants receive the wrong treatment. - Likelihood: Low - Impact: High	Risk: ● ●	Likelihood: Impact: Mitigation Approach:
---	---	-----------------	--

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)? - Mitigation Approach: BIT made sure that individuals were not assigned to both groups, and assigned a unique identifier to track them through the process. BIT also conducted a pilot test with City phone numbers to make sure that the treatment assignment was correct.

Risk: 211 responders are unprepared to handle callers who mention the SMS message when calling. - Likelihood: Medium

- Impact: Low

- **Mitigation Approach:** BIT notified the utility billing staff and 211 call takers when the project began so that they would be able to answer SMS related questions appropriately.

Risk: 211 responders seek to "help" the trial by providing better service to callers who mention the SMS message.

- Likelihood: Low
- Impact: Medium

- Mitigation Approach: 211 alerted call takers that the SMS messages are being implemented, but not that a trial was being run on them or that there are outcome measures that are being tracked.

They were told that "The City of Tulsa is testing using text messages to inform residents of 211 services. If a caller asks about a text received, let them know 'the City of Tulsa is using text messages to inform residents of services such as 211'."

Risk: Additional requests for eviction prevention services overwhelm providers.

- Likelihood: None, 211 had already been working on increasing its capacity and had added additional services thanks to emergency COVID-19-related funding sources.

Risk:The COVID-19 environment may have reduced the external validity of the trial results because (1) residents may have been more receptive to calling 211 since it had been heavily publicized during the crisis,

(2) their financial situation was more dire,

(3) seeking support services felt more socially acceptable. All of these could increase the treatment effect size. In addition, (4) residents may have been more likely to have already called 211 prior to the trial period for the same reasons as in point (1). They may therefore have been less likely to call during the trial period. This could have decreased the treatment effect size.
(5) The baseline rate of calls to 211 is higher during COVID19 than in prior

	periods. This may make it more difficult to detect a significant effect.	
Consider ethical risks How might participating in the trial harm people in your sample or others? How will you monitor the trial to ensure you can detect the harm early and change your implementation if needed?	 Risk: Backfiring effect, i.e. SMS recipients have adverse reactions to the messages. Likelihood: Low, we did not identify any risks in this trial. Mitigation Approach: BIT tracked who opted out of the messages and if they responded to the SMS message with a comment. BIT monitored the comments to see if people were indicating that they were having adverse reactions to the SMS messages. 	Your answer:
Consider race & equity implications If not already included above, how will you check to see if your intervention had differential effects by race? Disaggregating your data 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?)	We have described the racial makeup of trial participants that call 211 from the treatment and control groups in order to identify any concerning patterns in how the SMS impacts different racial groups. The next SMS trial that Tulsa conducts should use an administrative dataset that includes race, so that s quantitative subgroup analysis can be performed on the outcome measures. If you have access to this data we'd recommend you build it into your analysis.	Your answer:

Plan for Data Analysis

Understand and Specify Your Variable for Analysis

There are two parts to planning for your analysis. First, think through your different data and 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. 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 pre-specified "hypothesis" you are testing that will allow you to say what statistical test & analysis you will use to determine if the intervention worked. See below for an example of Tulsa's analysis plan. If you have questions on how to choose what test to run, check out <u>resources here</u>.

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 from the mean outcome for your control group, a regression to control for other factors, or 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	Туре	Source	Measurement
WaterBillingID	Unique ID	Water billing data	Many unique IDs
Treatment	Treatment assignment	From randomization	0 - control 1 - treatment
Bounceback	Understand attrition rate - DO NOT remove from sample for analysis	From SMS platform	0 - SMS delivered 1 - SMS not delivered
Called211	Primary outcome	UniteUs	0 - did not call 211 1 - called 211 at least once
Eviction Event	Exploratory outcome	OKPolicy data	0 - not involved in eviction event 1 - evolved in eviction event
Service Received	Exploratory outcome	UniteUs	0 - no services received (including "did not call 211") 1 - at least one service received
Housing Service Received	Exploratory outcome	UniteUs	0 - no housing services received 1 - at least one housing service received
Race*1	Exploratory descriptive analysis	UniteUs	0 - white, non-hispanic 1 - other

Sample Analysis Plan

We ran **t-tests** to determine whether the mean outcome for the treatment group was statistically different than the mean outcome for the control group.

Then we conducted some **descriptive analysis** for which we compared the number of people in the treatment and control groups with the following outcomes. Although we will not be able to determine whether the SMS was the cause of any differences between the groups, or that any differences are not due to chance, the descriptive analysis will help us understand the experience of trial participants.

- Received services from 211: If t-test above was not significant, we looked at descriptive differences.
- Received housing services from 211: If t-test above was not significant, we looked at descriptive differences.
- Racial make-up of participants who call 211: We will not be able to compare the racial make-up of treated participants who do and do not call 211 because water billing data does not include race. We will look at the racial makeup of callers from the treatment and control groups to identify any concerning patterns in how the SMS impacts different racial groups.
- Call rates among evicted treatment group members: This might indicate whether this group needs more outreach.
- Types of services received by callers: This gave us insight into which services our sample needs most.

For outcomes for which there was a statistically significant difference in mean outcomes for the treatment and control groups, we calculated the cost of each additional call and service received generated by the intervention.

(# of outcomes in treatment group - # of outcomes in control group) / Total cost of sending SMS messages

Notes on the data:

According to the data sharing agreement signed by the City of Tulsa and the Community Service Council (CSC), CSC agrees to provide the City result data which does not include confidential information of the Community Service Council but does provide data pertaining to Randomized Control Trial results.

- A City of Tulsa intern will provide review and quality assurance for the code. They will ensure that:
- Data merges were performed correctly
- Missing values were identified and dealt with as needed

¹ Due to data constraints from 211, this data was not included or analyzed.

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.

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
Protocol drafted	ВІТ	June 17
Protocol approved	Tulsa	June 17
 Round 1 of SMS Sample pulled and randomized Implementation checked Trial launched: Round 1 SMS sent Responses to SMS reviewed for adverse reactions and opt-out rates 	Tulsa	June 22-26 • June 22 • June 23 • June 23 • June 26
 Round 2 of SMS Sample pulled and randomized Trial launched: Round 2 SMS sent Responses to SMS reviewed for adverse reactions and opt-out rates 	Tulsa	July 20-24 • July 20 • July 24 • July 24
 Round 3 of SMS Sample pulled and randomized Trial launched: Round 3 SMS sent Responses to SMS reviewed for adverse reactions and opt-out rates 	Tulsa	August 17-21 • August 17 • August 21 • August 21
Data received from OK Policy and 211	Tulsa	September 28
Data analysis completed	Tulsa	October 12
Quality assurance of data analysis	BIT	October 19
Final Report drafted	Tulsa	October 26
Final Report approved	Tulsa/ BIT	October 30

Your Approach:

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

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:
 - o 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:
 - o 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)
 - o 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 Tulsa scaled up their text message campaign, head back to the Replication Guide for some advice on how to scale up your result if successful!