# EFFECTIVENESS STATISTICAL ANALYSIS PLAN (SAP) – REVISED 04/20/2023

## Summary of revision

This revision of the analytic plan was motivated by several external disruptions during the conduct of the study, and by a brief examination of treatment-blinded outcome data. The main changes to the original Statistical Analysis Plan (SAP) involve the method used to model the primary PDC outcome, detailed specification of adjustment covariates, and adjustment for differing lengths of follow-up due to death, changes in prescriptions, or external disruptions. No changes to primary or secondary outcomes are proposed. The changes of substance occur only in Sections 3.1.a-e.

## 1. Study summary

The primary outcome will be medication adherence defined as the proportion of days covered (PDC) in the 12 months post randomization. This will be a composite measure across a patient's multiple medications and will be assessed using pharmacy refill data. Secondary outcomes will include intermediate clinical measures (e.g., BP control), CV clinical events (e.g., hospitalization for myocardial infarction) and procedures (e.g. PCI), medication-associated clinical events (e.g., syncope in patient on anti-hypertensive therapy), healthcare utilization, and costs. These outcomes will also be assessed at 12 months, with cost and utilization also assessed for specific components as described below.

**Study description:** This individually randomized three-center trial will compare three behavioral nudges based on text and chatbot messaging with usual care to improve adherence to cardiovascular medications.

**Setting:** The study will be conducted in 17 primary care clinics within 3 health care systems (University of Colorado Health, VA Eastern Colorado Health Care System, and Denver Health Medical Center).

**Design:** The study will be an individually randomized controlled study with four treatment arms (details of interventions below). Patients become eligible for the study when identified through pharmacy refill data to have a 7-day gap in any prescribed CV medication refills. Eligible patients will be randomized to one of four arms. Randomization will be stratified within each of the clinics, and further stratified within clinics by patients with 1-2 vs 3 or more active CV medication classes at baseline, and using blocks of 4 patients to ensure balance within clinics over time. Thus, within each clinic and number of other medication stratum, each set of 4 consecutively enrolled subjects will be randomized to the four study arms. Treatments will be initiated immediately upon randomization, in response to the 7-day gap.

**Intervention:** The four treatment arms will be 1) usual care (control); 2) generic text message reminder; 3) text message behavioral nudges only; or 4) text message behavioral nudge plus a pre-programmed AI interactive chat bot designed to identify and resolve barriers to medication refill and adherence.

Delivery of treatments will be as follows:
For patients randomized to one of the three active intervention arms:

- If a patient gaps on a medication at baseline, they receive a text to refill that medication.
- If a patient gaps on multiple medications at baseline, they receive only one text about all of the medications they gapped on.

- If a patient later gaps on any study medication, they will receive a reminder, regardless of whether they had gapped on or been prescribed the medication at baseline.
- If a patient is removed from a medication, no further texts are sent for that medication.

Patients randomized to the control arm receive no texts or other reminders.

## 2. Study outcomes

**Primary and secondary outcomes:** The study outcomes have been selected based on input directly elicited from patients and other stakeholders. The primary outcome is adherence to CV medications as measured by 12-month composite proportion of days covered (PDC). Secondary outcomes include alternate measures of CV medication adherence, clinical events (e.g., event times for stroke, MI, mortality), utilization of care (e.g., hospitalizations or clinic visits for CV-related reasons), and costs of healthcare utilization. Subjects will be followed for at least 12 months following randomization to assess these secondary outcomes. Subjects who have more than one year of follow-up (up to 3 years depending on when they are enrolled during years 2-3) will continue to be followed for secondary outcomes.

**Ascertainment of Outcomes:** Data for the primary outcome PDC will be obtained using pharmacy records from each of the healthcare systems during the 365-day follow-up. The medication refill data needed to assess PDC is routinely collected in the pharmacy databases of each of the participating sites.

During the UG3 phase, the data management workgroup developed definitions and specifications for the secondary outcomes that will be captured from the EHR at each of the three healthcare systems. International Classification of Diseases, Ninth and Tenth Revision (ICD 9 and 10), CPT and DRG codes identifying CV clinical events, CV procedures, and adverse medication associated clinical events have been compiled to ensure accurate identification these outcomes. Outcome measures for BP and LDL have also been standardized using NIH Collaboratory definitions for these standardized EHR data elements (**Appendix V)**. The data core successfully used EHR and administrative data to assess for each of these outcomes during the UG3 pilot year.

Several sources of cost data will be used. To capture costs of development, implementation and maintenance of the intervention, we will develop instruments (e.g. time logs) and procedures to prospectively capture resource use associated with the intervention including what was done, who did it, how long it took, and what nonhuman resources were required. This is further discussed in greater detail in the health economics plan. In brief, intervention costs will be the cost of implementing the intervention excluding research and development costs. Costs associated with healthcare utilization will also be estimated using a resource-based method previously developed to assign costs to encounter data. Inpatient utilization will be measured using diagnostic-related groups (DRGs), outpatient utilization using relative value units (RVUs), and cost of pharmacy utilization using the midpoint between the Federal Supply Schedule (FSS) and the National Average Drug Acquisition Cost (NADAC). Inpatient costs will be estimated by applying national payment weights to DRGs, outpatient costs by applying a national conversion factor to RVUs, and pharmacy costs as the median between the FSS and NADAC. Please refer to health economic plan analysis for further details.

## 3. Statistical analysis plan

Analyses will be based on the intent to treat principle, using all patients who were randomized.

### 3.1 Analysis of primary composite PDC outcome

#### *3.1.a Summary of modeling approach*

We propose to calculate a composite PDC ratio of days covered for each subject and analyze it using a standard longitudinal model for monthly PDC values, with a linear mean model, and with robust standard error estimation and associated inference based on weighted Generalized Estimating Equations (GEE), e.g. Robins et al., 1995; Preisser et al., 2002. In more detail and with rationale:

*Outcome distribution*: Since data are in the form of a discrete proportion, our original proposal was based on Bernoulli/binomial models. However, treatment-blinded examination of a sample of outcome data showed that the binomial assumption will not apply due to spikes caused by common prescription lengths (e.g. 30 and 90 days), and strong overdispersion or equivalently lack of independence of daily outcome data over time. We considered alternative approaches such as Poisson with offset for number of observation days but similar challenges of spikes, overdispersion and lack of independence remain. Ultimately, with the large number of subjects available for analysis (~10,000) and the bounded distribution of PDC (between 0 and 1), central limit theorem arguments make standard linear modeling valid in this situation (e.g. Lumley et al., 2002), and robust inference methods described below provide additional assurance of valid inference even with heteroscedastic variance in the outcomes.

*Missing and truncated follow-up:* Missing data and truncated follow-up for PDC have occurred for several reasons. Patients with inpatient stays have medications supplied by the hospital and thus don't deplete their own supplies, so such days are omitted from both numerator and denominator of PDC. Other situations resulted in early termination of PDC calculation. Prescribed medications are sometimes terminated by providers, or patients may die during follow-up. Additionally, initial study procedures did not allow for collection of outcome data on patients who opted out, and while this was fixed early in the study a small percentage of early patients have shorter PDC observation. Finally, treatment delivery was disrupted by two external situations specific to health care systems, one involving a new VA messaging system that would confound delivery of Nudge treatment messages, and one involving gap identification for patients at UCHealth. In both cases, when the issue occurred our enrollment targets had been surpassed at each system. Delivery of Nudge treatments was stopped and collection of daily PDC outcome data was truncated. The disruptions due to these two exogenous events resulted in data that can be assumed missing completely at random (MCAR), however this is likely not true for the other cases of early termination of PDC observation. Further examination of the treatment-blinded outcome data showed an increasing trend over time from time zero in PDC during periods shortly after randomization, due to the initial gaps. This together with early termination could result in biased estimation.

*Longitudinal models and GEE:* To account for the situations of shortened PDC observation and possible resulting biases we will use a longitudinal model with PDC calculated for monthly intervals. We will use GEE with identity link and independence with unequal variances for the covariance structure. This approach provides robust and likely conservative (Hernán et al., 2002) estimation using empirical (sandwich) variance estimates, but is valid only for MCAR data. However, use of weighting can extend GEE to data missing at random (MAR). In this approach, observation-specific weights are equal to the inverse probability that the longitudinal value was observed. Probability of observation will be estimated by a logistic regression model for whether the value was observed (y/n), with covariates listed in section 3.1.c below. Weights greater than the 95$^{th}$ percentile of weights will be set to the 95$^{th}$ percentile weight. These methods can be implemented for example in SAS PROC GEE (Lin & Rodriguez, 2015) or R package `wgeesel` (Xu et al., 2018). We will use multiple imputation to impute missing covariate data. We will carry out the recommended sensitivity analyses to the MAR assumption using methods based on pattern mixture models and imputation, by assuming a range of

perturbations of imputed values and assessing differences in model conclusions (e.g. White et al., 2011; Fiero et al., 2017).

*Treatment comparisons:* Discussions with clinicians have indicated expressing results on a linear scale (PDC differences), as opposed to odds ratios or risk ratios, will be most appropriate for interpretation. This occurs naturally with the linear specification in 3.1.a and identity link in GEE. Estimation and inference will be carried out using the parameter estimates from the robust GEE estimation to construct expected PDC treatment differences. The estimand comparisons will be 12-month PDC, calculated by summing the 12 monthly longitudinal parameters, which incorporate the adjustments for early termination of some subjects using the weighted GEE approach. Primary hypotheses involve pairwise comparisons between each of the four study arms, and will be conducted using a multistage gatekeeper approach to account for the multiple treatment comparisons. In stage 1 of this approach, each of the three active intervention arms is compared with the control arm using significance level 0.05/3. In stage 2, if any of the three stage 1 tests is significant, the three pairwise comparisons among active intervention arms are tested with the Holm method using significance level (R/3)*(0.05/3), where R is the number of stage 1 tests that were significant (Dmetrienko et al., 2008).

### *3.1.b Defining composite PDC for multiple medications*

There are 13 CV medication classes considered in this study, and the primary outcome is composite PDC across medications a patient is prescribed. There are several considerations in defining composite PDC, including which of a patient's medications to include in the composite PDC calculation, and calculation of composite PDC with differing lengths of follow-up for different medications.

*Medications to include in composite PDC:* We have considered three ways of selecting medications to include in the primary PDC outcome: PDC1) all medications on which a patient gapped at baseline, PDC2) all medications a patient ever gaps on, calculating PDC from the time of gap, and PDC3) all medications a patient was prescribed at baseline. Medications prescribed after the baseline gap and enrollment will not be included in any of these definitions of PDC, though they will be considered in secondary analyses. Each definition (1-3) has benefits and shortcomings. PDC1 is closest to a pure effect of the intervention, but exploratory analyses of pilot data have shown that many patients who gap in a medication and are randomized also gap in another medication shortly thereafter. Specifically, preliminary data indicate that ~16% of patients gap on an additional medication within 30 days of baseline and ~37% within 90 days. PDC1 will omit these other medications prescribed but not gapped on at baseline. PDC2 provides a similar estimate incorporating these later-gapped medications but excludes medications the patient may never gap on due to reminders for other medications. PDC3 provides an estimate of the overall effect of the interventions on medication patterns, incorporating indirect effects of reminders, but risks inflating PDC and diluting intervention effects by including medications the patient never gaps on. Our primary analyses will use PDC1, while PDC2 and PDC3 will be considered in secondary analyses.

*Composite PDC with differing lengths of follow-up:* In general, PDC is the sum of observation days covered divided by the sum of observation days. Due to variations from the planned 365 days of assessment, there are several ways of calculating composite PDC for multiple medications. Information on hospitalizations, and on medication changes and cancellations, will be available from the patient's electronic health record. Note that at-risk days is medication specific, as medications may change during the one-year assessment period. Our primary PDC calculation will be PDC-C1 = (sum of numerators of medication-specific PDCs) / (sum of denominators of medication-specific PDCs). This can also be viewed as a weighted sum of medication-specific PDCs with weights equal to the proportion of the 365 target days for which PDC was observed, and equally weights each day on each medication. PDC-C1 will be the outcome definition in the longitudinal models described in section 3.1.a. As a sensitivity

analysis we will also consider an alternate definition, PDC-C2 = average of medication-specific PDCs, which equally weights medications regardless of their length of observation. When all medications are observed for 365 days these definitions are equal. For the longitudinal analyses planned, these definitions will be applied within each longitudinal interval. We will carry out the recommended sensitivity analyses based on pattern mixture models, by assuming various values for difference in means between observed and unobserved data and assessing differences in model conclusions.

***3.1.c Adjustment/propensity covariates***

The linear weighted GEE model for PDC and the logistic model for propensity of observation will each contain terms for randomization stratification variables health care system (VA, Denver Health or CUHealth), number of CV medications prescribed at baseline (1-2 or 3+), calendar month of randomization; treatment arm, follow-up month from randomization in longitudinal analysis, and treatment by follow-up month interaction; patient demographics age, race, ethnicity, insurance status, and marital status; and comorbidity variables hypertension, hyperlipidemia, coronary artery disease, diabetes, atrial fibrillation, chronic heart failure, chronic kidney disease, cerebrovascular disease, prior myocardial infarction, prior revascularization, depression, PTSD, and substance abuse.

***3.1.d Alternate measures of medication adherence***

We will consider several alternate analyses for medication adherence.

*Alternate definitions of PDC:* Alternate definitions PDC2 and PDC3 and alternate composite PDC-C2 defined above will be analyzed using the same methods as for the primary definition PDC1-C1. These analyses will provide a better understanding of indirect effects of treatments, and of effects of treatments on all medications a patient is prescribed.

*Sensitivity to inactive medications:* Among medications gapping at baseline, we will repeat the primary analysis limiting our analyses to those medications that had at least one fill during the follow-up period. This will allow us to assess treatment effects among medications that are known to be active, removing medications that were cancelled but not indicated as such by the medical record.

*Medication-specific adherence:* We will calculate and analyze PDC for each medication individually, to allow analyses of heterogeneity of treatment effect (HTE) by drug class. Multiple comparisons will be adjusted for separately for each analysis using the gatekeeper approach described above.

*Medication gaps:* We will consider alternate ways of describing medication adherence behavior, including length of initial gap, time to subsequent gap, average length of gaps, and number of 7-day gaps.

***3.1.e Secondary examinations of medication adherence***

In secondary analyses we will examine several other questions related to medication adherence, including:

*Predictors of medication gaps:* We will examine predictors of patients having an initial gap of at least 7 days, using the initial pool of candidate patients, to identify types of patients or prescription characteristics for patients at highest risk for non-adherence. We will also examine predictors of subjects having a gap of at least 7 days while enrolled in the study.

*Mechanism of treatment effects:* We will carry out analyses examining mechanisms or mediators of treatment effect by considering direct responses to reminders, including time from first reminder to refill, and measures of patient engagement, e.g. number of patient text responses to reminders (intensive text and chatbot arms only).

*Heterogeneity of treatment effect*: We will carry out analyses examining HTE, specifically i) HTE between the three health care systems, and ii) HTE for patient characteristics or subgroups

of particular interest. HTE analyses will be carried out using interactions as recommended in Kent et al., 2010 [23] and will be considered exploratory.

***3.1.f Original power analysis***

We first estimate sample size required to achieve 80% power for the desired change in the primary outcome, then used current data to show we will be able to achieve this sample size.

*Power and sample size:* Required sample size was estimated for the primary outcome PDC during the 12 months following randomization. Preliminary data from the VA were used for these estimates. We made the following assumptions: a) Significance using two-sided level 0.05 tests, b) Power at least 80%; c) Difference between treatments in PDC of 10 percentage points; d) Bonferroni adjustment for the 6 pairwise comparisons among the 4 study arms, resulting in adjusted level 0.05/6 (a conservative alternative of the sequential gatekeeper approach to be used in the final analysis as described above, to simplify the calculation of power); e) Analysis stratified by health care system; and f) Within-system and within-treatment residual standard deviation of 12 month PDC equal to 0.22 (mean 0.732), obtained by analysis of 2,859 veterans during the period 01/01/2017 – 12/31/2017 who were prescribed the medications of interest in this study. The large SD is due to a highly left-skewed PDC distribution. (We base sample size estimation on t-tests rather than the proposed binomial models since we do not have estimates of quantities needed to carry out power simulations. We may be able to obtain data for the PDC outcome for a set of baseline patients, in which case we will consider a refined power analysis.) Using these assumptions and estimates, and comparing any two treatments using a linear model with the above residual standard deviation of PDC, we estimate using sample size functions in R that we will need N=119 subjects per treatment arm, total across the three health care systems, for a total of 476 subjects to be randomized across the three health care systems.

*Available sample sizes:* We obtained data from each of the three health care systems on the number of patients at the seventeen specific clinics (VA, 4 clinics; UCHealth, 5 clinics; DH 8 clinics) to be included in this study. Figure 1, which we believe to present conservative estimates of enrollment, shows estimated numbers of patients with CVD conditions and prescribed CVD medications across the 3 HCS. Patients will be sent a letter with the opportunity to opt-out. In addition, care providers will be provided lists of their patients who are potentially eligible for the study to see if there are patients that they feel should not be included in the study. Assuming that 75% of patients have a gap, another 15% of patients opt-out of the study following randomization, and 10% of patients do not have usable outcome data, we expect to have usable outcome data for about 7,740 patients across the four study arms.

*Planned enrollment:* In the patient accrual proposal sent to NHLBI (March 20, 2019), we proposed to enroll 5,000 patients which is a conservative estimate in case the number of patients opting out is higher than in the pilot or the number of patients with 7-day gaps is lower than estimated. Even with this conservative estimate, we expect to have ample subjects (nearly ten times as many as needed) to achieve the necessary power for the primary analysis of PDC. Additional subjects will provide power for secondary analyses, and for analyses of secondary outcomes.

*Achieved sample sizes 3/2023*

We surpassed our overall enrollment goal of 5,000 patients, enrolling 1,235 patients the VA, 7,266 patients at Denver Health, and 1,000 patients at UCHealth.

**3.2 Analysis of secondary outcomes**

The secondary outcomes of clinical events, care utilization and cost will be analyzed using similar approaches as for the primary outcome but based on appropriate models, e.g. Cox survival models for time to clinical event or rehospitalization, generalized gamma regression for cost, etc. Standardization methods allow results to be expressed on interpretable scales such

as risk difference (e.g. Sjolander, 2016).[22] Data will be analyzed using SAS (SAS Institute Inc., Cary, NC) and R software. In the table below, we provide an estimate of the secondary outcomes of interest based on a cohort of patients who would have been potentially eligible for the study from 2 of our health systems. This illustrates that we are able to capture these outcomes.

**Outcomes of interest for the pragmatic trial. This is based on the cohort of patients identified from 2017-2018 who would have been eligible for the study and we followed them over time to assess outcomes**

| Outcome | Patient population based on the presence of specific comorbidity for which the outcome is relevant | Outcomes of interest (DH) | Outcomes of interest (VA) |
|---|---|---|---|
| Systolic BP - Mean (SD) mm Hg | All | 131.2 (17.9) | 133.8 (18.6) |
| Diastolic BP - Mean (SD) mm Hg | All | 78.7 (10.8) | 79.4 (10.3) |
| LDL - Mean (SD) | All | 85.5 (38.6) | 90.1 (33.1) |
| Hemoglobin A1c - Mean (SD) | All | 7.8 (1.9) | 7.0 (1.6) |
| All Cause Hospitalization (1 Yr.) | All | 8.7% (792/9149) | 13.6% (332/2447) |
| Cause Specific Hospitalization | | % (number of patients with events/number of eligible patients for the outcome) | % (number of patients with events/number of eligible patients for that outcomes) |
| Hypertension Emergency | Hypertension | 0.2% (14/7364) | 0.4% (8/1877) |
| Myocardial infarction (MI) | HTN/Hyperlipidemia/Diab/CAD | 0.2% (22/9119) | 0.3% (7/2402) |
| Stroke | HTN/Hyperlipidemia/Diab/CAD/AF | 0.1% (12/9149) | 0.2% (4/2447) |
| Heart Failure | HTN/CAD/AF | 0. 8% (58/7546) | 1.1% (23/2052) |
| Hyperglycemia | Diabetes | 0.1% (5/5122) | 0.1% (1/1048) |
| Atrial fibrillation (AF) | AF | 2.3% (8/352) | 2.9% (9/311) |
| All Cause ED Visit (1 Yr) | All | 1.9% (1700/9149) | 4.5% (1093/2447) |
| Cause Specific ED Visit | | | |
| Hypertension Emergency | Hypertension | 0 (1/7364) | 0.2% (3/1877) |
| MI | HTN/Hyperlipidemia/Diab/CAD | 0 (0/9119) | 0.1% (3/2402) |
| Stroke | HTN/Hyperlipidemia/Diab/CAD/AF | 0 (0/9149) | 0.1% (2/2447) |
| Heart Failure | HTN/CAD/AF | 0 (3/7546) | 1.2% (25/2052) |
| Hyperglycemia | Diabetes | 0.9% (48/5122) | 2.1% (22/1048) |
| AF | AF | 1.7% (6/352) | 5.8% (18/311) |
| Procedures | | | |
| PCI | HTN/Hyperlipidemia/Diab/CAD | 0.5% (45/9119) | 0.8% (20/2402) |
| CABG | HTN/Hyperlipidemia/Diab/CAD | 0 (0/9119) | 0.2% (4/2402) |
| Cardioversion | AF | 2.6% (9/352) | 1.6% (5/311) |

**References**


Lumley T, Diehr P, Emerson S, Chen L. (2002) The importance of the normality assumption in large public health data sets. Annual Review of Public Health 23:151-169.

Bell ML, Horton NJ, Dhillon HM, Bray VJ, Vardy J. (2018) Using generalized estimating equations and extensions in randomized trials with missing longitudinal patient reported outcome data. Psycho-Oncology, 27:2125-2131.

Dmitrienko A, Tamhane AC, Wiens BL. (2008) General multistage gatekeeping procedures. Biometrical Journal 50: 667-677.

Fiero MH, Hsu C-H, Bell ML. (2017) A pattern-mixture model approach for handling missing continuous outcome data in longitudinal cluster randomized trials. Statistics in Medicine 36:4094-4105.

White IR, Horton NJ, Carpenter J, Pocock SJ. (2011) Strategy for intention to treat analysis in randomized trials with missing outcome data. British Medical Journal 342:d40.

Hernán, M. A., Brumback, B. A., & Robins, J. M. (2002). Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. Statistics in medicine, 21(12), 1689-1709.

Preisser, J.S., Lohman, K.K. and Rathouz, P.J., 2002. Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. Statistics in medicine, 21(20), pp.3035-3054.

Robins, J.M., Rotnitzky, A. and Zhao, L.P., 1995. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association, 90(429), pp.106-121.

Lin G & Rodriguez RN. (2015) Weighted methods for analyzing missing data with the GEE procedure. SAS paper 166-2015, https://support.sas.com/resources/papers/proceedings14/SAS166-2014.pdf.

Xu C, Li Z, Wang M (2018). wgeesel: Weighted Generalized Estimating Equations and Model Selection. R package version 1.5, https://CRAN.R-project.org/package=wgeesel.