Australian Government

Services Australia

Pre-call SMS Pilot

Pre-analysis Plan

June 2025

Contents

Summary table

Project title	Pre-call SMS Pilot
Partners	Services Australia The Australian Centre for Evaluation (ACE)
Evaluator (Institution)	Services Australia
Trial design (including number of arms)	Three-arm randomised controlled trial (RCT)
Unit of randomisation	Team
Stratification variables (if applicable)	Program area, work magnitude
Target group	Centrelink customers for whom a work item has been delegated to staff in the two selected Services Australia Smart Centres to finalise.
Number of clusters (if applicable)	24 teams, 300–350 staff
Anticipated number of participants	24 teams, 300–350, 6100 outbound contacts (if a three-week period)
Primary outcome measure	Call answer rates following sending of Pre-call SMS Call answer rates comparing wording in each SMS group
Secondary outcome measures	Difference across payment types

Introduction

Services Australia (the agency) have established a dedicated impact evaluation team to deliver an impact evaluation trial. One of the key objectives of the team is to deliver an impact evaluation pilot in collaboration with key internal business areas and align to industry best practice.

Pre-call SMS wording was identified as a suitable first use case. The proposal was assessed against critical questions centred on risk, impact, feasibility, resourcing and priority. The agency will design and deliver a pilot to provide indicative results on the effectiveness of updating the pre-call SMS wording.

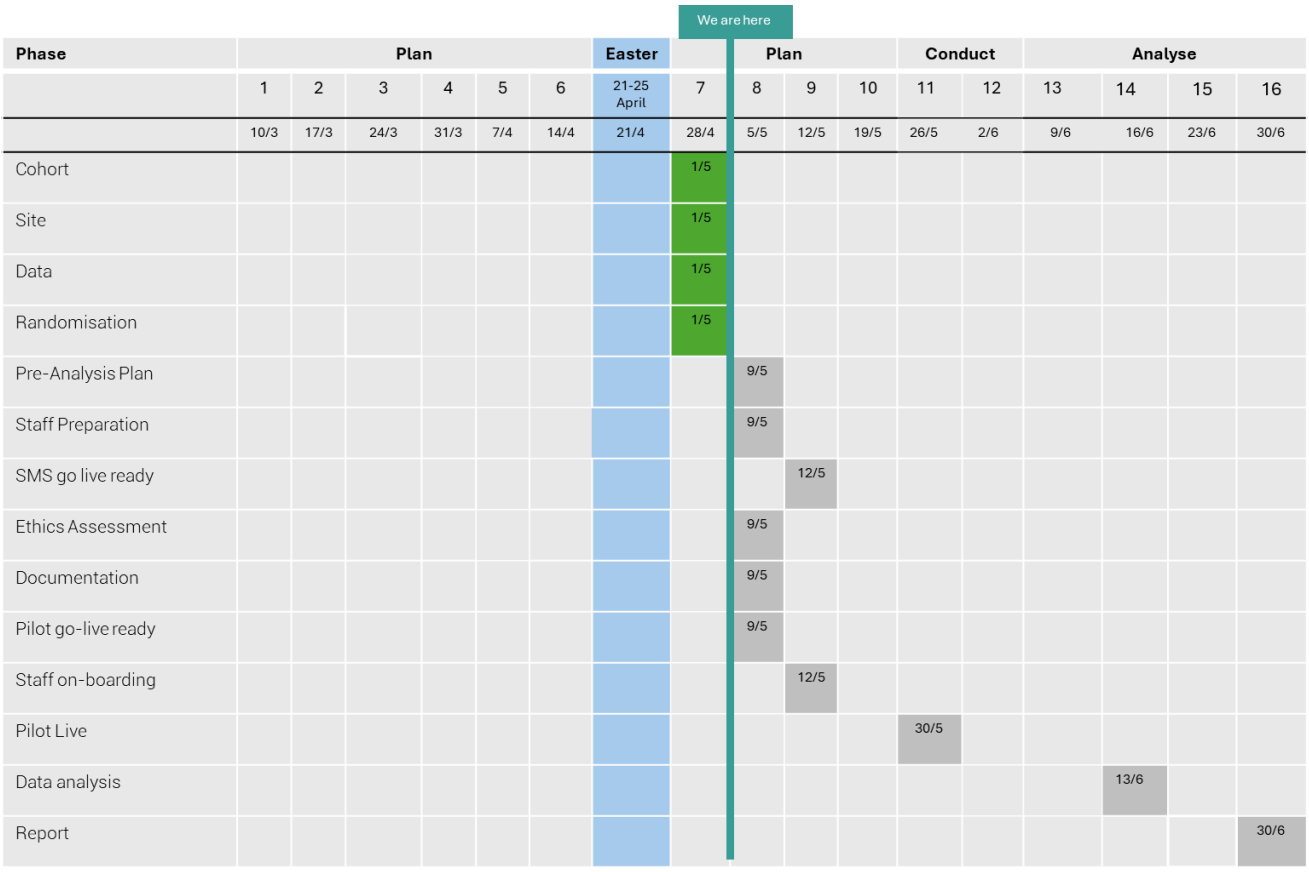
Approach

A Working Group (WG) has been established to deliver the pilot, consisting of specialist agency staff. The Australian Centre for Evaluation (ACE) will also provide weekly coaching to provide capability uplift and align to industry best practice. It is anticipated the WG will meet weekly and provide advice out-of-session. The timing of these meetings will depend on availability and alignment with key pilot deliverables. Appendix A outlines stakeholder roles and responsibilities.

Timeline

The project was initially scheduled for a 12-week duration. However, unforeseen delays occurred due to emergency management responses following Cyclone Alfred and the Easter public holidays.

Figure 1 – Pre-call SMS Pilot Timeframe highlighting key milestones to be achieved.



Note: Dates have been adjusted, including due to ex-tropical cyclone, Easter public holidays and SMS testing requirements.

Background

Pre-call SMS are currently being used for a number of different services and payments across the agency. However, there is limited evidence about the effectiveness of the pre-call SMS in prompting customers to answer the phone.

For the purpose of this pilot, the in-scope population is Centrelink customers who require an outbound call in the Families and Students program areas.

Pilot Purpose and Aim

The purpose of the pilot evaluation is to obtain indicative results on the effectiveness of sending a pre-call SMS, and whether altering the wording increases customer call answer rates.

This evaluation will inform business areas of the most effective SMS wording, whether there are differences in the impact of the pre-call SMS across customer cohorts.

More specifically, the primary aim of this pilot is to:

1. Evaluate whether sending a pre-call SMS increases customer call answer rates
2. Evaluate the impact of updating SMS wording on customer call answer rates

The secondary aim is to:

3. Explore whether customer call answer rates differ across customer cohorts

Results may be used to generate evidence about whether pre-call SMS process-guidance should be updated, or whether it may be beneficial to incorporate pre-call SMS into other processes in the agency.

This pilot will also support capability uplift across the agency, in particular designing, implementing and analysing randomised controlled trials or other forms of impact evaluation. Key lessons learned and future opportunities will be identified to support embedding impact evaluation across the agency.

Previous Research

Pre-call SMS Data

Previous analysis of agency pre-call SMS data has utilised a quasi-experimental approach. This pilot will build upon this analysis through the use of a control group for a more reliable counterfactual.

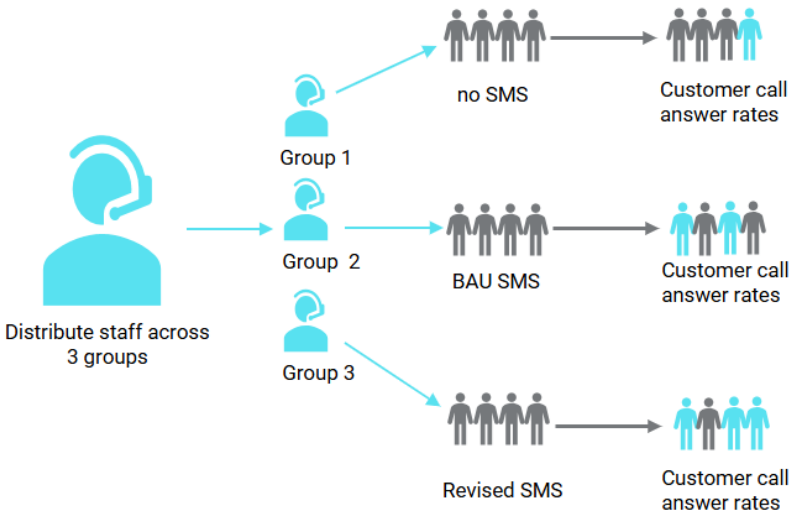
The following insights from previous analysis will guide trial design:

- SMS effectiveness slightly varies across payment types and customer age.
- There are some inconsistencies in staff adherence to guidelines on when to send the SMS

Design

The evaluation will use a Randomised Control Trial (RCT) approach (see Figure 2), whereby staff will be distributed randomly to one of three groups. Group 1 will send no message prior to making an outbound call, Group 2 will send the existing Business as Usual (BAU) message, and Group 3 will send a revised message.

Figure 2 – Visual representation of RCT approach



Intervention(s)

The table below provides the SMS wording being sent for each trial group.

Table 2 – Summary of trial arms

Control group	No SMS
Intervention group 1 (BAU SMS)	Centrelink (Alphatag) A Services Australia Service Officer will call you today from a private number. Please answer this call. Do not reply by SMS
Intervention group 2 (revised SMS)	Centrelink (Alphatag) We will call you today from a private number with some important information. Please answer this call. Do not reply by SMS

Methodology

Services Australia staff manually send the pre-call SMS to customers before calling a customer.

Outcome measure(s)

The table below provides a description of the outcome variables to be captured for the three groups.

Table 3 – Summary of outcomes captured

Outcome type	Detail	Variable type
Primary outcome	Customer call answer rate: A successful 'call answer' includes any call answered by a customer in scope for this trial that lasts over 30 seconds.	Dichotomous variable
Sub-group analysis	Customer call answer rate for each payment type with conditions as above.	Dichotomous variable
Unintended consequences	**Complaints data:** Non-anonymous complaints will be linked to individual customers, who can then be linked to the treatment arm, up to date of the data extraction. **Inbound calls:** Inbound calls will be linked to the related treatment arm, up to date of the data extraction	Dichotomous variable

Exploratory analysis will be completed post pilot to explore outcome variables of adherence to sending the SMS, time between SMS and call, length of call and work finalisation.

It is expected some trial actions will lead to subsequent activity, such as complaints and inbound calls. Without a clear unique identifier linking trial and subsequent actions, possible links can be identified by looking for subsequent actions shortly after the trial action, related to the same customer. Unless qualified, this approach is expected to produce false positives. This is usually treated by qualitative analysis, where analysts assess relevance and relationship between actions.

Population of interest and sample collection

As the trial is designed to have minimal impact on service delivery, certain exclusions can be applied in the analysis stage.

Our population of interest is Centrelink customers who require an outbound call related to the Families and Students program areas.

Inclusion Criteria

a. Work has been allocated through the Workload Management System (WLM) to service officers with the appropriate skills in the selected teams and/or Smart Centre participating in the trial.

b. Customer requires an outbound call, based on service officer's application of relevant processes.

Exclusion Criteria

Some exclusions were applied by service officers during the trial.

a. Customers not registered for electronic messaging (the system cannot send SMS to customers not registered)

b. Customer has opted out of Services Australia research

c. Customers experiencing Family Domestic Violence if indicated on their record

d. Level 1 and Level 2 Complaints and Feedback

e. Customer Identified Risk of Homelessness

f. Vulnerable Customer activities

g. Internal reviews

h. Urgent Payments

Other exclusions were applied by analysts when curating trial data.

a. Service officers that do not log into softphone will not be captured.

b. Dialled phone numbers that cannot be matched to a single customer. Outbound calls are not consistently associated with a particular customer, so the customer is identifiable if the dialled phone number can be matched exactly or with high confidence.

c. Some phone numbers are truncated in the source data asset. A data matching process can be used to match some with high confidence, but not all will be included.

d. Some SMS are sent to customers, though have not been matched them to any particular call or work item.

e. All calls from service officers that are outside of the trial teams are out of scope.

Assumptions

a. We assume customers receive an SMS at the time recorded in correspondence data. In some instances, there may be a network delay, though the extent to which this occurs is unknown.

b. Service officers may send the Pre-Call SMS multiple times to the same customer. Multiple SMS from the same service officer to the same customer within one calendar day are considered part of a single overall outbound attempt; if any of the calls are successful then the outcome will be a 1.

c. Outbound contact processes generally require service officers to attempt all current phone numbers. Multiple calls from the same service officer to the same customer within one calendar day are considered part of a single overall outbound attempt; if any of the calls are successful then the outcome will be a 1.

Hypotheses

Ha: There will be an increase in the proportion of [primary outcome = calls successfully answered by the customer] for Treatments 1/2 (standard/new SMS) compared to Control (T1 and T2 > C, one-sided test).

Hb: There will be a difference in the proportion of calls successfully answered by the customer for the T1 group compared to T2 (T1 ≠ T2, two-sided test).

Secondary (if pooled effect is statistically significant,)

Hc: There will be an increase in the proportion of [primary outcome = calls successfully answered by the customer] for Treatment 1 (new SMS) compared to Control (T1 > C, one-sided test).

Hd: There will be an increase in the proportion of [primary outcome = calls successfully answered by the customer] for Treatment 2 (new SMS) compared to Control (T2 > C, one-sided test).

Randomisation

Clustered randomisation will occur at the team level, where each cluster applies its assigned trial approach to outbound contacts it makes. After consideration of both staff and team level randomisation, it was determined randomisation at the team level is the lowest level feasible, which is expected to dampen the power of what observations we capture.

'Work items', each of a particular 'work type', are assigned to staff using a workload management program. 'Program Area', which takes either 'Families' or 'Students' values, is identified as a stratification variable for randomisation because prior knowledge demonstrates differences in SMS success across program areas. Staff cannot control which work items are assigned to them.

Some teams or staff do more relevant work than others. Stratification using a 'Work Volume' data point will avoid one trial approach having more samples than another. To simplify stratification, we will use the ordinality of the continuous Work Volume measure to create 'Work Size', which takes either 'Small', 'Medium', or 'Large' values. The smallest 1/3 of the data is assigned 'Small', the middle 1/3 'Medium' and the largest 1/3 'Large'. This is expected to account for cases where staff are completing limited work of interest, such as working limited hours and/or performing non-target work. Exclusion criteria referred to the data capture issue of outbound contacts made while not logged in; outbound contacts will be used as the 'Work Volume' data point, which may handle this issue.

Randomisation will be performed by establishing a list of teams to randomise. The list will be imported into the SAS Enterprise Guide program. The list will be processed through the SURVEYSELECT procedure, through which a pseudo-randomisation algorithm will equally split teams/staff to one of the three trial groups, also balancing program area and work volume across the groups. The list of teams and trial group assignments will be exported and distributed to relevant team leaders for distribution with their teams.

During randomisation preparation, it was noted that only three students teams were expected to be involved in the trial. Normally, two stratification variables would be specified as part of the SURVEYSELECT procedure. Where only three 'Students' teams appear available, the procedure would fail, given we require nine 'Students' teams: one team in each trial arm for every strata (unique combinations of stratification variables). The solution, as advised by the Australian Centre of Evaluation, is to create a hybrid stratification variable, including the values 'Small', 'Medium', 'Large' and 'Students'. All except the 'Students' group constitute the 'Families' teams. By effect, the procedure will randomise 'Families' teams stratified by the 'Work Size' values and simultaneously 'Students' teams without stratification.

The trial relies on the integrity of randomisation. The standard is for all staff randomised into to a trial arm to participate. Some staff may leave or enter; the extent to which this occurs and may influence results will be discussed in an additional analysis.

Sample size calculations

A power analysis was performed with respect to Ha using the POWER SAS statistical procedure.

Due to changes in the timing of the trial, and possible impacts of other work priorities that could reduce the volume of calls, we have run a range of scenario tests to determine what the required sample size should be. Statistical power tapers off once the call volumes are above 6000, which during normal work periods should be approximately 3 weeks for the 24 teams in scope. Therefore, trial length should be guided by reaching a call target of no less than 6000 calls, and preferably closer to 8000 if possible.

Intracluster correlation (ICC) captures the potential effect clustering may have on inference.

Table 4 – Power by scenario

Observations (Calls)	ICC 0.01	ICC 0.02
4000	72%	54%
6000	77%	57%
8000	80%	58%

Table 5 – Power analysis settings

Power Analysis Parameter	Selection
Test Type	Proportion/mean difference
Statistical Test	Pearson’s chi-square test / t-test
Significance Level	5% (statistical default)
Desired Power	80% (statistical default)
Test Sides	1 (superiority trial)
Group Proportions	1:2 (two groups; three arms with treatment pooling)
Null Proportion	0.56 (prior exploratory research)
Minimum Detectable Effect	0.06 (prior exploratory research)

Method of analysis

Primary outcome analysis

The primary analysis is the effect of pre-call SMS on contact success.

Point estimates, confidence intervals and p-values will be derived from an OLS regression model with the following specification:

$$Y_i = \beta_0 + \beta_t X_{it} + \beta_c Z_{ic} + \beta_{tc} X_{it} Z_{ic} + \epsilon_i$$

Where:

- Subscript *i* refers to the *i*th observation from the trial, from the first observation to the number of observations made. Like a simple average calculation, the model brings together all aspects across observations and gives us a view of the trial topic according to what we've specified in the model.
- The dependent variable Y_i is the binary contact success outcome. For each observation, if Y_i = 1 the customer answers the phone and maintains connection for at least 30 seconds, or Y_i = 0 if not.
- β_0 is the intercept, which will measure success without the presence of the pre-call SMS. β_0 can also be written as $\beta_0 X_{i0}$. The success measure without an SMS can be calculated as $Y = \beta_0$.
- The X_{it} term is the treatment assignment indicator, where {0} = control and {1} = treatment 1/2. The β_1 coefficient represents the average treatment effect for the pooled interventions relative to control. The success measure can be calculated as $Y = \beta_0 + \beta_1$.
- The Z_{ic} terms refer to covariates we expect to relate to both the outcome and the treatment. Covariates include stratification variables used when randomising into treatment and control groups. Covariates should be mean-centred where appropriate. Stratification covariates should have no statistical significance in the model.
- $X_{it} Z_{ic}$ represents the multiplication of X_{it} and Z_{ic}. When included in the regression model, the model generates parameter β_{tc}, which represents influence of covariates on treatment/control groups.
- ϵ_i is the cluster-robust standard error term.

The OLS regression model provides point-estimates and standard errors required to perform the hypothesis tests.

The primary analysis of the effect of the pre-call SMS will be intent-to-treat (ITT) and will consist of a covariate-adjusted comparison of our primary outcome.

Subgroup analysis

The purpose of this subgroup analysis is to ascertain whether the treatment effect differs for each of these subgroups of different Services Australia customers. This subgroup analysis is being conducted because there is some existing evidence that call answer rates, and the impact of the pre-call SMS differ across different customer cohorts. The subgroup analysis will build on the model specified in the primary analysis section but will add a variable for program. We will not correct for multiple comparisons due to the exploratory nature of these analysis and the pilot nature of this trial. We are conscious that this may inflate the Type I error rate but will caveat this appropriately in the reporting of these analyses.

There is potential for customer characteristics to be skewed across control/treatment groups. Another subgroup analysis topic can focus on age and gender. Another subgroup analysis is the difference in outcomes between claim and non-claim related work.

Additional analysis

Results and statistical inference will be performed with an OLS regression model. We will simultaneously use a logistic regression model, checking whether results differ between model specifications. If the analyses differ, we will use the OLS regression model and discuss differences.

The trial design is intent-to-treat (ITT). We will undertake a Local Average Treatment Effect (LATE) analysis, comparing the difference between primary results with results limited to staff who followed the direction of their trial arm.

Staff may leave the trial. We will discuss the extent to which this occurred and what effect it may have had on the results.

Trial Threats

Missing data

Missing outcome data may occur due to:

- Attrition: some staff may drop out of service delivery during the implementation of the trial.
- Technical issues: there could be system errors in the online system.
- Natural disaster: in the event of a natural disaster, prioritisation and emergency procedures may impact on trial delivery.
- Incorrect phone number: some customers may not receive the SMS or answer the call due to an incorrect phone number recorded.
- Incorrect address: Incorrect address means we may not be aware of customer's time zone.

Mitigation strategies:

Mitigation strategies for missing data include data validation, quality assurance, randomisation and aiming for a sample size.

Blinding and evaluation-driven effects

Since the trial involves sending staff-triggered (i.e. not automated or system generated) pre-call SMS reminders, it is impossible to blind the staff to their treatment status.

Mitigation strategies:

As a mitigation, the analysts will only know the treatment assignment during the data analysis phase, and do not have direct communication with the staff participating in the trial.

SMS is a standard method of communication between Centrelink and customers, making it unlikely that customers will be aware they are part of a trial. Therefore, we do not believe there will be significant evaluation-driven effects in this study.

Non-adherence

Staff non-adherence may occur in multiple forms and may include:

- Staff may participate in administrative sharing in that they may talk about their particular treatment process, and invertedly impact other staff issuing their assigned SMS (or no SMS).
- Staff may not adhere to the recommended timing as per process instructions.
- Staff allocated to a treatment group may not send the required message
- Staff in the control group (no message) may send a message
- Staff in any group may send a message after call is made
- Staff send the message to an out-of-scope customer cohort

These would result in spillovers or contamination.

Mitigation strategies:

To mitigate these potential issues, staff and Site Supervisors will receive extensive pre-briefing detailing the instructions to follow. Additionally, a process will be established to enable participating staff to clarify concerns or queries throughout the trial period.

Quality Assurance (QA) Process

QA will attempt to:

- Conduct rudimentary code review to verify cohort extraction
- Independently replicate, for both primary outcome analysis and subgroup analysis:
  - Contact success rates for each cohort
  - Significance levels
  - Statistical power
- QA will also assess the extent to which customers in each sample appear to represent their respective cohorts.

Interpretation of results

The primary analysis will summarise if sending a pre-call SMS improved the outbound contact outcome. It will also summarise if different types of pre-call SMS give different outcomes. These results will be compared to previous exploratory analyses and trials.

The power analysis will be repeated to calculate power based on the actual number of samples per group. This trial aims to use statistical rules-of-thumb: 80% power and a 5% significance level.

An interpretation of results will discuss the extent to which we achieved a randomised control trial. This should include, but not limited to, robustness checks and sensitivity analyses, missing data, multiple comparisons, outliers and exclusions. This may be the subject of a separate and more comprehensive report.

Pre-analysis plan commitments

No trial data has been collected, and no analysis has been undertaken prior to the completion of this pre-analysis plan. Any additions or omissions from this plan will be recorded in the Findings Report.

The pre-analysis plan and evaluation plan will be sent for informal pre-registration.

servicesaustralia.gov.au