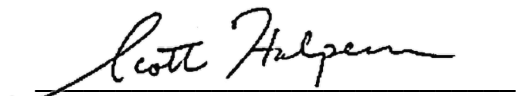
Statistical Analysis Plan

Study Name	Comparing Smoking Cessation Interventions among Underserved Patients Referred for Lung Cancer Screening
Version Number	2
Date	11 November 2025
ClinicalTrials.gov #	NCT04798664
Sponsor Information	Patient-Centered Outcomes Research Institute 133 New Hampshire Ave NW Suite 1200 Washington, DC 20036

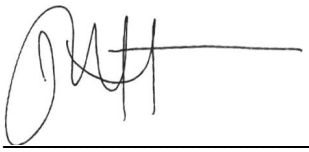
Comparing Smoking Cessation Interventions among Underserved Patients Referred for Lung Cancer Screening

SIGNATURE PAGE

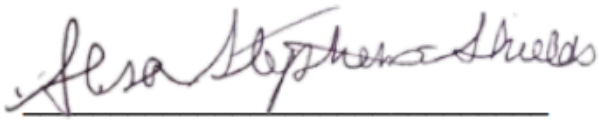
Approved by:



Scott D. Halpern, MD, PhD
Principal Investigator
University of Pennsylvania School of Medicine
Palliative and Advanced Illness (PAIR) Center



Joanna Hart, MD, MSHP
Co-Investigator
University of Pennsylvania School of Medicine
Palliative and Advanced Illness (PAIR) Center



Alisa Stephens-Shields, PhD
Statistical Investigator
University of Pennsylvania School of Medicine
Palliative and Advanced Illness (PAIR) Center

TABLE OF CONTENTS PAGE

LIST OF ABBREVIATIONS

The following abbreviations and special terms are used in this Statistical Analysis Plan (SAP).

Abbreviation or special term	Explanation
AAR	Ask-Advise-Refer
Co-I	Co-Investigator
DSMB	Data and Safety Monitoring Board
EFT	Episodic Future Thinking
GLM	Generalized Linear Model
ITT	Intension To Treat
LDCT	Low Dose Computed Tomography
Penn	University of Pennsylvania
PI	Principal Investigator
SAP	Statistical Analysis Plan
SSN	Social Security Number

INTRODUCTION

This statistical analysis plan (SAP) is a comprehensive and detailed description of the strategy, rationale and statistical techniques that will be used in the smoking cessation study.

1. STUDY DETAILS

1.1 Study Design

We will conduct a 4-arm randomized trial comparing 4 interventions-

i. Basic usual care: Ask-Advise-Refer (AAR) approach,

ii. Enhanced usual care: usual care plus free nicotine replacement therapy and FDA-approved pharmacotherapies. The intervention is to facilitate access to pharmacotherapies -- e.g., encourage to use, encourage to go to doctor and financial assistance through voucher,

iii. Enhanced usual care plus financial incentives to stop smoking,

iv. Enhanced usual care plus financial incentives plus a mobile health application that motivates patients to think about their future health by promoting episodic future thinking (EFT) to promote sustained, biochemically confirmed smoking abstinence for 6 months among smokers in underserved demographic groups.

The participants to be enrolled will be current smokers who are black, Hispanic, and/or have low socioeconomic status (defined as household income <200% of the federal poverty line or a high school education or less) or rural residence who are referred for low-dose computed tomography (LDCT) screening at 4 large health systems. Patients with LDCT orders will be identified via the electronic health record and further screened for eligibility. Eligible patients will enroll and complete the study using the NIH-funded *Way to Health* portal. All of the following patient inclusion criteria must be met:

1) Age ≥ 18 years

2) Current smoker (≥ 1 cigarettes per day, not including e-cigarettes)

3) Has a LDCT scan ordered by his or her physician

4) Able to receive study invitation and screening, by virtue of showing up to a radiology location affiliated with 1 of the 4 health systems for the LDCT, or having a valid email address or telephone number on file with the health system

5) Underserved, defined as one or more of the following:

i. Black

ii. Hispanic

iii. Rural residence

iv. Low socioeconomic status, defined as one or both of: a. ≤ high-school education or less

b. household income < 200% of the federal poverty line

6) Access to a phone with text messaging or the internet

7) Has a lung cancer screening order, meet our criteria for being underserved, and self-report smoking at least 1 cigarette per day on average for the past month

It is expected that LDCT orders will be placed by physicians in line with national screening guidelines, i.e., patients with a ≥ 30 pack-year smoking history and tobacco use within the past 15 years. The majority of enrolled patients are expected to be ≥ 55 years; however, patients < 55 years meeting all other eligibility criteria are eligible for this clinical trial.

1.2 Study Endpoints, Hypotheses, and Corresponding Outcomes

The primary outcome will be biochemically confirmed, sustained abstinence from smoking tobacco for 6 months following participants' selected quit dates. For patients not reporting the use of NRT, a urinary sample with cotinine levels at 50 ng/ml or less will confirm cessation. Biochemical levels of cotinine between 21-50 ng/ml will show up as positive on test results, but will be considered a false positive test. For participants reporting the use of NRT, a urinary sample with anabasine levels less than 3 ng/ml will confirmation cessation. For patients only using e-cigarettes, a blood carboxyhemoglobin level less than or equal to 4% will confirm cessation. Secondary smoking-related outcomes include point-prevalent quit rates at 2 weeks and 3 months, and relapse rates at 12 months. Patients who become deceased prior to completing the program will be removed from the denominator of future quit rate outcomes.

Secondary patient-reported outcomes include:
• Motivation to quit
• Self-efficacy related to cessation efforts
• Perceived barriers to cessation
• Health-related quality of life

Hypothesis 1a: Each of the three intervention arms will increase the rate of 6-month sustained smoking abstinence compared to Ask-Advise-Refer.
Hypothesis 1b: Financial incentives plus free access to pharmacotherapy (nicotine replacement therapy and FDA-approved pharmacologic cessation aids) will increase the rate of 6-month sustained smoking abstinence compared to free access to pharmacotherapy alone
Hypothesis 1c: Episodic future thinking (EFT) plus financial incentives plus free access to pharmacotherapy (nicotine replacement therapy and FDA-approved pharmacologic cessation aids) will increase the rate of 6-month sustained smoking abstinence compared to financial incentives plus free access to pharmacotherapy

Hypothesis 2: There will be significant treatment heterogeneity (hypothesis test of $\beta = 0$) in the foregoing contrasts by pre-specified patient-level covariates at $\alpha = 0.10$ level significance.
Hypothesis 3: Each of the three intervention arms will improve health related quality of life and will improve rates of 12-month smoking abstinence.

1.3 Randomization Scheme

We plan to enroll 3200 patients to four arms (usual care, enhanced usual care, enhanced usual care plus financial incentives and enhanced usual care plus financial incentives plus a mobile health application EFT). Participants will be randomized individually, stratified by health system, using random number generation within the *Way to Health* platform. At the Penn, Geisinger, and Henry Ford sites, 25% of enrolled participants will be assigned to each of the 4 arms. Participants enrolled at Kaiser sites will only be assigned to Arms 2-4, each with a 33% probability. This is because Kaiser provides free cessation aids to all its patients, and as such, there exists no basic usual care (control) arm. Based on the expected accrual at each of the health systems, we expect that overall there will be 470 patients in basic usual care (Arm 1) and 910 patients in each of intervention Arms 2-4.

2. ANALYSIS SETS

Intention-to-treat analyses will include all randomized patients regardless of their adherence to protocol or the actual treatment received. Patients will be analyzed using their assigned treatment arm. All events occurring within the specified time periods will be included in the analysis regardless of whether the patient adheres to the protocol. Information from patients who withdraw their consent to participate in the study will be included up to the time when they withdraw their consent. The primary analysis for all primary, and secondary endpoints will be performed as intent-to-treat analyses.

3. ANALYSIS METHODS

3.1 General Principles

Descriptive Statistics

Numerical variables will be summarized using standard summary statistics including the number of participants, mean, standard deviation, median, 10^{th} and 90^{th} percentile and range (i.e., minimum and maximum) as appropriate. For categorical data, proportions will be presented in a frequency table format.

Demographic and baseline clinical and biological variables will be summarized and compared by treatment arms. Continuous variables will be summarized using means, medians, ranges, standard deviations, and interquartile ranges. Frequency distributions will be presented for categorical and ordinal variables. A CONSORT diagram displaying the number of patients screened, eligible, and consented along with reasons for ineligibility will be provided. Graphical tools such as histograms, boxplots, and scatterplots will be created to assess quality of data and to display patterns over time. We expect that missing data will be minimal; however, we will fully describe any missing data.

3.2 Analysis Methods

Primary Analysis

The analysis of the primary outcome will use a generalized linear model (GLM) with a logit link function to estimate the effect of treatment arms to usual care. All models will be adjusted for health system, and patient level covariates.

The following model will be used to test Hypothesis 1a.

$$g\big(E(Smoke_i)\big) = \beta_0 + \beta_{11} patient\ covariate_{1i} + \beta_{12} patient\ covariate_{2i} + \cdots + \beta_2 arm_{2i} + \beta_3 arm_{3i} + \beta_4 arm_{4i} + \cdots + \gamma_2 HS2_i + \cdots + \gamma_4 HS4_i + \gamma_5 HS5_i \quad (1)$$

where *g()* is the logit link function, *E()* represents the expected value, the subscript *i* corresponds to patient *i*, arm_{ki} equals 1 when patient *i* is randomized to the arm *k* of the three treatment arms, and $HS2_i$, …, $HS4_i$, $HS5_i$ equals 1 when patient *i* is from health system 2, …, 5, respectively (note that patients from the largest health system will make up the reference group). We will test $\beta_k = 0, k = 2,3,4$ using a two-sided Wald test at the 0.05 significance level to decide whether to reject Hypothesis 1a in the final analysis. We will do adjustment for multiple tests to maintain family-wise type I error rate according to Proschan (1999)[1].

Table 1: Patient level covariates

Variable	Definition
Age	Continuous
Gender	Male vs. not male
Race	Categorical (Black/not)
Ethnicity	Hispanic vs. Not Hispanic
Education	Categorical (5 levels)
Income (%FPL)	<200% FPL vs >= 200% FPL
Rurality (self-report)	Rural vs non-rural (use RUCA if self-report is missing)
Tobacco/ Nicotine dependence	cigarettes per day from eligibility survey

LDCT screening results	Reassuring vs. Concerning (must be on or before final quit date)
Number of prior LDCT screening tests (prior to enrollment)	Integer
Presence of high burden of chronic comorbid illnesses	Elixhauser score
Time of accrual	Pandemic era (launch through 9/18/22) vs non (1,0)
Insurance	Medicare vs Medicaid vs Commercial

To address Hypotheses 1b & 1c we will use the structure of the model (1) where we will include 1 dummy variable for treatment arm variable in each model.

Heterogeneous treatment effects (HTE) will be explored by testing statistical interactions between pre-selected patient characteristics (listed in Table 2) and each contrast between 4 interventions in logistic regression models. This will help guide future targeting of optimal interventions for individual patients. We will evaluate each interaction term separately in a model adjusted for all main effects and health system to estimate stratum-specific effects of each characteristic on the interventions' effectiveness. We will also report results from a fully adjusted model that retains all significant interaction terms. This analysis will inform our hypothesis 2. In analyzing the interaction terms we do not plan to incorporate correction for multiple testing.

Table 2: Potential effect modifiers

Variable	Definition
Age	Continuous
Gender	Male vs. not male
Race	Black vs not Black
Ethnicity	Hispanic vs. Not Hispanic
Rurality (Self Report)	Rural vs non-rural (use RUCA if self-report is missing)
Income (%FPL)	<200% FPL vs >= 200% FPL
Smartphone with Internet Use	Phone w/ data plan vs not
Tobacco/ Nicotine dependence	cigarettes per day from eligibility survey
LDCT screening results	Reassuring vs. Concerning (must be on or before final quit date)
Number of prior LDCT screening tests (prior to enrollment)	0 scans vs 1 or more scans
Presence of high burden of chronic comorbid illnesses	Elixhauser score
Mental Health Diagnoses (EHR)	Diagnoses vs not

Sensitivity analysis

We plan to analyze the primary outcome with ITT sample structured similar to (1) but will only include health system as fixed effects and no patient-level covariates. If we find any patient covariate that we prespecified for inclusion in the model and has missing values > 15% then the above unadjusted analysis would become our primary analysis.

As we are considering levels of cotinine between 21 and 50 ng/ml to be a false positive, we will also do a sensitivity analysis coding those false positives as non-quits. We will also evaluate rates of false positives across arms.

We will perform a sensitivity analysis without patients that were enrolled via the last chance to opt out (LCOO) system, as those patients might have different rates of baseline completion. We will also examine the characteristics of this group.

Subgroup analysis

Light smokers might be impacted differently by the interventions than heavier smokers. In order to test this, we will perform a subgroup analysis of our primary outcome with patients who smoke 1-4 cigarettes per day. We also plan to do a subgroup analysis among the participants with smart phones with data plan vs. not if the interaction analyses above are significant.

Restricted sample analysis

- We anticipate that some trial participants at UPHS will receive pharmacotherapy even though they are assigned to control arm. To tackle this discrepancy, we plan to conduct primary outcome analysis by only including arms 2-4 for UPHS. The structure of the analysis will be similar to equation (1).
- LGH Wellness clinic is offering NRTs (grant funded) to the participants regardless of the arm assignment. To explore the effect of this, we will conduct primary outcome analysis using restricted analysis that excludes LGH.
- On 1/27/23, KP added wellness coaching messaging in order to raise participants' access to NRT. We will explore an analysis of the primary outcome restricting KP patients to only after the messaging has changed

Secondary Outcome Analysis

To analyze the secondary endpoints listed in table 3 we will use suitable modelling methods such as linear models, count models, Quantile regression, etc. depending on the structure of the outcome. We will use logistic regression model to analyze and report binary outcomes.

Fitness of the model will be checked for all models. If model diagnostics indicate poor model fit, an appropriate transformation of the response and/ covariates will be done. Also, we will consider alternative modeling approaches for analysis. When disseminating results from these alternative models, they would be clearly described as deviating from the planned approach.

Table 3: Secondary outcomes

	Tools of measurement	Type/ Definition	Timeframe	Test For Trend?
Smoking related outcomes				
2 weeks quit status	Biochemical test	Binary, yes/ no		No
3 months quit status	Biochemical test	Binary, yes/ no		No
12 months relapse	Biochemical test	Binary, yes/ no		No
2 weeks self-report cigarette use	Self-report	Binary, yes/no		No
3 months self-report cigarette use	Self-report	Binary, yes/no		No
6 months self-report cigarette use	Self-report	Binary, yes/no		No
12 months self-report cigarette use	Self-report	Binary, yes/no		No
2 weeks self-report other tobacco use	Self-report	Binary, yes/no		No
3 months self-report other tobacco use	Self-report	Binary, yes/no		No
6 months self-report other tobacco use	Self-report	Binary, yes/no		No
12 months self-report other tobacco use	Self-report	Binary, yes/no		No
Patient reported outcomes				
Motivation to quit	Stages of Change questionnaire		0, 6, 12 months	No
Self-efficacy to quit smoking	Situational measure of self-efficacy	Each question answer ranges between 1 = *not at all sure*, 4 = *completely sure*. Total Score range: 10-40.	0, 6, 12 months	Yes
Perceived barriers to cessation	Challenges to Stopping Smoking Scale (CSS-21)	The score consists of two subscales: 9-item‘intrinsic factors’ and 12-item ‘extrinsic factors’. The total scores of the ‘intrinsic scale’ ranged from 9 to 36 and the ‘extrinsic scale’ from 12 to 43[4]. A higher score indicates greater challenges. Missing values were	0, 6, 12 months	Yes

		replaced with the mean of answered items for participants with ≤20% items missing. Participants with >20% missing data were excluded from the analysis.		
Temporal Discounting	5-Trial Adjusting Delay Task	Discounting rate (k) obtained from result of final (fifth) question in task. k scores match results from Table 1 in Koffarnus & Bickel, 2014[5]. The higher k, the more the present is valued over future gain. K is calculated from the equation V = A/(1+kD), where V is the present value, A is the delayed value, and D is the delay in days.	0, 6, 12 months	Yes
Health-related quality of life (HRQOL)	EuroQuol Group's EQ-5D	Score ranging from -1 to 1 calculated by SAS package provided by the author. We will also calculate Misery score from the survey responses which is simply sum of the scores ranging from 5 to 25.	0, 6, 12 months	Yes

Other Secondary Analyses

Due to a Chantix recall, there might have been changing rates of general pharmacotherapy use over time. We will examine responses of patients indicating whether or not they have used pharmacotherapy to see if these rates have declined. We will look at this rate by quarter year of enrollment.

3.3 Handling of Missing Data

Incidence of missing data, particularly the primary endpoint, is expected to be low because recommended practice in smoking cessation RCTs is to consider participants who do not

submit samples for biochemical confirmation as ongoing smokers[6]. We also used this standard in our 3 prior trials of smoking cessation interventions[7-9]. However, a small number of patients may have missing data on the primary outcome due to mortality or requests to not be contacted further for this study.

For secondary outcomes, we will fully describe all data that are missing at each time point. If possible, analyses of secondary outcomes will use multiple imputation methods that assume data are missing at random by including all baseline characteristics, treatment assignment, and reasons for missingness in the imputation model. This approach allows use of baseline and available follow-up data (e.g., at baseline and 6 months) to inform the generation of values for subsequent missing data points (e.g., at 12 months), and properly adjusts for the uncertainty in the resulting imputed values[10,11]. However, because the accuracy of these methods of imputation declines with higher rates of missingness, any secondary outcomes for which data are missing for >20% of the eligible sample for that analysis will be described in the results rather than analyzed inferentially (i.e., with imputation and statistical significance testing).

We also plan to conduct sensitivity analyses using pattern-mixtures models to explore and potentially adjust for patterns of missingness.

4. INTERIM CHECKS

We have no plan to do any interim statistical tests of the primary outcome, unless these are specifically requested by the DSMB. Alisa Stephens, PhD (Co-I), Assistant Professor of Biostatistics at Penn, will serve as a blinded statistician overseeing interim data presentation (descriptive statistics) and communicating these to the DSMB.

5. SAMPLE SIZE CONSIDERATIONS

5.1 Sample Size Considerations for Primary and Secondary Aims

Based on our previous trials, and considering the population to be studied in the present trial, we anticipate a six-month prolonged abstinence rate of 2.5% with basic usual care. Although somewhat higher rates of cessation were observed in usual care during the National Lung Screening Trial, that trial used self-report to determine smoking status and enrolled patients who were interested in improving their health, rather than the non-selected patients to be

enrolled in this pragmatic trial. We wish to detect absolute differences of at least 5 percentage points in abstinence rates between any of the three intervention arms and basic usual care or between any two of the three intervention arms (six total contrasts). This difference is clinically important given the dramatic health benefits of stopping smoking and the low rates of success of available interventions, and is identical to that used to determine power in our recent pragmatic RCTs.

Contrast	Health systems contributing data to the analysis
AAR vs. AAR + pharmacotherapy	3 health systems, excluding Kaiser
AAR vs. AAR + pharmacotherapy + incentives	3 health systems, excluding Kaiser
AAR vs. AAR + pharmacotherapy + incentives + EFT	3 health systems, excluding Kaiser
AAR + pharmacotherapy vs. AAR + pharmacotherapy + incentives	All 4 health systems
AAR + pharmacotherapy vs. AAR + pharmacotherapy + incentives + EFT	All 4 health systems
AAR + pharmacotherapy + incentive vs. AAR + pharmacotherapy + incentives + EFT	All 4 health systems

We used statistical simulations to understand the power under various settings with 1,000 simulations at each setting. All simulations were performed in R version 3.5.0[12].

5.2 Sample Size and Power Considerations for Treatment Heterogeneity Analyses

Achieving the target sample size will provide >80% power to detect differences of at least 10 percentage points for statistical interactions between the comparative effects of two interventions and dichotomous patient characteristics (e.g., black vs. white) in which the less commonly observed characteristic is represented among at least 25% of participants. This calculation of power to detect statistical interactions uses an alpha level of 0.10, as in our prior study of modifiers of the effects of smoking cessation interventions.

5.3 Protocol Deviations

All important protocol deviations will be summarized and reviewed by treatment arm and site during the active enrollment period. The protocol deviations monitored will include patients who received incorrect treatment at any time, patients violating inclusion/exclusion criteria, patients assigned a randomization code according to incorrect baseline stratification factors, and non-compliance with protocol.

References

1. Proschan M. A. (1999). A multiple comparison procedure for three- and four-armed controlled clinical trials. Statistics in medicine, 18(7), 787–798. https://doi.org/10.1002/(sici)1097-0258(19990415)18:7<787::aid-sim77>3.0.co;2-m
2. Cai, B., Small, D. S., & Have, T. R. T. (2011). Two-stage instrumental variable methods for estimating the causal odds ratio: Analysis of bias. *Statistics in Medicine*, *30*(15), 1809–1824. https://doi.org/10.1002/sim.4241
3. Terza, J. V. (2017). Two-stage residual inclusion estimation: A practitioners guide to Stata implementation. *The Stata Journal*, *17*(4), 916-938.
4. Thomas, D., Mackinnon, A. J., Bonevski, B., Abramson, M. J., Taylor, S., Poole, S. G., Weeks, G. R., Dooley, M. J., & George, J. (2016). Development and validation of a 21-item challenges to stopping smoking (CSS-21) scale. BMJ open, 6(3), e011265. https://doi.org/10.1136/bmjopen-2016-011265
5. Koffarnus, M. N., & Bickel, W. K. (2014). A 5-trial adjusting delay discounting task: accurate discount rates in less than one minute. Experimental and clinical psychopharmacology, 22(3), 222–228. https://doi.org/10.1037/a0035973
6. West, R., Hajek, P., Stead, L., & Stapleton, J. (2005). Outcome criteria in smoking cessation trials: proposal for a common standard. *Addiction (Abingdon, England)*, *100*(3), 299–303. https://doi.org/10.1111/j.1360-0443.2004.00995.x
7. Volpp, K. G., Troxel, A. B., Pauly, M. V., Glick, H. A., Puig, A., Asch, D. A., Galvin, R., Zhu, J., Wan, F., DeGuzman, J., Corbett, E., Weiner, J., & Audrain-McGovern, J. (2009). A

randomized, controlled trial of financial incentives for smoking cessation. The New England journal of medicine, 360(7), 699–709. https://doi.org/10.1056/NEJMsa0806819

8. Halpern, S. D., French, B., Small, D. S., Saulsgiver, K., Harhay, M. O., Audrain-McGovern, J., Loewenstein, G., Brennan, T. A., Asch, D. A., & Volpp, K. G. (2015). Randomized trial of four financial-incentive programs for smoking cessation. *The New England journal of medicine*, *372*(22), 2108–2117. https://doi.org/10.1056/NEJMoa1414293
9. Halpern, S. D., Harhay, M. O., Saulsgiver, K., Brophy, C., Troxel, A. B., & Volpp, K. G. (2018). A Pragmatic Trial of E-Cigarettes, Incentives, and Drugs for Smoking Cessation. *The New England journal of medicine*, *378*(24), 2302–2310. https://doi.org/10.1056/NEJMsa1715757
10. Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Wiley.
11. Buuren, S. van, & Groothuis-Oudshoorn, K. (2010). *MICE: Multivariate Imputation by Chained Equations in R* [Article]. Journal of Statistical Software; University of California, Los Angeles. http://localhost/handle/1874/44635
12. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Statistical Analysis Plan

Study Name	*Comparing Smoking Cessation Interventions Among Underserved Patients Referred for Lung Cancer Screening*
Version Number	1
Date	3 March 2020

Comparing Smoking Cessation Interventions among Underserved Patients Referred for Lung Cancer Screening

SIGNATURE PAGE

Approved by:

_____
Scott D. Halpern, MD, PhD
Principal Investigator
PAIR Center

_____
xxxxxxxxxx
Co-Investigator
xxxxxxxxxx

_____
xxxxxxxxxx
Co- Investigator
xxxxxxxxxxxx

_____
xxxxxxxx
Co- Investigator
xxxxxxxxx

Revision history

Revision	Date	Section/Page	Changes Made -- Reasons for the Change

TABLE OF CONTENTS PAGE

LIST OF ABBREVIATIONS

The following abbreviations and special terms are used in this Statistical Analysis Plan (SAP).

Abbreviation or special term	Explanation
AAR	Ask-Advise-Refer
Co-I	Co-Investigator
DSMB	Data and Safety Monitoring Board
EFT	Episodic Future Thinking
GLM	Generalized Linear Model
ITT	Intension To Treat
IV	Instrumental Variable
LDCT	Low Dose Computed Tomography
Penn	University of Pennsylvania
PI	Principal Investigator
SAP	Statistical Analysis Plan
SSN	Social Security Number

INTRODUCTION

This statistical analysis plan (SAP) is a comprehensive and detailed description of the strategy, rationale and statistical techniques that will be used in the smoking cessation study.

1. STUDY DETAILS

1.1 Study Design

We will conduct a 4-arm randomized trial comparing 4 interventions-

i. Basic usual care: Ask-Advise-Refer (AAR) approach,

ii. Enhanced usual care: usual care plus free nicotine replacement therapy and FDA-approved pharmacotherapies,

iii. Enhanced usual care plus financial incentives to stop smoking,

iv. Enhanced usual care plus financial incentives plus a mobile health application that motivates patients to think about their future health by promoting episodic future thinking (EFT) to promote sustained, biochemically confirmed smoking abstinence for 6 months among smokers in underserved demographic groups.

The participants to be enrolled will be current smokers who are black, Hispanic, and/or have low socioeconomic status (defined as household income <200% of the federal poverty line or a high school education or less) or rural residence who are referred for low-dose computed tomography (LDCT) screening at 4 large health systems. Patients with LDCT orders will be identified via the electronic health record and further screened for eligibility. Eligible patients will enroll and complete the study using the NIH-funded *Way to Health* portal. All of the following patient inclusion criteria must be met:

1) Age ≥ 18 years

2) Current smoker (≥ 5 cigarettes per day, not including e-cigarettes)

3) Has a LDCT scan ordered by his or her physician

4) Able to receive study invitation and screening, by virtue of showing up to a radiology location affiliated with 1 of the 4 health systems for the LDCT, or having a valid email address or telephone number on file with the health system

5) Underserved, defined as one or more of the following:

i. Black

ii. Hispanic

iii. Rural residence

iv. Low socioeconomic status, defined as one or both of: a. ≤ high-school education or less

b. household income < 200% of the federal poverty line

6) Access to a phone with text messaging or the internet

It is expected that LDCT orders will be placed by physicians in line with national screening guidelines, i.e., patients with a ≥ 30 pack-year smoking history and tobacco use within the past 15 years. The majority of enrolled patients are expected to be ≥ 55 years; however, patients < 55 years meeting all other eligibility criteria are eligible for this clinical trial.

1.2 Study Endpoints, Hypotheses, and Corresponding Outcomes

The primary outcome will be biochemically confirmed, sustained abstinence from smoking tobacco for 6 months following participants' selected quit dates. Secondary smoking-related outcomes include point-prevalent quit rates at 2 weeks and 3 months, and relapse rates at 12 months.

Secondary patient-reported outcomes include:
• Nicotine dependence
• Motivation to quit
• Self-efficacy related to cessation efforts
• Perceived barriers to cessation
• Health-related quality of life

Hypothesis 1a: Each of the three intervention arms will increase the rate of 6-month sustained smoking abstinence compared to Ask-Advise-Refer.
Hypothesis 1b: Financial incentives plus free access to pharmacotherapy (nicotine replacement therapy and FDA-approved pharmacologic cessation aids) will increase the rate of 6-month sustained smoking abstinence compared to free access to pharmacotherapy alone
Hypothesis 1c: Episodic future thinking (EFT) plus financial incentives plus free access to pharmacotherapy (nicotine replacement therapy and FDA-approved pharmacologic cessation aids) will increase the rate of 6-month sustained smoking abstinence compared to financial incentives plus free access to pharmacotherapy
Hypothesis 2: There will be significant treatment heterogeneity (hypothesis test of $\beta = 0$) in the foregoing contrasts by pre-specified patient-level covariates at $\alpha = 0.10$ level significance.
Hypothesis 3: Each of the three intervention arms will improve health related quality of life and will have improve rates of 12-month smoking abstinence

1.3 Randomization Scheme

We plan to enroll 3200 patients to four arms (usual care, enhanced usual care, enhanced usual care plus financial incentives and enhanced usual care plus financial incentives plus a mobile health application EFT). Participants will be randomized individually, stratified by health system, using random number generation within the *Way to Health* platform. At the Penn, Geisinger, and Henry Ford sites, 25% of enrolled participants will be assigned to each of the 4 arms. Participants enrolled at Kaiser sites will only be assigned to Arms 2-4, each with a 33% probability. This is because Kaiser provides free cessation aids to all its patients, and as such, there exists no basic usual care (control) arm. Based on the expected accrual at each of the health systems, we expect that overall there will be 470 patients in basic usual care (Arm 1) and 910 patients in each of intervention Arms 2-4.

2. ANALYSIS SETS

Intention-to-treat analyses will include all randomized patients regardless of their adherence to protocol or the actual treatment received. Patients will be analyzed using their assigned treatment arm. All events occurring within the specified time periods will be included in the analysis regardless of whether the patient adheres to the protocol. Information from patients who withdraw their consent to participate in the study will be included up to the time when they withdraw their consent. The primary analysis for all primary, and secondary endpoints will be performed as intent-to-treat analyses.

3. ANALYSIS METHODS

3.1 General Principles

Descriptive Statistics

Numerical variables will be summarized using standard summary statistics including the number of participants, mean, standard deviation, median, 10^{th} and 90^{th} percentile and range (i.e., minimum and maximum) as appropriate. For categorical data, proportions will be presented in a frequency table format.

Demographic and baseline clinical and biological variables will be summarized and compared by treatment arms. Continuous variables will be summarized using means, medians, ranges, standard deviations, and interquartile ranges. Frequency distributions will be presented for categorical and ordinal variables. A CONSORT diagram displaying the number of patients screened, eligible, and consented along with reasons for ineligibility will be provided. Graphical tools such as histograms, boxplots, and scatterplots will be created to assess quality of data and to display patterns over time. We expect that missing data will be minimal;

however, we will fully describe any missing data. Importantly, we will use statistical methods consistent with the intent-to-treat principle.

3.2 Analysis Methods

Primary Analysis

The analysis of the primary outcome will use a generalized linear model (GLM) with a logit link function to estimate the effect of treatment arms to usual care. All models will be adjusted for health system, and patient level covariates.

The following model will be used to test Hypothesis 1a.

$$g\big(E(Smoke_i)\big) = \beta_0 + \beta_{11} patient\ covariate_{1i} + \beta_{12} patient\ covariate_{2i} + \cdots + \beta_2 arm_{2i} + \beta_3 arm_{3i} + \beta_4 arm_{4i} + \cdots + \gamma_2 HS2_i + \cdots + \gamma_4 HS4_i + \gamma_5 HS5_i \quad (1)$$

where *g()* is the logit link function, *E()* represents the expected value, the subscript *i* corresponds to patient *i*, arm_{ki} equals 1 when patient *i* is randomized to the arm *k* of the three treatment arms, and $HS2_i$, …, $HS4_i$, $HS5_i$ equals 1 when patient *i* is from health system 2, …, 5, respectively (note that patients from health system 1 make up the reference group). We will test $\beta_k = 0, k = 2,3,4$ using a two-sided Wald test at the 0.05 significance level to decide whether to reject Hypothesis 1a in the final analysis. We will do adjustment for multiple tests to maintain family-wise type I error rate according to Proschan (1999).

Table 1: Patient level covariates

Variable	Definition
Age	Continuous
Gender	Male vs. not male
Race	Categorical (Black/not)
Ethnicity	Hispanic vs. Not Hispanic
Education	Categorical (5 levels)
Income (%FPL)	<200% FPL vs >= 200% FPL
Rurality (self report)	Rural vs urban
Temporal discounting	Continuous
Tobacco/ Nicotine dependence	Fagerstrom score (6 Questions, score range: 1-10) at baseline
LDCT screening results	Reassuring vs. Concerning
Number of prior LDCT screening tests	Integer
Presence of high burden of chronic comorbid illnesses	Charlson score
Time of accrual	Calendar month/ Season

To address Hypotheses 1b & 1c we will use the structure of the model (1) where we will include 1 dummy variable for treatment arm variable in each model.

Heterogeneous treatment effects (HTE) will be explored by testing statistical interactions between pre-selected patient characteristics (listed in Table 2) and each contrast between 4 interventions in logistic regression models. This will help guide future targeting of optimal interventions for individual patients. We will evaluate each interaction term separately in a model adjusted for all main effects and health system to estimate stratum-specific effects of each characteristic on the interventions' effectiveness. We will also report results from a fully adjusted model that retains all significant interaction terms. This analysis will inform our hypothesis 2. In analyzing the interaction terms we do not plan to incorporate correction for multiple testing.

Table 2: Potential effect modifiers

Variable	Definition
Age	Continuous
Gender	Male vs. not male
Race	Black vs not Black
Ethnicity	Hispanic vs. Not Hispanic
RUCA Rurality	Rural vs urban (RUCA2+ vs RUCA1)
Rurality (Self Report)	Rural vs urban
Income (%FPL)	<200% FPL vs >= 200% FPL
Smartphone with Internet Use	Phone w/ data plan vs not
Temporal discounting	Continuous
Tammemagi	Continuous
Tobacco/ Nicotine dependence	Fagerstorm score (6 Questions, score range: 1-10) at baseline
LDCT screening results	Reassuring vs. Concerning
Number of prior LDCT screening tests	0 scans vs 1 or more scans
Presence of high burden of chronic comorbid illnesses	Charlson score
CFPB Financial Well Being Score	Score range: 19-90
Insurance	Medicare (part D) vs Others

Mental Health Diagnoses (EHR)	Diagnoses vs not Diagnosis of the mental health condition will be defined using ICD 10 codes •F20-F29 Schizophrenia, schizotypal, delusional, and other non-mood psychotic disorders •F30-F39 Mood [affective] disorders •F40-F48 Anxiety, dissociative, stress-related, somatoform and other

Sensitivity analysis
We plan to analyze the primary outcome with ITT sample structured similar to (1) but will only include health system as fixed effects and no patient-level covariates. If we find any patient covariate that we prespecified for inclusion in the model and has missing values > 15% then the above unadjusted analysis would become our primary analysis.

Restricted sample analysis
We anticipate that some trial participants at UPHS will receive pharmacotherapy even though they are assigned to control arm. To tackle this discrepancy, we plan to conduct primary outcome analysis by discarding them from the ITT sample. The structure of the analysis will be similar to equation (1).

LGH Wellness clinic is offering NRTs (grant funded) to the participants regardless of the arm assignment. To explore the effect of this, we will conduct primary outcome analysis using restricted analysis that excludes LGH.

Instrumental Variable Analysis
This analysis examines the causal effect of the interventions on outcomes accounting for non-compliance behavior. This approach entails a two-stage residual inclusion model in which the randomization to the treatment components is used as an instrumental variable in causal odds ratio treatment effect analysis[1,2]. This analysis will also be adjusted for clinic and health system. Like the ITT analysis, these analyses use data on all randomized patients to estimate the effects of treatment regardless of group assignment, and after accounting for the possibility that intervention adherence rates may differ among the four arms. Thus, the estimated effect of the interventions is adjusted for the percentage of assigned participants who stick to their intervention arms, and the percentage who opt out from their assigned intervention.

Arms	Health systems contributing data to the analysis
AAR (Ask-Advise-Refer)	3 health systems, excluding Kaiser
AAR + pharmacotherapy	4 health systems

AAR+ pharmacotherapy + incentives	4 health systems
AAR + pharmacotherapy + incentives + EFT	4 health systems

Each treatment has some/ all components out of the 4 components. The four components are,1) AAR; 2) Pharmacotherapy; 3) Incentives; 4) EFT. But we cannot assume someone who complied with a higher ranked component would always comply with a lower ranked component, e.g., someone who complied with EFT would also comply with AAR, pharmacotherapy and incentives. This violates the monotonicity assumption if we treat randomization as single IV. To avoid such violation, we will treat randomization to each component as a separate IV. Without additional strong assumptions, one needs one instrument for each treatment to get identification, so with four arms (three instruments), we can identify the effect of at most three levels of treatment.

Randomization to 4 components- described above will be presented using 3 IV variables. Each IV identifies the average treatment effect (causal odds ratio) for its group, such that each IV is identifying the average treatment effect for a different subgroup.

Assuming homogeneity of treatment effects across classes and no interactions between the three treatments of medication, rewards and EFT, we can implement Two Stage Residual Inclusion (2SRI) method. In the first stage, we regress (Logistic regression) each treatment *D* on IVs *Z*1, *Z*2 and Z3 (as well as covariate **X**) separately. Then use the residuals from three first stage models in the second stage regression model as added covariates along with the observed value of each treatment. Bootstrap confidence intervals will be reported with the point estimates.

Each subject has eight potential outcomes. Only one of the potential outcomes can be observed, the outcome corresponding to the actual treatment the subject received. The outcome can be defined as follows:

$Y_i^{\{m,r,e\}}$, where m, r, e can be 0 or 1 to indicate taking medication or not, rewards 0 or 1 and EFT 0 or not.

Our approach assumes the exclusion restriction that a subject's potential outcome depends only on the treatment received. In other words, the treatment assignment only influences the potential outcome through the treatment received.

Secondary Outcome Analysis

To analyze the secondary endpoints listed in table 3 we will use suitable modelling methods such as linear models, count models, Quantile regression, etc. depending on the structure of the outcome. We will use Logistic regression model to analyze and report binary outcomes.

Table 3: Secondary outcomes

	Tools of measurement	Type/ Definition	Timeframe	Test For Trend?
Smoking related outcomes				
2 weeks quit status	Biochemical test	Binary, yes/ no		No
3 months quit status	Biochemical test	Binary, yes/ no		No
6 months quit status	Biochemical test	Binary, yes/ no		No
12 months relapse	Biochemical test	Binary, yes/ no		No
2 weeks self-report cigarette use	Self-report	Binary, yes/no		No
3 months self-report cigarette use	Self-report	Binary, yes/no		No
6 months self-report cigarette use	Self-report	Binary, yes/no		No
12 months self-report cigarette use	Self-report	Binary, yes/no		No
2 weeks self-report other tobacco use	Self-report	Binary, yes/no		No
3 months self-report other tobacco use	Self-report	Binary, yes/no		No
6 months self-report other tobacco use	Self-report	Binary, yes/no		No
12 months self-report other tobacco use	Self-report	Binary, yes/no		No
Patient reported outcomes				
Motivation to quit	Stages of Change questionnaire		0, 6, 12 months	No
Self-efficacy to quit smoking	Situational measure of self-efficacy	Each question answer ranges between 1 = *not at all sure*, 4 = *completely sure*. Total Score range: 10-40.	0, 6, 12 months	Yes
Perceived barriers to cessation	Challenges to Stopping Smoking Scale (CSS-21)	The score consists of two subscales: 9-item‘intrinsic factors’ and 12-item ‘extrinsic factors’. The total scores of the ‘intrinsic scale’ ranged from 9 to 36 and the ‘extrinsic scale’ from 12 to 43. A higher score indicates greater challenges. Missing values were replaced with the mean of answered items for participants with ≤20% items missing. Participants	0, 6, 12 months	Yes

		with >20% missing data were excluded from the analysis.		
Temporal Discounting	5-Trial Adjusting Delay Task	Discounting rate (k) obtained from result of final (fifth) question in task. k scores match results from Table 1 in Koffarnus & Bickel, 2014. The higher k, the more the present is valued over future gain. K is calculated from the equation V = A/(1+kD), where V is the present value, A is the delayed value, and D is the delay in days.	0, 6, 12 months	Yes
Health-related quality of life (HRQOL)	EuroQuol Group's EQ-5D	Score ranging from -1 to 1 calculated by SAS package provided by the author. We will also calculate Misery score from the survey responses which is simply sum of the scores ranging from 5 to 25.	0, 6, 12 months	Yes

Fitness of the model will be checked for all models. If model diagnostics indicate poor model fit, an appropriate transformation of the response and/ covariates will be done. Also, we will consider alternative modeling approaches for analysis. When disseminating results from these alternative models, they would be clearly described as deviating from the planned approach.

3.3 Handling of Missing Data

Incidence of missing data, particularly the primary endpoint, is expected to be low. Missingness in primary outcome can occur in following situations:

- No response to short survey
- Self-reported quit but no lab sample submitted in time

We will fully describe all data that are missing at each time point. Our secondary analyses will use multiple imputation methods that assume data are missing at random by including all baseline characteristics, treatment assignment, and reasons for missingness in the imputation

model. This approach allows use of baseline and available follow-up data (e.g., at baseline and 6 months) to inform the generation of values for subsequent missing data points (e.g., at 12 months), and properly adjusts for the uncertainty in the resulting imputed values[3,4].

We also plan to conduct sensitivity analyses using pattern-mixtures models to explore and potentially adjust for patterns of missingness.

4. INTERIM CHECKS

We have no plan to do any interim statistical tests of the primary outcome, unless these are specifically requested by the DSMB. Alisa Stephens, PhD (Co-I), Assistant Professor of Biostatistics at Penn, will serve as a blinded statistician overseeing interim data presentation (descriptive statistics) and communicating these to the DSMB.

5. SAMPLE SIZE CONSIDERATIONS

5.1 Sample Size Considerations for Primary and Secondary Aims

Based on our previous trials, and considering the population to be studied in the present trial, we anticipate a six-month prolonged abstinence rate of 2.5% with basic usual care. Although somewhat higher rates of cessation were observed in usual care during the National Lung Screening Trial, that trial used self-report to determine smoking status and enrolled patients who were interested in improving their health, rather than the non-selected patients to be enrolled in this pragmatic trial. We wish to detect absolute differences of at least 5 percentage points in abstinence rates between any of the three intervention arms and basic usual care or between any two of the three intervention arms (six total contrasts). This difference is clinically important given the dramatic health benefits of stopping smoking and the low rates of success of available interventions, and is identical to that used to determine power in our recent pragmatic RCTs.

Contrast	Health systems contributing data to the analysis
AAR vs. AAR + pharmacotherapy	3 health systems, excluding Kaiser
AAR vs. AAR + pharmacotherapy + incentives	3 health systems, excluding Kaiser
AAR vs. AAR + pharmacotherapy + incentives + EFT	3 health systems, excluding Kaiser
AAR + pharmacotherapy vs. AAR + pharmacotherapy + incentives	All 4 health systems
AAR + pharmacotherapy vs. AAR + pharmacotherapy + incentives + EFT	All 4 health systems
AAR + pharmacotherapy + incentive vs. AAR +	All 4 health systems

pharmacotherapy + incentives + EFT	

We used statistical simulations to understand the power under various settings with 1,000 simulations at each setting. All simulations were performed in R version 3.5.0[9].

5.2 Sample Size and Power Considerations for Treatment Heterogeneity Analyses

Achieving the target sample size will provide >80% power to detect differences of at least 10 percentage points for statistical interactions between the comparative effects of two interventions and dichotomous patient characteristics (e.g., black vs. white) in which the less commonly observed characteristic is represented among at least 25% of participants. This calculation of power to detect statistical interactions uses an alpha level of 0.10, as in our prior study of modifiers of the effects of smoking cessation interventions.

5.3 Protocol Deviations

All important protocol deviations will be summarized and reviewed by treatment arm and site during the active enrollment period. The protocol deviations monitored will include patients who received incorrect treatment at any time, patients violating inclusion/exclusion criteria, patients assigned a randomization code according to incorrect baseline stratification factors, and non-compliance with protocol.

References

1. Cai, B., Small, D. S., & Have, T. R. T. (2011). Two-stage instrumental variable methods for estimating the causal odds ratio: Analysis of bias. *Statistics in Medicine*, *30*(15), 1809–1824. https://doi.org/10.1002/sim.4241
2. Terza, J. V. (2017). Two-stage residual inclusion estimation: A practitioners guide to Stata implementation. *The Stata Journal*, *17*(4), 916-938.
3. Buuren, S. van, & Groothuis-Oudshoorn, K. (2010). *MICE: Multivariate Imputation by Chained Equations in R* [Article]. Journal of Statistical Software; University of California, Los Angeles. http://localhost/handle/1874/44635
4. Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Wiley.

STATISTICAL ANALYSIS PLAN AMENDMENTS LOG

Protocol Number: Penn IRB 833713

Protocol Title: Comparing Smoking Cessation Interventions among Underserved Patients Referred for Lung Cancer Screening: A Randomized Clinical Trial

Principal Investigator: Scott D. Halpern

Modification Date	Summary of Changes	Change Justification
06/30/2021	We now plan to enroll all underserved patients who have a lung cancer screening order, meet our criteria for being underserved, and self-report smoking at least 1 cigarette per day on average for the past month. We will perform a subgroup analysis for light smokers.	We believe this change will better align our clinical trial with methods used in other smoking cessation trials in which the criteria were “daily smoking” without requiring a minimum number of cigarettes. This proposed change comes after discussion with our Stakeholder Advisory Committee of health system leaders and smoking cessation experts; the stakeholders in attendance of a June 2021 meeting unanimously supported this change.
06/30/2021	Nicotine dependence was removed as a secondary outcome.	The Fagerstrom (measuring nicotine dependence) is only collected at baseline; therefore, nicotine dependence cannot be used as a secondary outcome.
02/7/2022	Added a sensitivity analysis coding levels of cotinine between 21 and 50 ng/ml as non-quits. Patients will be considered quits in primary analyses between this level.	Numbers between that level likely show up as false positives on the tests.
04/19/2022	Carboxy test levels for patient quit status changed from less than 4 to less than or equal to 4.	Carboxy tests return from the lab as whole numbers. Our original threshold was less than 4. We started to see a number of test results that read 4.0 for carboxy, despite participant attestation that they quit tobacco. Investigators suspect that their results were actually <4.0 (meeting the original carboxy threshold), but due to rounding on the part of labs, were reported as 4.0 (thus failing to meet the original threshold, which would kick them off the quit path).
5/31/2022	Added tests to look at trends in Chantix use, self-reported vs prescription.	Nationwide Chantix recall.
10/31/2022	Added sensitivity analysis analyzing patients randomized through the LCOO (Last chance to opt out) program	LCOO baseline completion rate is far below the rates of other participants, so these patients might be fundamentally different than other enrolling patients
3/6/2023	Secondary analysis added restricting the KP sample to only after the change was made regarding Wellness Coaching processes.	Patients at KP were accessing NRT less than expected due to processes at the hospital system. Wellness Coaching was added in order to get more eligible participants access to NRT.

4/30/2024	Fagerstrom score, as both a patient-level covariate and effect modifier, was changed to only cigarettes per day.	Fagerstrom had high missingness due to being asked at the baseline questionnaire, so instead we are using cigarettes per day, as we had more data from the eligibility questionnaire, and it handles the same construct.
4/30/2024	Added insurance as a covariate (Medicare, Medicaid, Commercial) and dropped it as an effect modifier.	Insurance was changed from an effect modifier to a covariate to better capture variance. It was also changed from a binary variable (Medicare Part D vs others) to a categorical one (Medicare, Medicaid, Commercial).
4/30/2024	Time of accrual changed from calendar month/season to pandemic era (launch through 9/18/22 = 1, after = 0).	Wanted to adjust for the difference in temporality based on pandemic era versus not.
6/11/2024	Temporal discounting was removed as a patient-level covariate and effect modifier.	The variable was only collected at the baseline survey, and due to not all patients having the baseline survey, we decided to drop it so that we would not lose too many patients from our models.
6/11/2024	Charlson score changed to Elixhauser.	Charlson score is grouped into categories instead of a continuous variable, and there were some concerns over the categorizations. Elixhauser is continuous and has been used in other trials, so we decided this would be a better variable for comorbidities.
6/11/2024	Clarified the Rurality variable definition.	Rural vs urban was changed to Rural vs non-rural. Analysis used self-report when available and pulled RUCA only when self-report was not available.
6/9/2025	Removed instrumental variable analysis	Upon completion of data collection and beginning our analysis, it was realized by the team that the assumptions typically required for the CATE analysis, which are often employed in simpler settings with fewer, non-overlapping arms, were not plausible. In the original SAP, the team noted that due to the design of cumulatively adding the previous arm's intervention to the next arm, we could not treat randomization as a single IV without violating the monotonicity assumption. Thus, we stated we would treat randomization to each component as a separate IV. During data processing, we then realized that the exclusion restriction, which states that the effect of randomization to each component on outcomes goes only through the component of interest, is not plausible given the co-occurring nature of the interventions based on the cumulative design of the study arms. Therefore, this analysis would not be causally interpretable as intended.