Analysis Plan

Project Name: Decreasing abandonment of calls to the 988 Suicide and Crisis Lifeline

Project Code: 2309

Date Finalized: August 10, 2023

Project Description

This project seeks to decrease the proportion of callers to the 988 Suicide and Crisis Lifeline ("988") who abandon their calls (i.e., hang up) before being connected to a person at a local call center. Calls may be abandoned during the initial automated message and prior to making an initial routing selection, or after the automated message when the call is routed to a local call center and the caller is waiting for a connection. The objective is to reduce the abandonment rate at both points in the call process and increase the proportion of callers who are connected to a crisis counselor on the phone.

The Department of Health and Human Services (HHS), Substance Abuse and Mental Health Services Administration (SAMHSA) and Vibrant Emotional Health (who implements and administers the 988 system) aim to make changes to the integrated voice response (IVR) system and messages that callers to 988 hear when calling 988 and attempting to reach a crisis counselor. The intervention involves changing the message script and voice recording that callers hear during the initial IVR (before a call is routed) and the script, voice recording, and waiting music that callers hear while waiting to be connected to a counselor. We expect changes to the IVR and waiting message content could increase call connections through four pathways: shorter duration in an automated system, more human voice, increasing process transparency, and reducing anxiety with more calming music.

The primary evaluation question is whether the updated caller experience – including changes to content and reduced IVR length – increases the likelihood that a caller stays on the line and is connected to a crisis counselor. To answer this question the evaluation will compare outcomes for incoming callers who hear either the current version of the 988 IVR and queue messages (the control condition) or a revised version of the messages (the treatment condition).

Control condition: Callers initially hear an automated message with options the caller can choose by pressing the appropriate number. The initial IVR message is approximately 44 seconds long, at which point if the caller is still on the line they will be routed to a queue while the system finds an available counselor. After a brief pause the caller hears music after the call is routed and hears cut-in messages at 18-second intervals.

Treatment condition: Callers initially hear an automated message with options that is less than 30 seconds long. If the caller remains on the line they will be routed to a queue while the system finds an available counselor. After a brief pause the caller hears a message with affirming messages, additional resources, and that they will hear music while the system is checking for an available counselor. Music plays after the initial post-routing message with cut-in messages 30-second intervals. The voice recording was changed to a voice that research participants thought better matched the IVR objectives of being warm, calm, reassuring. . Minor changes to the IVR script have been made to reduce length and improve clarity, and a description of the option to press "0" to be routed to a counselor has been added. Post-routing cut-in messages have been revised, and the option for texting 988 has been added. The wait music has been changed.

A comparison of the treatment and control messages is shown in the Table below:

Initial 988 IVR scripts (pre-routing) – current and intervention	
Current	Intervention
You've reached the 988 Suicide & Crisis Lifeline. We are here to help. Para español oprima el número dos. To reach the Veterans Crisis Line if you are a US Veteran or service member or are calling about one, press 1. To connect to specialized support for LGBTQ+ under the age of 25, press 3 Otherwise please remain on the line while we route your call to a Lifeline crisis counselor. [Record as separate voice file from main]: Your call may be monitored and recorded for quality assurance purposes. [Number of spoken words: 81]	You've reached the 988 Suicide & Crisis Lifeline. Para español marque el número dos. To talk to the Veterans Crisis Line press 1. To connect to support for LGBTQI+ youth and young adults, press 3. Otherwise, to talk with a counselor, stay on the line or press 0. To improve 988, your call may be monitored and recorded. [Number spoken words: 67]
Waiting queue scripts (post-routing) – current and intervention	
Current	Intervention
[Music plays after a brief pause] [Music breaks at 18 second intervals for the	[Brief pause after being routed]

following message] Please stay on the line while your call is directed to an available responder.	We are checking for a counselor who is available to talk. You'll hear music while we do this, and we'll give you an update in 30 seconds. You are not alone. We care and want to support you. Someone will be on the line with you soon. [30 seconds of music] We are continuing to check with counselors for someone who is available to talk. You'll continue to hear music, and we'll give you another update in 30 seconds. If you are in the process of a suicide attempt, please call 911. Otherwise, please stay on the line. We want to help and will have someone available as soon as possible. [30 seconds of music] We know your call is urgent and thank you for your patience. If you prefer, you can text 988 to reach a counselor, or please stay on the line to be connected. [30 seconds of music] We're still working to find someone for you to talk to. While you are waiting, you can find more resources or reach a counselor through chat if you prefer at 988LIFELINE.ORG. Otherwise, please stay on the line. [30 seconds of music, then loop back to first message]

SAMHSA and Vibrant expect to deploy the new (treatment) version of the IVR and wait messages alongside the current (control) versions on August 3, 2023. Callers will be assigned to either the control or treatment condition based on the last four digits of their 10-digit phone number. OES randomized all 10,000 possible 4-digit strings, stratifying on the 2nd, 3rd, and 4th digits (corresponding to the 8th, 9th, and 10th digits of phone numbers), into intervention and comparison groups, with half in each. This random allocation was done in Stata by sorting the list of 4-digit strings by the 4th, then 3rd, then 2nd digits, and then by a randomly generated number. The first five within each set of ten was designated as intervention, and the second five were

designated as comparison. Once the study period begins, callers whose last four digits are assigned to intervention will hear the new version of the IVR and queue messages while the rest will hear the original version of the IVR and queue messages. This assignment mechanism will ensure that repeat callers using the same phone number during the evaluation field period will always be assigned to the same treatment condition.

The evaluation will observe call events when people call the 988 Suicide and Crisis Lifeline. All phone callers to the line will be included in the evaluation during the study period, except for data exclusions noted below. The current volume of calls to the 988 line indicates we can expect between 15,000 - 18,000 calls per day. The target number of observations for the full study is expected to be approximately **N = 315,000** (21-day study duration @ 15,000 calls per day; we will aim to meet the minimum sample size numbers, which may result in a slightly longer or shorter field period depending on daily call volumes).

Preregistration Details

This Analysis Plan will be posted on the OES website at oes.gsa.gov before outcome data are analyzed. In addition, the project is preregistered in the American Economics Association RCT Registry at https://www.socialscienceregistry.org/trials/11638.

Hypotheses

The primary hypothesis is that changes to the caller experience will increase the likelihood that a caller will connect to a crisis counselor.

We will examine several secondary hypotheses related to whether changes in the IVR and waiting messages and content result in different caller behavior. These include the effect of changes on:

- The likelihood that the caller stays on the line at least until the call is routed.
- The likelihood that the caller connects with a counselor via phone within 24 hours of an initial 988 call.
- The likelihood that a caller connects with a counselor via phone or text within 24 hours of an initial 988 call.

Data and Data Structure

This section describes variables that will be analyzed, as well as changes that will be made to the raw data with respect to data structure and variables.

Data Source(s):
Call event records
Records of calls to the 988 line are collected by Vibrant for all callers. For each call the following information is captured:
- Phone number-level unique identifier (i.e., a randomly generated unique id based on the 10-digit phone number and consistent across call and text data)

- Last four digits of the phone number
- An indicator for whether the call was assigned to hear the control or treatment version
- Date and time of call
- Call events (disconnect, menu selection, connected)
- Time stamp for call events
- State and county level FIPS codes associated with the phone number.

<u>Text event records</u>

Records of texts to the 988 text line are collected in a similar way as calls. These data will be used to match to call records from the same phone number and observe when callers also texted the 988 line. Relevant information captured includes:
- Phone number-level unique identifier 10-digit phone number (i.e., a randomly generated unique id based on the 10-digit phone number and consistent across call and text data)
- Date and time of text
- Call events (disconnect, connected)
- Time stamp for call events

Outcomes to Be Analyzed:

Table 1. Primary and secondary outcomes

Outcome	Primary or secondary	Description
Connect to a counselor by phone	primary	binary; =1 if a caller is connected to a counselor, =0 otherwise.
Call routed	secondary	binary; =1 if caller is routed to call center queue, =0 if disconnects prior to routing.
Connect to counselor via phone within 24 hours of initial call	secondary	binary; =1 if a call from a phone number that calls 988 is connected to a counselor on any attempt via phone, within 24 hours of first contact, , =0 otherwise.
Connect to counselor within 24 hours of initial call	secondary	binary; =1 if a call or text from a phone number that calls 988 is connected to a counselor on any attempt via any method within 24 hours of first contact, =0 otherwise.

Imported Variables:

None

Transformations of Variables:
- *NewIVR* (Intervention group status): This will be built with a fixed mapping from the last digits of the callers' phone number to the binary indicator for intervention status.
- *Answerin24:* These are created by looking through the 50 calls following the initial call from the same number, creating a variable for whether any of those within 24 hours were answered. Sort calls by caller ID and date-time of call. Loop through the next 50 calls from the same number, record the earliest one answered. This variable will be 1 if the initial call was answered or if any subsequent call within 24 hours is answered and 0 otherwise, including if there are no subsequent calls. This will be missing for all calls within the last 24 hours of the dataset.
- *Contactin24:* This will be created similarly to *Answerin24*, but in this case, we will sort all calls *and texts,* look through up to 50 contacts (calls and text messages) after the initial call, and create a variable equal to 1 if the initial call is answered, if any of the subsequent contacts are answered, and 0 otherwise. This will be missing for all calls within the last 24 hours of the dataset.

Transformations of Data Structure:
None.

Data Exclusion:
The following observations will be removed from the analysis sample:
- For all outcomes: Drop calls the research team made to confirm fidelity with the study plan, drop incoming text messages (except as used to construct other variables for observations representing calls)
- For all outcomes in main analysis: Drop calls from numbers that call an average of more than 5 times per day during the study period, which is the equivalent of 105 over the full three week trial. The goal here is to drop from the analysis calls from automated systems (i.e., robocalls) or other sources that are less likely to be from real people seeking help.
- For all outcomes in main analysis: Drop calls from numbers flagged by Vibrant as abusive or harassing or potential robo-calls.
- For all outcomes except Selects IVR Option: Drop calls in which the caller chooses to be routed through the Veterans Crisis Line, as these will leave the 988 phone system and therefore not be included in the main dataset.
- For the outcomes 1) Called back within 24 hours, and 2) Reached a person within 24 hours: Drop calls made within 24 hours of the end of the study
- For the outcome Texted within 1 hour: Drop calls within 1 hour of the end of the study.

Treatment of Missing Data:
With the automated production of the dataset from phone calls, we do not anticipate missing data, except as described above when particular groups will have missing outcome data for some secondary outcomes (for example, for callers in the last 24 hours of the study, we will not have

information about whether they called back within 24 hours). In such cases, these observations will be dropped from the sample for the relevant outcome.

Descriptive Statistics, Tables, & Graphs

- A table reporting descriptive statistics and balance tests for the intervention and comparison groups, including day of call and time of call. We will test the significance of the difference between each of these along based on an omnibus F-test of all of the differences.
- Tables summarizing results from the statistical models described below
- To illustrate the results, we will create graphs of treatment effects for each of our primary and secondary outcomes

Statistical Models & Hypothesis Tests

This section describes the statistical models and hypothesis tests that will make up the analysis — including any follow-ups on effects in the main statistical model and any exploratory analyses that can be anticipated prior to receiving the data.

Statistical Models:

Validity tests:

The validity of the experimental design relies on the last four digits of phone numbers not independently predicting the likelihood of a call reaching a crisis counselor, or any of the other outcomes of interest.

We will confirm balance by checking for differences in our primary and secondary outcomes in the three weeks before the intervention. To do this, we will estimate the same equations used in our main analysis (with primary and secondary outcome variables), using pre-intervention data. The details of this estimation are described below in the next subsection. We expect to see no more statistically significant coefficients than would be predicted by chance.

We will also confirm balance based on calls during the intervention - by comparing pre-treatment variables (time and day of call) for calls with numbers assigned to treatment and control conditions.

We will estimate the following equation, followed by an F-test of the joint significance of the coefficients:

$$W_{ij} = \alpha + \beta \ 1\{NewIVR = 1\}_{ij} + \gamma_j + \varepsilon_{ij}$$

W_{ij} represents time markers of incoming calls for call i calling from a phone number with the set of last 4 digits, j. This includes binary indicators for calls initiated between 12am and 7:59am, calls between 8am and 5:59pm, and calls between 6pm and 11:59pm, and binary indicators for calls on

each of the 7 days of the week. Then $1\{NewIVR = 1\}_{ij}$ is an indicator for being assigned to the treatment condition. Sets of fixed effects for each of the last 3 digits are represented by γ_j. If the randomization has succeeded, the coefficient, β, should not be statistically significantly different from zero any more frequently than would be expected by chance.

Effect of intervention on outcomes:
To estimate of the effect of the new IVR system on the primary outcome, we will estimate the following model using OLS:

$$Answered_{ij} = \beta_0 + \beta_1 NewIVR_{ij} + X_{ij}\delta + \gamma_j + \varepsilon_{ij}$$

where $Answered_{ij}$ represents the primary outcome of interest, an indicator variable for whether a 988 call reached a person, for call i with last-digit of phone number j. New_IVR_j is an indicator variable equal to 0 if the last four digits of the phone number were assigned to comparison and therefore the call was routed through the previous IVR system and 1 if the last four digits of the phone number were assigned to the intervention and therefore the call was routed through the new IVR. Therefore, β_1 represents the coefficient of interest, or the difference in the likelihood of a call connecting with a crisis counselor between the intervention and comparison groups, and our null hypothesis is that $\beta_1 = 0$.

X_{ij} represents a vector of pre-treatment control variables, included to increase the precision of the estimates, including day of the call (with each day of the study included as a separate indicator variable) and time of the call (with each hour of a 24 hour clock included as a separate indicator, based on the start time of the call). Sets of fixed effects for each of the last 3 digits are represented by γ_j. Observations will be clustered by the last four digits of the callers' phone number, as this is the unit of treatment assignment. Due to the high number of clusters, we do not anticipate the clustering of standard errors will affect the statistical significance of our estimates.

Confirmatory Analyses:
We will test the primary confirmatory hypothesis that exposure to the new 988 caller experience affects the likelihood of a caller being connected to a counselor. Using the estimating equation in the previous section, the null hypothesis is $\beta_1 = 0$.

We will also analyze the effect of the intervention on a set of secondary outcomes by repeating the analysis described above, estimating the same equation, but replacing $Answered_{ij}$ with each of the secondary outcomes:

Thus, we will estimate the following equation using OLS:

$$Y_{ij} = \beta_0 + \beta_1 NewIVR_{ij} + X_{ij}\delta + \gamma_j + \varepsilon_i j$$

where Y_{ij} represents each of the secondary outcome variables, and each of the other variables and conditions are as described just above. For each outcome, our null hypothesis will be that $\beta_1 = 0$.

Exploratory Analysis:
We anticipate undertaking exploratory analyses to better understand the main findings and to address descriptive questions from key partners. Below we describe several exploratory analyses and note that the details will be chosen based on the primary findings and which additional questions they suggest.

Additional outcomes:
We will explore the effect of an additional set of outcomes by repeating the main analysis with an additional set of outcomes:

- The time elapsed between the start of the call to either disconnection or connection to a counselor.
    - *CallinSeconds*: This will be built by adding *time_to_route*, *time_to_abandon*, *time_to_answer*, and *talk_time*. We will then create a set of binary indicators for whether this variable is larger than 10, 20, 30, 60, and 120 seconds, and these will be the outcome variables we will use.
- The likelihood that a caller uses the 988 text line within one hour of the initial call.
    - *Useof988TextLine:* This will be created similarly to *Answerin24*, but in this case, we will sort all calls *and texts*, look through up to 50 contacts (calls and text messages) after the initial call, and create a new variable equal to 1 if any of them are texts within 60 minutes of the initial call, and 0 otherwise. This will be missing for all calls within the last 60 minutes of the dataset.
    - We will look at this for the full sample, and also separately for the sample of callers who are connected and the sample who are not.
- The likelihood that a caller initiates a follow-up call to 988 within 24 hours of an initial 988 call.
    - *Callbackin24:* Sort calls by caller ID and timing of call, record the timing until the subsequent call. *Callbackin24* is 1 if the time to the next call is less than 24 hours and 0 otherwise, including if there is no subsequent call recorded. This requires creating a single variable for the date-time of call using the parsed year, month, date, hour, minutes of the given variable, *initiated_datetime.* This variable will be recorded as missing for calls within the last 24 hours of the dataset.
    - We will look at this for the full sample, and also separately for the sample of callers who are connected and the sample who are not.
- The likelihood that a caller selects the IVR option "0" during the call.
- The likelihood that a caller selects the IVR option "1" during the call.
- The likelihood that a caller selects the IVR option "2" during the call.

- The likelihood that a caller selects the IVR option "3" during the call.
- The likelihood that a caller selects the IVR option "4" during the call (for callers using area codes for Washington State).
    - *SelectIVR1, SelectIVR2, SelectIVR3, SelectIVR4:* These will be constructed by creating new variables based on the given variable *caller_ivr_selection.*

Thus, we will estimate the following equation:

$$Y_{ij} = \beta_0 + \beta_1 NewIVR_{ij} + X_{ij}\delta + \gamma_j + \varepsilon_j$$

where Y_{ij} represents each of the secondary outcome variables, and each of the other variables and conditions are as described just above. For each outcome, our null hypothesis will be that $\beta_1 = 0$.

Heterogeneous effects:
We will examine heterogeneous changes in abandonment rates by time of day, and day of week. This will be done by interacting each of these sets of indicators with the variable, $NewIVR_{ij}$, in the primary estimating equation.

Common call drop points:
We will use information about the precise timing of when calls were dropped and what the caller would have been hearing at that moment to investigate whether there are points in the IVR system where callers are more/less likely to hang up. To do this, we will plot the number of calls dropped at each point on a graph with time on the x-axis, labeling the x-axis with what would have been happening on the call. This will be done separately for the intervention and comparison groups. This may highlight areas for future consideration if certain wording is associated with more dropped calls.

Of particular interest is the possibility of dropped calls when the message states that the call will be "monitored and recorded." We will look for spikes immediately following this point in the intervention and comparison groups, and - if they exist - we will compare the two to see whether the intervention dampened the magnitude of the spike.

Option to press zero:
One potentially important change to the new system is to announce at the end of the IVR that a caller can press zero ("0") to speak to a counselor (whereas previously the IVR message did not announce the option to press zero). This information provision could change the action of repeat callers on subsequent calls. To better understand how this changes the experience and how callers respond to this option, we will conduct the following tests:

Are callers more likely to press zero? We will estimate the main equation on all calls with the following options:
- Pressed zero

- Pressed zero before the number of seconds have elapsed at which the option to press zero is announced in the intervention group.

Then to understand whether repeat callers behave differently, we will estimate the two equations described above along with an indicator for calling more than once within the three weeks before and the three weeks during the study period and this indicator interacted with treatment status.

We will also compare the time to connection to a crisis counselor for those who press 0 and those who do not. To do this, we will estimate the following equation:

$$SecondsToConnect_i = \beta_0 + \beta_1 Pressed0_i + X_i\delta + \varepsilon_i$$

$SecondsToConnect_i$ is the number of seconds until a call is connected to a crisis counselor, and $Pressed0_i$ is an indicator for whether the caller pressed zero at any point before being connected to a counselor (to be repeated using an indicator for whether the caller pressed zero before XX seconds, as specified above). The sample for this estimation will be all calls that ultimately were connected to a crisis counselor. This will be done separately for the intervention group and the comparison group.

Calls coming from multi-phone systems:
Preliminary evidence suggests there may be a disproportionate number of incoming calls from phone numbers ending in 0. This is consistent with calls coming from multi-phone systems in which many extensions are seen as ending in 000. With the final data, we will report what fraction of incoming calls end in 000 to describe how common calls from such multi-phone systems are.

We will repeat the primary analysis separately for calls coming from numbers that end with 000 to estimate the treatment effects for this particular sample and for calls from numbers not ending with 000.

Reconnects:
As described above, we will estimate treatment effects on whether callers called back within 24 hours, reach a counselor within 24 hours, and whether they text within 1 hour. One possibility is that the treatment increases the speed and likelihood of reaching a counselor *during the first call*, which could reduce the need for subsequent calls. We will include 2 exploratory analyses to examine this.

1) We will create an outcome variable representing whether a counselor was reached within 24 hours. This variable will be equal to one if a caller either connected to a counselor in the initial call (ie: *answered*=1) or they connected to a counselor within 24 hours. It will be zero if they both did not initially connect and did not reconnect within 24 hours.
2) We can also look separately at treatment effects on call-back rates for those who initially did connect and initially did not connect. If there is no treatment effect on initial connection, then this will be straight-forward. However, if the intervention group is more (or less) likely to initially connect, then this will need to be interpreted with caution, as

there will be differential selection by treatment status into the two groups (having connected and not having connected).

Sensitivity analysis:
We will examine the sensitivity of the results for the confirmatory hypotheses to different exclusion criteria. The main analysis will be conducted on call data that excludes calls from phone numbers that called more than 5 times per day on average during the study period. We will run similar analyses with different cutoffs, greater and lesser than 5 times per day (2, 4, 10, 24).

Inference Criteria, Including Any Adjustments for Multiple Comparisons:
For all OLS models, we will use CR2 standard errors for statistical inference, using cluster indicators for the last four digits of the callers' phone number. In all models, we will reject the null hypothesis if $p < 0.05$ for a two-tailed test on the coefficient of interest.

For the primary outcome of interest (*Answered*), we will not adjust estimates for multiple hypothesis testing.

For the analysis of the secondary outcomes we will adjust for multiple hypotheses as needed using a simulation-based approach as outlined in OES guidance for adjusting for multiple hypotheses.[1]

Limitations:
- Repeat calls or texts from different numbers: Some of the analyses rely on observing when multiple calls or texts are placed from the same phone number. We do not know if these calls are from the same individual, especially for phone numbers that may be multi-phone systems (where callers from different lines could show up as the same incoming phone number). We also do not know if an individual calls or texts multiple times but uses different phone numbers. The structure of the data and our analysis treat incoming calls from different numbers as unique callers. This limits our ability to know whether findings related to repeat callers can be fully attributed to individuals making repeat calls (from the same number), or whether other call behavior and the structure of the data masks repeat caller behavior.
- Discerning high-volume callers from robocalls: There are a small number of incoming calls from numbers that are high-volume repeat callers. Some of these may be coming from automatically dialed numbers (i.e., robocalls). But anecdotal evidence suggests that there are a small number of people seeking help who are very high volume callers. The data do not allow us to perfectly discern which numbers are robocalls and which are high-volume calls from people. We use a relatively high threshold for excluding calls from high-volume numbers (5 calls per day on average, or 105 total calls during the expected 21-day study period). But it is possible with any exclusion threshold that we will include some robocalled numbers in our analyses and exclude some real callers.

---

[1] https://oes.gsa.gov/assets/files/multiple-comparison-adjustment.pdf