Statistical Analysis Plan (SAP)

TRIAL FULL TITLE	Using a Simplified Tool to Predict Lung Cancer Screening (LCS) Eligibility: Experiment to Test Effects of Messaging on Response Rates
SAP VERSION	V1
SAP VERSION DATE	November 25, 2023
TRIAL STATISTICIAN	Kristin Linn, PhD
TRIAL PRINCIPAL INVESTIGATOR	Katharine Rendle (PI), Anil Vachani (Co-PI), Justin Bekelman (Co-PI)
SAP AUTHOR(s)	Katharine Rendle (PI), Kristin Linn (Statistician)

1. TABLE OF CONTENTS

2. ABBREVIATIONS AND DEFINITIONS

AE	Adverse Event
EHR	Electronic Heath Record
LCS	Lung Cancer Screening
LDCT	Low-Dose Computed Tomography
SAP	Statistical Analysis Plan
USPSTF	United States Preventive Services Task Force

3. INTRODUCTION

a. Preface

Despite growing evidence that lung cancer screening (LCS) reduces lung cancer-specific mortality, LCS across the United States and at Penn Medicine is remarkably low (Fedewa et al. 2022). This is due in part to challenges with identifying adults who meet eligibility criteria for lifetime smoking intensity (i.e., 20 pack-years or greater), which is often missing from the electronic medical record. We developed a simplified tool (using yes/no questions) that has been shown to estimate pack-year eligibility accurately (Rendle et al. 2023). In this study, we aim to test the impact of different strategies (i.e., introductory and norming messages and incentives) on survey response rates when the survey is delivered via text message.

b. Scope of the analyses

These analyses will compare the effect of three different factors (introductory messages, norming messages, and incentives) on survey response rates using a 3x2x2 factorial design.

4. STUDY OBJECTIVES AND ENDPOINTS

a. Study Objectives

The primary objective of this analysis is to assess how the combination of different messaging content and incentives impact response rates to the survey.

b. Endpoints

The primary endpoint is full completion of the survey defined as responding to all three survey questions within 7 days of initial outreach. The secondary endpoint is partial completion of the survey (defined as responding to at least one survey question within 7 days of initial outreach.

5. STUDY METHODS

a. General Study Design and Plan

This single-blind study uses a 3x2x2 factorial design. Upon confirmation of their name, patients are randomized to one of the 12 potential study arms immediately based on two stratification variables (race and sex). If the patient does not confirm their name via text message, the patients will not be randomized nor included in the study. Upon randomization, the participants immediately receive the assigned interventions via text message.

3x2x2 Study Design	*Tobacco Use Message A*		*Tobacco Use Message B*	
	Incentive	*No Incentive*	*Incentive*	*No Incentive*
Introductory Message A	Arm 1	Arm 2	Arm 3	Arm 4
Introductory Message B	Arm 5	Arm 6	Arm 7	Arm 8
Introductory Message C	Arm 9	Arm 10	Arm 11	Arm 12

b. Inclusion-Exclusion Criteria and General Study Population

Inclusion Criteria: Based on available EMR data at the time of enrollment, patients will be eligible if they:

- meet age eligibility (50-80 years old) for LCS based on 2021 USPSTF guidelines
- have completed at least one primary care visit at Penn Medicine in 2020-2025
- have a mobile phone listed in their medical record
- confirm identify at the time of initial outreach via text message.

Exclusion Criteria: Based on available EHR data at the time of enrollment, patients will be ineligible if they:

- have a documented history of lung cancer
- have a documented history of completing LCS at Penn Medicine
- are listed as not wanting to be contacted or solicited for research
- do not otherwise meet inclusion criteria.

c. Randomization and Blinding

Participants will be randomized into one of 12 study arms at the time of name confirmation. Randomization will occur using the randomization module within the Way-to-Health platform and be stratified by sex (female vs non-female) and race (Black adult vs non-Black adult). Single blinding (statistician) will be used in this study. Participants are randomized in blocks of 24 (12 study arms).

d. Study Assessments

Index date will be defined by the date that the participant is initially sent the first outreach text message (defined as the message that asks patients to confirm their name). At this point, the participant has seven days to respond to this text message, upon which they will be randomized immediately and given the assigned intervention. If they do not confirm their name or if they decline within 7 days, the patient will not be randomized or receive the intervention. For primary analysis, we will use the analysis time window below (7 days). To understand potential long-term reach, we will also assess the number of patients that responded after 7 days, including those who responded with name confirmation after 7 days and those who attempted to complete the survey after 7 days.

Primary and Exploratory Analysis Time Windows

Days From Index Date	Lower bound (days)	Upper bound (days)
Day 0 (Primary Analysis)	0	7
Day 8 (Exploratory Analysis)	8	60

6. SAMPLE SIZE

To determine the sample size for the trial, we considered a range of hypothesized response rates from 3% to 12% for the 12 arms representing different combinations of messaging and incentives and based on prior research studies. We hypothesized this range of response rates would be plausible for many of the arms and chose the final sample size based on our ability to distinguish statistically two arms towards the extremes of the range. Figure 1 below shows a plot of the 95% confidence intervals for the range of hypothesized response rates (plotted on both x- and left y-axes). With N=6,000 participants total (500 per arm), we would be able to detect a 5-percentage point difference between two arms if one of the arms had a response rate of 3% to approximately 7%. In comparison, with N=3,600 we would only be able to detect a 5-percentage point difference between two arms if one of the arms had a response rate of 3% or lower. Although larger samples would allow us to detect smaller differences, piloting on a larger sample would decrease the number of participants available for future trials. Balancing this logistical concern with power to detect differences between arms, we chose to include N=6,000 participants in this pilot study.

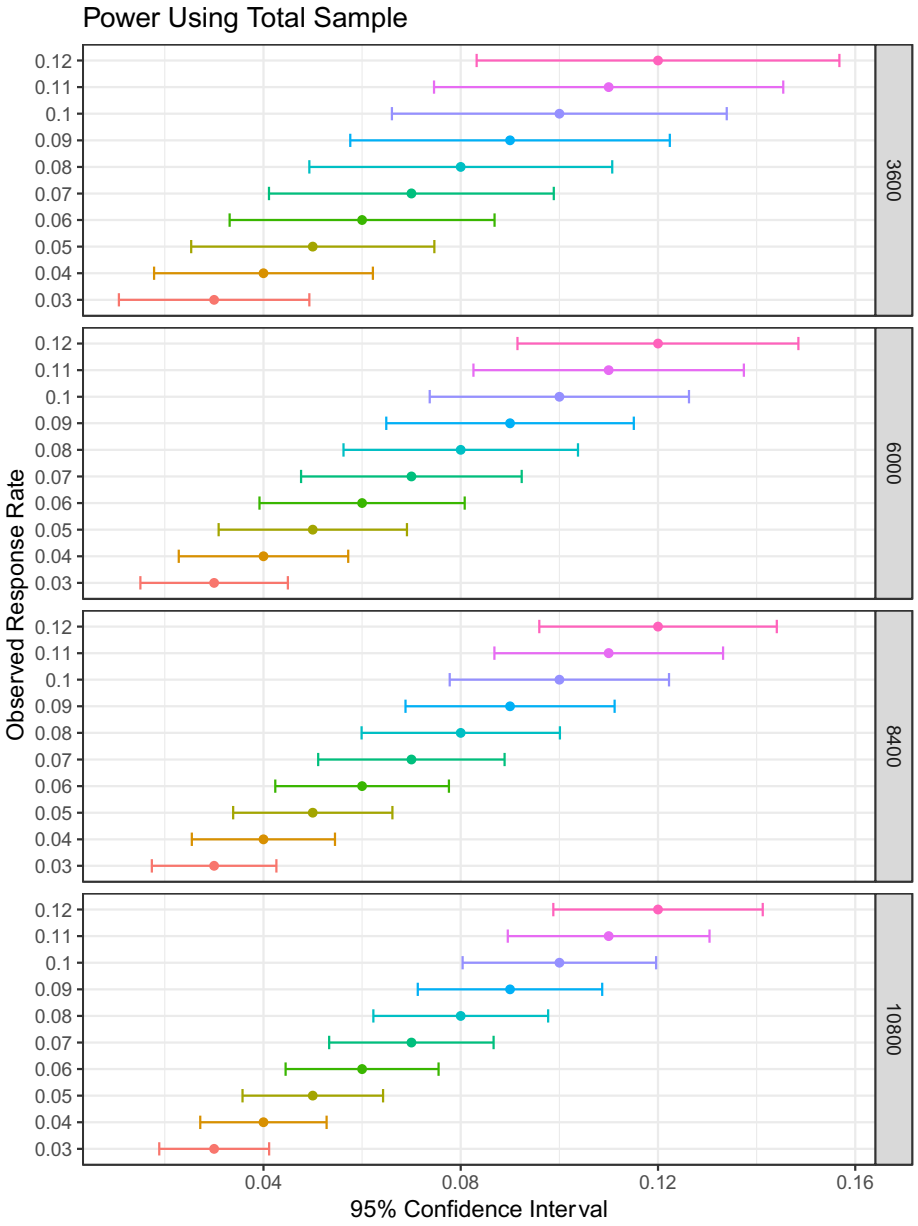


Figure 1: *Illustration of 95% confidence interval length for potential response rates between 3% and 12%. Response rates are plotted on both the x- and y-axes. Each panel plots confidence intervals for the total sample size (N) given on the right y-axis.*

7. GENERAL ANALYSIS CONSIDERATIONS

a. Timing of Analyses

The final analysis will be performed when at least 6,000 participants have randomized. The final analysis will be performed after all necessary cleaning and approval of statistician and after the finalization and approval of this SAP document.

b. Analysis Populations

i. Intention to Treat

All subjects who were randomized will be included in the intention to treat analysis. This will be the primary efficacy population for the purposes of the study.

ii. Per Protocol Population

All subjects who responded in any form (e.g., any text message response including but not limited to survey response) after receiving the intervention (i.e., after all interventions are delivered) will be included in the per protocol analysis. This will be the secondary efficacy population for the purposes of the study. This will exclude participants in the ITT that only confirmed their name and did not reply in any form after that.

c. Covariates and Subgroups

Covariates including baseline demographic (e.g., age, race, sex, ethnicity) and smoking history data (e.g., tobacco use status, lifetime smoking) will be extracted from the EHR. Variables used in randomization (sex and race) will be adjusted for in the primary analysis. We will assess baseline characteristics of each group by key covariates to assess any potential differences. Given persistent disparities in lung cancer by race and in LCS guidelines by race and sex, we will conduct exploratory subgroup analysis to assess potential treatment heterogeneity by race (Black adults vs Non-Black adults) and sex (female vs non-female). We do not hypothesize that our treatment will be different for each group, but given disparities, we conduct exploratory analysis to assess heterogeneity of treatment effects by race and sex.

d. Missing Data

The primary analysis model will adjust for race and sex. Since both these variables must be present in the EHR for randomization, there will be no individuals excluded due to missing race or sex in the analysis. There may however be participants in the dataset that have “unknown” race documented in the EHR. For randomization, these individuals are classified as “Non-Black” adults. For summary statistics, these individuals will be reported as unknown race. Due to the large number of patients to be randomized (N=500 per arm), we expect randomization to provide balance on all baseline covariates and therefore will not adjust for any other variables in the models. By definition, there will be no missing primary or secondary outcomes.

e. Multiple Testing

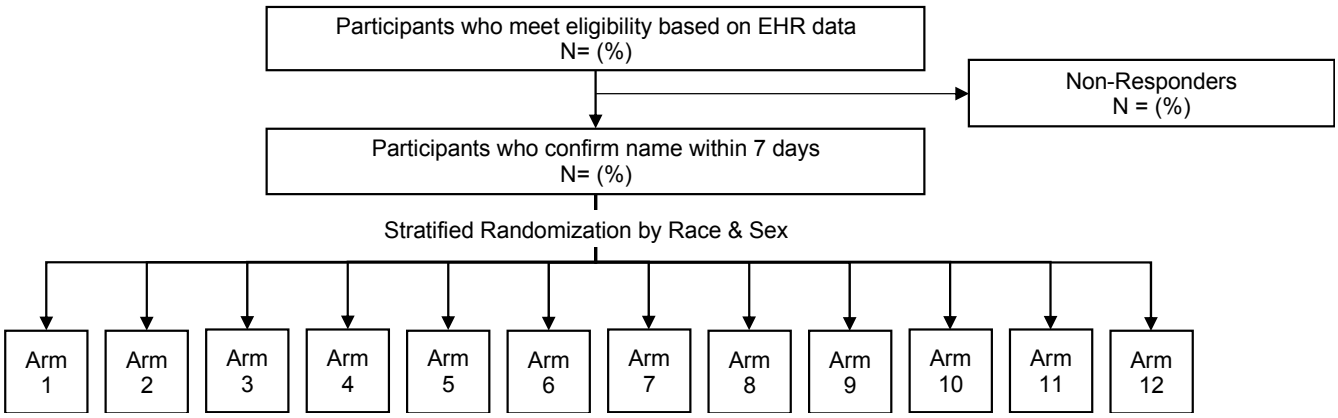
Since we are aiming to identify the best combination of factor levels, we are more concerned with making Type II errors than Type I errors. Therefore, we will not adjust for multiple testing.

8. SUMMARY OF STUDY DATA

All continuous variables will be summarized using the following descriptive statistics: n (non-missing sample size), mean, standard deviation, median, maximum and minimum (as applicable). The frequency and percentages (based on the non-missing sample size) of observed levels will be reported for all categorical measures. In general, all data will be listed, sorted by study arm and subject, and when appropriate by outcome within subject. All summary tables will be structured with a column for each intervention and will be annotated with the total population size relevant to that table, including any missing observations.

a. Subject Disposition

We will identify participants to be included in ITT based on randomization assignment among all those contacted. Allocation assignment will be assessed using categorical variable assigned (defined as Arm 1-Arm 12) assigned at the time of randomization and stored within the survey platform. In order to identify patients who have completed the survey (primary outcome) or partially completed the survey (secondary outcome), we will use variables assessed from the survey platform (WaytoHealth) for each subject. The survey questions are coded based on the response (Yes =1; No=0) for each of the three questions. We will also conduct visual evaluation (i.e., review text responses) for any question where the natural language processing is less than 90% confident (as assigned by the survey platform). Below is a draft consort diagram that to be populated with number and proportion of participants in each cell.



b. Derived variables

The survey platform (WaytoHealth) uses natural language processing to interpret the text-based response

into a derived variable (Yes or No), using a pre-specified confidence threshold of 0.9. For any response, that does not reach this confidence threshold, the platform asks for clarification (i.e., “Please answer Yes or No”) from the participant. To ensure quality of the derived variables, we will review a random sample of questions that classified as yes or no (50 each). If we identify that more than 5% are misclassified by the platform, then we will review all responses. For any survey responses with a confidence threshold lower than 0.9, two members of the analytic team will independently review each response for interpretation and classification. If no agreement is reached, the survey response will be marked as unknown.

c. Protocol Deviations

We do not anticipate any major protocol deviations that could impact analysis, but one that would require adjustments would be if participants receive a different intervention than they are assigned to receive. This unlikely event might happen if there is an error in the programming code within WaytoHealth. Prior to analysis, we will review a subset of 50 randomly selected participants to review the content of all intervention messages received and assess concordance with their assigned study arm. If we identify that more than 5% received an incorrect intervention message, we will review all participants to ensure we have correctly classified them in the right intervention arm. After this, we will conduct subgroup analysis with and without misclassified participants to understand the impact on analysis and report our findings. Any misclassified participants might be excluded primary analysis depending upon subgroup analysis and reported as such.

d. Demographic and Baseline Variables

Demographics and Baseline Variables	Source
Age at index date	EHR
Race	EHR
Ethnicity	EHR
Sex	EHR
Sexual Orientation and Gender Identity	EHR
Socioeconomic Status (Based on Census-Tract)	EHR
Tobacco Use (Current, Former, Never, Unknown)	EHR
Pack-Years (Lifetime Smoking)	EHR
Other Smoking Data (E.g., Quit Date, Types)	EHR

9. EFFICACY ANALYSES

a. Primary Efficacy Analysis

All summary statistics will be produced in accordance with section 8, and all primary efficacy analyses will use the ITT population. All tests will be two-sided and use a Type I error of 0.05. First, we will test for a main effect of incentive versus no incentive using an independent proportions Z-test. If this main effect is significant, we will proceed to study the two messaging factors using only data from individuals randomized to the incentive condition with higher response rate. If the main effect of incentive is not significant, we will proceed with studying the two messaging factors using all data and assuming incentive level does not interact with messaging or have any other effect on the outcome. Next, using either half or all the data, we will model the binary indicator of complete survey response or non-complete survey response within 7 days using a logistic regression. The model will adjust for race and sex and will incorporate 5 indicators of treatment arm (3 introductory messaging levels x 2 tobacco use messaging levels – 1 reference group). We will compute the marginal response probability for each arm, averaging over the observed distribution of sex and race. Standard errors and 95% confidence intervals will be obtained for each marginal response probability using the delta method. If one of the confidence intervals has a lower limit that exceeds the upper limit of all the others, its corresponding set of factor levels will be used in the primary study, pending confirmation that exploratory analyses do not raise red flags about disparities in response rates. If a set of confidence intervals have lower limits that surpass the upper limits of the other intervals, the final choice of factor levels to use will be based on discussions amongst the study team and key stakeholder, taking into consideration the point estimate, confidence interval length, and exploratory data on disparities.

i. Secondary Analyses of Primary Efficacy Endpoint

We will conduct secondary analyses of the primary efficacy endpoint using the per protocol, instead of the ITT population, to assess if documented engagement after receipt of the full intervention affected efficacy.

ii. Analyses of Secondary Endpoints

We will conduct similar analysis outlined in Section 9a to assess the effect of the interventions on the secondary outcome, partial completion of the survey.

b. Exploratory Efficacy Analyses

We will conduct exploratory efficacy analyses within each of the following pre-specified subgroups: 1) females vs non-females and 2) Black adults vs non-Black adults (using EHR data). We conduct primary and secondary efficacy analysis within each of these subgroups in accordance with section 9a and 9b above. The summary statistics will be produced in accordance with section 8. This subgroup analysis is justified because there are existing disparities in LCS guidelines in each of these groups and we aim to understand if our strategies may help to reduce or increase disparities before implementing them broadly.

10. SUMMARY OF CHANGES TO THE PROTOCOL AND/OR SAP

This statistical plan outlines analysis for the experimental trial, which is part of a larger, multiphase study that will be informed by this phase of study.

11. REFERENCES

Fedewa SA, Bandi P, Smith RA, Silvestri GA, Jemal A. Lung Cancer Screening Rates During the COVID-19 Pandemic. Chest. 2022 Feb;161(2):586-589. doi: 10.1016/j.chest.2021.07.030.

Rendle KA, Steltz JP, Cohen S, Schapira MM, Wender RC, Bekelman JE, Vachani A. Estimating Pack-Year Eligibility for Lung Cancer Screening Using 2 Yes or No Questions. JAMA Netw Open. 2023 Aug 1;6(8):e2327363. doi: 10.1001/jamanetworkopen.2023.27363.