STATISTICAL ANALYSIS PLAN (SAP)

High-Risk Polypharmacy: Reducing High-Risk Geriatric Polypharmacy via EHR Nudges

Version Date: April 13, 2026

Version: 2.0

Supported by: National Institute on Aging R01AG070054

IRB Number: STU00213231

NCT Number: NCT05656560

List of Abbreviations

Abbreviation	Definition
ADE	Adverse drug event
CCB	Calcium channel blocker
CKD	Chronic kidney disease
eGFR	Estimated glomerular filtration rate
EHR	Electronic health record
EDW	Enterprise data warehouse
HF	Heart failure
HFrEF	Heart failure, reduced ejection fraction
HRPP	High risk polypharmacy
NM	Northwestern Medicine
NSAID	Nonsteroidal anti-inflammatory drug
SNRI	Serotonin norepinephrine reuptake inhibitor
SSRI	Selective serotonin reuptake inhibitor
UPMC	University of Pittsburgh Medical Center; affiliated with the University of Pittsburgh Schools of the Health Sciences

STATISTICAL ANALYSIS PLAN (SAP)

High-Risk Polypharmacy: Reducing High-Risk Geriatric Polypharmacy via EHR Nudges

Principal Investigator: Stephen D. Persell, MD, MPH

Statistician: Lucia C. Petito, PhD

Table of Contents

1. INTRODUCTION

This document outlines the proposed analyses for the High-Risk Polypharmacy (HRPP) clinical trial, which aims to compare the effectiveness of two types of clinical decision support tools (commitment and justification nudges) to reduce high-risk prescribing and adverse events in older patients who take many (5 or more) pharmaceutical drugs, which can interact to increase risk of falls and other unintended consequences (high-risk polypharmacy). Inappropriate polypharmacy, defined as the administration of more drugs than clinically necessary, is highly prevalent among geriatric patients in the United States. Polypharmacy is associated with adverse drug reactions, drug interactions, medication nonadherence, adverse events such as falls and mortality, and higher costs of care. We plan to conduct a four-arm, pragmatic, 2x2 factorial cluster randomized controlled trial with equal allocation (1:1:1:1 clinics) in adult patients aged 65 years or older who have polypharmacy (use five or more systemic prescription medications) seen at participating clinics at two health systems: Northwestern Medicine (NM) and University of Pittsburgh Medical Center (UPMC). The purpose of this document is to provide details regarding the statistical analysis plan (SAP) for this study.

1.1 Study Aims

The overarching study aims are as follows:

Aim 1: To evaluate the effects of an EHR-based commitment nudge, a justification nudge, and the combination of both nudges on a composite measure of high-risk polypharmacy via a pragmatic randomized controlled trial. *We will use cluster randomization in which primary care clinics are randomized to receive 0, 1, or 2 nudges using a factorial design. The nudges will run for 18 months, followed by 12 months of observation to assess persistence of effects*.

We hypothesize that the intervention will decrease polypharmacy, may decrease emergency care and hospitalizations, and improve the care of older adults.

After the 18-month intervention period, we will collect efficacy outcomes for an additional 12 months. We hypothesize that any observed differences in the trajectories of the rates of HRPP present at the end of the intervention period between active intervention groups and control will diminish during the 12-month follow-up period, and that the decline in effect will be greater for the single intervention group than the dual intervention group. This dissipation of effects is expected, although it may occur quite slowly because medication changes are fairly infrequent.

Exploratory Aim 1: We will perform sub-group analysis to examine patients with and without dementia or mild cognitive impairment. We also plan to conduct pre-specified sub-group analyses to examine whether the effects of the interventions differ by age category, sex, race, ethnicity, and degree of polypharmacy.

We hypothesize that the interventions' effects may be stronger among patients with dementia or mild cognitive impairment, the very elderly, and those with more extensive polypharmacy.

Exploratory Aim 2: To qualitatively and quantitatively assess clinician experiences with the EHR-based nudges, including their acceptability and effects on workflow. *At the conclusion of the intervention period, we will perform semi-structured interviews and field a clinician survey*.

This SAP will focus on the details of the analyses for Aim 1 and Exploratory Aim 1; we reserve details of the exploration of clinician experiences (Exploratory Aim 2) for future work.

Study time points include baseline assessment of clinics (to evaluate annual performance for use in constrained randomization), a 12-month historical baseline period, an 18-month active intervention period, and a 12-month post-intervention period (to evaluate persistence of effects).

2. STUDY OUTCOMES

In the sections below, we include the relevant specific field names for variables within the study database as of the time of SAP creation. These are indicated by square brackets, and include: month, composite_HRPP, composite_HRPP_deno, dementia, fall_cndtn_rx, fall_cndtn_rx_deno, fall_rx_rx, fall_rx_rx_deno, hf_thia, hf_thia_deno, hf_nsaid, hf_nsaid_deno, hf_ccb, hf_ccb_deno, ckd_gly, ckd_gly_deno, ckd_nsaid, and ckd_nsaid_deno. All primary, secondary, and safety outcomes are produced from data abstracted from the EHR at each participating health system. The primary, and secondary outcomes were developed and piloted in the initial R21 study [R21AG057396].[1]

2.1 Primary Outcome

The primary effectiveness outcome at the patient level is an indicator of an HRPP Composite Outcome [composite_HRPP], which is experiencing one or more of the drug-conditions or drug-drug combinations listed in the HRPP Component Measures occurring in an eligible patient determined by an examination of the EHR active medication list and EHR-recorded discrete diagnoses. Eligible patients (the denominator for the primary outcome) will have been seen at a participating clinic by a participating clinician for an in-person or telehealth visit in the prior 365 days, be age 65 years or older at the beginning of the 365 days look-back period, and have 5 or more qualifying medications on the active medication list at the time of the outcome assessment [composite_HRPP_deno]. Each of the seven components of the HRPP composite is defined under “Secondary Outcomes.” This variable will be calculated on a monthly basis using the preceding 365 days of data. A detailed list of the specifications for each component of the HRPP composite outcome is available in Section 11.

2.2 Secondary Outcomes

Each individual measure component of the composite is a secondary outcome. The individual components of the HRPP Composite will use the same general eligibility criteria as the primary outcome along with component-specific denominator criteria. A brief description of the components of the HRPP composite outcomes is included in **Table 1**, with detailed descriptions for abstraction of data components from the EHR available in Section 11 below. These measures will be calculated on a monthly basis using the preceding 365 days of data, as for the primary HRPP composite outcome.

Table 1. Overview of Components of the HRPP Composite Outcome

Outcome Description		Database Information	
Measure	**Brief Description**	**Outcome Indicator**	**Eligibility Indicator**
Fall condition-drug interaction	≥1 active prescription for eligible fall risk	fall_cndtn_rx	fall_cndtn_rx_deno

	medications (Opioids are not included)		
Fall drug-drug interaction	≥ 3 active prescriptions for eligible fall risk mediations	fall_rx_rx	fall_rx_rx_deno
Heart failure-thiazolidinedione interaction	Among those with HF, ≥1 active prescription for eligible thiazolidinedione medications	hf_thia	hf_thia_deno
Heart failure-NSAID interaction	Among those with HF, ≥ 1 active prescription for eligible NSAIDs	hf_nsaid	hf_nsaid_deno
Heart failure with reduced ejection fraction-non-dihydropyridine calcium channel blocker interaction	Among those with HFrEF (EF < 40%), ≥ 1 active prescription for eligible CCB medications	hf_ccb	hf_ccb_deno
CKD-glyburide/glimepiride interaction	Among those with CKD, ≥ 1 active prescription for eligible glyburide-containing medications	ckd_gly	ckd_gly_deno
CKD-NSAID interaction	Among those with CKD, ≥ 1 active prescription for eligible NSAIDs	ckd_nsaid	ckd_nsaid_deno
Note: Prescriptions are active medications on the "medication list" available through the EHR.			

Additional secondary outcome measures at the patient level are: (1) emergency department visits – all cause [ed_visit_any], (2) emergency department visits – potentially adverse drug event-specific [ed_visit_ae], (3) hospital admissions – all cause [hosp_adm_any], and (4) hospital admissions – potentially adverse drug event specific [hosp_adm_ae]. Eligibility for these secondary outcomes will be the same as for the primary HRPP composite [composite_HRPP_deno]. Emergency department visits and hospitalizations will be assessed by queries of the NM or UMPC electronic medical records and will only include visits occurring within these health systems. Potentially adverse drug event-specific emergency visit and hospitalizations will be identified by the ICD-10 codes listed in Table 2. These variables will be calculated on a quarterly basis using the 365 days of data preceding the start of the quarter.

Table 2. Diagnosis Criteria for Potentially ADE-Specific ED Visits and Hospitalizations

ADE	ICD-10 values
Fall	W00-W19 and all sub-codes [all categories of falls] S72.0 and all sub-codes [fracture of head and neck of femur] S72.1 and all sub-codes [pertrochanteric fracture] S72.2 and all sub-codes [subtrochanteric fracture of femur]
Acute kidney injury	N17 and all sub-codes [acute kidney injury (nontraumatic)], N19 [unspecified kidney failure]
Hypoglycemia	E16.0, E16.1, E16.2 [hypoglycemia codes], E13.64 and all subcodes [DM2 with hypoglycemia], E10.64 and all subcodes

	[DM1 with hypoglycemia], T38.3 and all subcodes [poisoning by, adverse effect of, and underdosing of insulin and oral hypoglycemic drugs]
Fluid retention and heart failure exacerbation	I50.21 [acute systolic heart failure], I50.23 [acute on chronic systolic heart failure], I50.31 [acute diastolic heart failure], I50.33 [acute on chronic diastolic heart failure], I50.41 [acute combined systolic and diastolic heart failure], I50.43 [acute on chronic combined systolic and diastolic heart failure], I50.811 [acute right heart failure], I50.813 [acute on chronic right heart failure], J81.0 [acute pulmonary edema]

2.3 Safety Outcomes

Described below are the safety measures that we will assess during this trial using EHR data. We will conduct clinician chart review for events possibly or probably related to the clinical decision support interventions. We will report the denominator and numerator counts (and the rate per 10,000 patients per year) in the month prior to the intervention start and during the last month of the active intervention period in each study arm. These represent unlikely but conceivable adverse consequences of discontinuing or reducing the high-risk medications targeted by the nudges. Patients will be considered eligible for the denominator of a safety measure if they met that specific drug-condition or drug-drug high-risk polypharmacy criteria during the 365 days prior to the safety assessment period. We will collect safety assessments by quarter. For example, patients who met the safety-measure-specific high-risk polypharmacy criteria at any time from 1/1/2023 to 12/31/2023 would be in the denominator of patients assessed for the safety outcome from 1/1/2024 through 3/31/2024. We will also investigate all clinician-reported adverse events or unanticipated problems for all study clinical decision support (CDS).

1. **Safety measure for discontinuing or reducing benzodiazepines and anticonvulsants: hospitalization or emergency department (ED) visit for seizure.** Event is the occurrence of a hospitalization or ED visit for seizure, detected through a query of the Northwestern and UPMC EDWs. The denominator for this measure is all patients who met criteria for fall condition-drug interaction or fall drug-drug interaction with a benzodiazepine or anticonvulsant medication on the active medication list during the 365 days prior to the safety assessment period.
2. **Safety measure for discontinuing or reducing tricyclic antidepressants, gabapentinoids, serotonin and norepinephrine reuptake inhibitors, opioids, and NSAIDs: hospitalization or ED visit for pain.** Event is the occurrence of a hospitalization or ED visit for pain, detected through a query of the Northwestern and UPMC EDWs. The denominator for this measure is all patients who met criteria for fall condition-drug interaction or fall drug-drug interaction with a tricyclic antidepressant, gabapentinoid, SNRI or opioid on the active medication list or the CKD-NSAID interaction or Heart failure-NSAID interaction with an NSAID on the active medication list during the 365 days prior to the safety assessment period.
3. **Safety measure for discontinuing or reducing tricyclic antidepressants, serotonin and norepinephrine reuptake inhibitors and selective serotonin reuptake inhibitors: hospitalization or ED visit for depression or suicidality.** Event is the occurrence of a hospitalization or ED visit for depression or suicidality, detected through a query of the Northwestern and UPMC EDWs. The denominator for this measure is all

patients who met criteria for fall condition-drug interaction or fall drug-drug interaction with a tricyclic antidepressant, SNRI or SSRI on the active medication list during the 365 days prior to the safety assessment period.

4. **Safety measure for discontinuing or reducing benzodiazepines, serotonin and norepinephrine reuptake inhibitors and selective serotonin reuptake inhibitors: hospitalization or ED visit for anxiety.** Event is the occurrence of a hospitalization or ED visit for anxiety, detected through a query of the Northwestern and UPMC EDWs. The denominator for this measure is all patients who met criteria for fall condition-drug interaction or fall drug-drug interaction with a benzodiazepine, SNRI, or SSRI on the active medication list during the 365 days prior to the safety assessment period.
5. **Safety measure for discontinuing or reducing benzodiazepines and Z-drugs (GABA receptor modulators): hospitalization or ED visit for sedative withdrawal.** Event is the occurrence of a hospitalization or ED visit for withdrawal from sedative, hypnotic, or anxiolytic drug, detected through a query of the Northwestern and UPMC EDWs. The denominator for this measure is all patients who met criteria for fall condition-drug interaction or fall drug-drug interaction with a benzodiazepine or Z-drug on the active medication list during the 365 days prior to the safety assessment period.
6. **Safety measure for discontinuing or reducing opioids: hospitalization or ED visit for opioid withdrawal.** Event is the occurrence of a hospitalization or ED visit for opioid withdrawal, detected through a query of the Northwestern and UPMC EDWs. The denominator for this measure is all patients who met criteria for fall condition-drug interaction or fall drug-drug interaction with an opioid drug on the active medication list during the 365 days prior to the safety assessment period.
7. **Safety measure for discontinuing or reducing antipsychotics: hospitalization or ED visit for behavioral disturbance in dementia or psychosis.** Event is the occurrence of a hospitalization or ED visit for dementia with behavioral disturbance, or psychosis detected through a query of the Northwestern and UPMC EDWs. The denominator for this measure is all patients who met criteria for fall condition-drug interaction or fall drug-drug interaction with an antipsychotic drug on the active medication list during the 365 days prior to the safety assessment period.
8. **Safety measure for discontinuing or reducing glyburide, glimepiride, and thiazolidinediones: hospitalization or ED visit for hyperglycemia.** Event is the occurrence of a hospitalization or ED visit for hyperglycemia detected through a query of the Northwestern and UPMC EDWs. The denominator for this measure is all patients for who met criteria for CKD-glyburide/glimepiride interaction with glyburide or glimepiride, on the active medication list at the time of the event or in the year preceding the event or heart failure-thiazolidinedione interaction with a thiazolidinedione on the active medication list during the 365 days prior to the safety assessment period.
9. **Safety measure for discontinuing or reducing non-dihydropyridine calcium channel blockers: hospitalization or ED visit for tachycardia (other than ventricular) or hypertensive emergency.** Event is the occurrence of a hospitalization or ED visit for tachycardia (other than ventricular), detected through a query of the Northwestern and UPMC EDWs. The denominator for this measure is all patients who met criteria for heart failure with reduced ejection fraction-non-dihydropyridine calcium channel blocker interaction - with a non-dihydropyridine calcium channel blocker on the active medication list during the 365 days prior to the safety assessment period.

10. **Clinician reported adverse events/unanticipated problems.** At any time, participating clinicians can report an adverse event or unanticipated problem potentially related to study participation.

We will assess additional exploratory and safety outcomes, as recommended by the Data Safety Monitoring Board (DSMB), on an as-needed basis.

3. PATIENT AND CLINICIAN DEMOGRAPHICS AND CLINIC-LEVEL CHARACTERISTICS

The following are specific patient and clinician demographics and clinic-level baseline assessments of interest for analyses. Primary analyses will adjust for fixed clinic-level covariates included in the constrained randomization procedure (detailed in Section 5 of this SAP). We plan to report marginal intervention effect estimates after adjustment for said covariates, regardless of significance.

3.1 Clinic-level Characteristics at Randomization

1) Practice region: At NM (NMG – Central; NMG – North; RMG); At UPMC these will be split into five existing administrative groups

2) Number of clinicians attributed to each participating clinic. Each clinician is attributed to a clinic based on where they provided the plurality of their patient care in the 12-month randomization period

3) Number of patients eligible for the composite HRPP outcome (detailed in “Primary Outcome” above) in the 12-month randomization period

4) Rate per 100 eligible patients of the composite HRPP outcome (detailed in “Primary Outcome” above) in the 12-month randomization period

Data to calculate 2-4 were abstracted from the EHR at NM on 5/25/2023 for the period 3/1/2022 – 2/28/2023 and at UPMC for the period 3/1/2023 – 2/29/2024.

3.2 Patient Demographics

Additional patient demographics of interest, abstracted from the EHR, will be used to describe the patient population and conduct subgroup analyses. These include:

1) Sex

2) Age

3) Race

4) Ethnicity

5) Presence or absence of dementia diagnosis, or diagnosis of mild cognitive impairment (defined using ICD-10 codes)

6) Osteoporosis (defined using ICD-10 codes)

7) AHRQ Elixhauser comorbidity index

- Modified to exclude the following sensitive diagnoses: acquired immune deficiency syndrome, alcohol abuse, depression, drug abuse, psychoses)
- Operationalized as the person-specific count of the remaining conditions used to define the index as of the date the intervention started (October 9, 2023 at NM; November 11, 2024 at UPMC).
- Definitions of conditions are available in Moore et al. (2017).[2]

ICD-10 codes will be taken from the active problem list or the encounter diagnosis at the encounter.

3.3 Clinician Demographics

The following clinician demographics, obtained from the two health systems or from public facing websites, will be used to describe the clinicians and to conduct pre-specified subgroup analyses:

1) Gender (assumed from public-facing websites)

2) Specialty (family medicine, geriatrics, internal medicine)

3) Type of clinician (MD/DO versus other)

4. DATA STORAGE

Data will be collected and managed using a variety of tools. For NM data abstracted from the EHR (baseline clinic-level characteristics, patient-level demographics, and patient-level outcomes) will be housed in the secured file server hosted by Feinberg School of Medicine (FSM) at Northwestern University (NU) in the excel format. For the UPMC data, the limited data set will be securely transferred to the secured FSM file server at NU for the analysis.

Clinician survey data will be collected and managed using Research Electronic Data Capture (REDCap) housed at Northwestern University's Clinical and Translational Sciences Institute (CTSA), NUCATS. REDCap is a secure, web-based application designed for research studies that provides an intuitive interface for validated data entry, audit trails for tracking data manipulation and export procedures, and automated export procedures for seamless data downloads to common statistical packages, and procedures for importing data from external sources.[3] Refer to the study Data Management and Sharing Plan (DMSP) for details.

5. RANDOMIZATION METHODS

This study uses a 2 x 2 factorial cluster-randomized design with equal allocation (1:1:1:1) across study arms. We will use a modified constrained randomization process, stratified by healthcare system (NM and UPMC), to randomize practices and maintain exchangeability between clinics assigned to each study arm. The clinic-specific characteristics, calculated in the randomization period detailed above in Section 3.1, include 1) rate of HRPP composite outcome, 2) size of eligible patient population, 3) number of clinicians attributed to the clinic, and 4) region within the health system. Participating clinics will then be randomized to 1 of 4 study arms:

- The ‘Control’ arm will not receive any interventions other than the invitation to a brief educational module
- The ‘Commitment’ nudge plus brief educational module,
- The ‘Justification’ nudge plus brief educational module, or
- The Commitment and Justification nudges (‘Both’ nudges) plus brief educational module.

The study interventions will be delivered to clinicians based on their primary practices (i.e., the practices where they provide the majority plurality of their patient care), with clinical decision support keyed to the clinicians’ EHR identifiers. These are assigned during the randomization period and treated as fixed through the study period.

We will use a modification of the methodology proposed in Siddique et al. (2024) to implement the constrained randomization algorithm, in a manner similar to our approach in previous studies.[4-6] First, the study statistician (Dr. Petito) will generate possible 1:1:1:1 randomization schemes. Since there are many ways to randomize the participating practices into 4 groups, the study statistician will create 100,000 possible sequences. Second, the balance of the four prespecified variables (which were “constrained” upon) will then be evaluated in the 100,000 possibilities. Schemes will be deemed candidate schemes if they had “adequate balance,” adapting the minimal sufficient balance approach. For the three continuous variables, the balance metric will be the total inverse variance weighted Euclidean distances in each study arm, summed across clusters. We then will call a candidate scheme ‘adequately balanced’ if the balance metric was in the top 5% of all candidates. Then we will further restrict the space of candidate schemes to allow no more than a difference of one clinic across each arm for each category of the “region” variable. Last, the final sequence will be selected from the candidate schemes (which were adequately balanced) at random with equal probability.

As the inverse variance weighted balance metric is dependent on the normality of the continuous variables included in the constrained randomization, pre-randomization analysis will include visual inspection of the empirical distribution of each of the three continuous variables. If they are not normally distributed, a natural log transformation will be considered.

The study statistician (Dr. Petito) will generate the final study randomization sequence after setting a seed number to ensure reproducibility, separately for each study site. The study statistical analyst (Ms. Lee) will upload a “Production” randomization list into REDCap for use when deploying the baseline and follow-up clinician surveys. The randomization lists will be housed on Northwestern University’s “FSMResFiles” with restricted access so only the statistical team (Dr. Petito, Ms. Lee) can access. This is an unblinded pragmatic trial, so the clinics and attributed clinicians will be aware of the intervention they receive.

Randomization occurred on June 8, 2023 at NM and on May 16, 2024 at UPMC.

6. STATISTICAL METHODS

6.1 General considerations

Descriptive statistics will summarize participant demographics and baseline clinical outcomes overall and across arms: proportion (percentages) for categorical variables; mean (± standard deviation) for continuous variables; and median (interquartile range) for skewed or count variables. Analyses in general will employ normal theory methods and residual diagnostics will

evaluate validity of assumptions; where appropriate (i.e., in the event of low cell counts for categorical data or questions of normality), transformation of variables, nonparametric methods, or exact tests may be employed. All primary effectiveness and safety analyses will be pre-specified as outlined in this SAP, and deviations from planned analyses will be labeled as such in any reports or dissemination materials.

Multiple testing. The primary and secondary outcome analyses will assume a two-sided 5% type I error rate. We will employ a Bonferroni correction for multiple hypothesis tests, as each intervention arm (justification, commitment, both) will be compared to the same control group. Thus, all reported "95% confidence intervals" will correspond to an alpha of 0.017 (0.05 / 3) and be calculated as 98.3% confidence intervals. Safety outcome analyses will not be Bonferroni-corrected. Interaction analyses will be interpreted at a relaxed two-sided 10% type I error rate.

Software. All analyses will be conducted in SAS or R (version ≥ 4.3) using the lme4 package for the logistic mixed models. Analysis code for the primary analysis will be made available alongside the publication.

6.2 Estimand

The primary estimand is the difference between treatment arms in difference in the monthly change in the predicted probability (or log-odds) of the outcome between the pre-intervention period and the intervention period. In essence, this corresponds to the treatment x time x I(time > 0) interaction term in a repeated-measures model. Under the DiD framework, this is:

$$DiD = [\Delta Rate_{pre} - \Delta Rate_{post} \mid A = 1] - [\Delta Rate_{pre} - \Delta Rate_{post} \mid A = 0]$$

where Δ denotes the monthly change in the rate of the outcome within each group. The parallel trends assumption underpins the causal interpretation: in the absence of treatment, the rate of change in outcome risk would have been the same across arms.

In this study, we will target 2 DiD estimands – one for the commitment intervention, one for the justification intervention. We will also investigate whether, when both present, these interventions interact to further reduce the risk of the outcome.

6.3 Planned Primary, Secondary, and Safety Analyses

To test our hypotheses that the 'commitment', 'justification', and 'both' interventions are effective at reducing the composite HRPP outcome, we will employ a mixed-effects hierarchical logistic regression model to estimate the adjusted marginal effect over time of each intervention on the binary outcome. Fixed effects will include intervention assignment (indicator for each of the 3 active intervention arms), time period (historical baseline versus intervention period), time period interacted with intervention assignment, time (continuous, linear), and a 3-way interaction term between time period, intervention, and time (continuous). Clinicians will be included as random effects. To isolate the effect of the intervention on the primary outcome, we will also include fixed effects for the clinic-level variables included in the constrained randomization procedure. The underlying statistical model will be

$$logit\left[\Pr\left(Y_{ijkt} = 1\right)\right] = \beta_0 + \beta_{0jk} + \beta_1 C_k + \beta_2 J_k + \beta_3 t + \beta_4 Post_{ijkt} + \beta_5 C_k t + \beta_6 J_k t + \beta_7 C_k J_k + \beta_8 C_k J_k t +$$

$$\beta_9 C_k Post_{ijkt} + \beta_{10} J_k Post_{ijkt} + \beta_{11} C_k J_k Post_{ijkt} +$$

$$\beta_{12}C_k tPost_{ijkt} + \beta_{13}J_k tPost_{ijkt} + \beta_{14}C_k J_k tPost_{ijkt} + \overrightarrow{\beta_{15}}\overrightarrow{W_k}$$

Where Y_{ijkt} corresponds to the primary outcome for the *i*th person seen by the *j*th clinician at the *k*th clinic at time *t* (in months). $C_k, J_k \in [0,1]$ correspond to treatment indicators for the commitment and justification interventions, respectively. $Post_{ijkt} = 1(t > 0)$ corresponds to whether the observation occurred pre- or post-baseline. $\overrightarrow{W_k}$ is the vector of clinic-level covariates as measured at randomization that were included in the constrained randomization procedure. $\beta_{0jk} \sim N(0, \sigma_\beta^2)$ is the clinician-level random intercept. The coefficient on the 3-way interaction term, β_{12}, will represent whether the log-odds of the monthly rate of change in the primary outcome differs between the commitment intervention and control patients during the intervention period. The coefficient on the 3-way interaction term, β_{13}, will represent whether the log-odds of the monthly rate of change in the primary outcome differs between the justification intervention and control patients during the intervention period.

Interaction effects. β_{14} tests whether the effect of C on the time trend differs by level of J (and vice versa). If this term is statistically significant at the pre-specified threshold of $\alpha = 0.10$ (a relaxed threshold reflecting lower power for interaction terms), results will be reported separately by subgroup (i.e., four-arm analysis) rather than as pooled main effects. If not significant, the interaction term will be retained in the model as specified above, but main effects will be the focus of inference.

Assumption of linear time. Time (t) will be modeled as linear by default. Linearity will be assessed visually by plotting observed proportions by arm across time points with smoothed curves (loess). If non-linearity is detected, a **natural cubic spline** with knots at tertiles of the follow-up period will be used, and the underlying DiD model will be summarized as the average monthly slope over the 18-month intervention period.

Subgroup analyses. The primary outcome will be examined in the following pre-specified subgroups:

- Sex
- Race: Asian, Black/African-American, White, American Indian/Alaskan Native, Native Hawaiian/Other Pacific Islander, multiracial/other, unknown
- Ethnicity: Hispanic or Latino vs not
- Age categories: 65-74, 75-84, 85+
- Polypharmacy (number of systemic prescription meds): 5-9, 10-14, 15+
- Osteoporosis: present vs absent
- Dementia or mild cognitive impairment: present vs absent

The secondary and safety outcomes will be examined in a similar fashion. If they are sufficiently rare (for example, no events in a given month), model convergence may become an issue. First, we plan to expand the measurement period from monthly to quarterly, and fit the same model on a coarser time scale. If that fails, we will abandon the time-varying aspect of the analysis and do one pull for the historical baseline period and one pull for the intervention period. This corresponds to a comparison of mean outcomes during entirety of the historical baseline and active intervention periods. The mixed effects logistic regression model would then be run without including continuous time, only an indicator for active intervention. If that still fails to

converge, we will exclude the random effects for providers and use a fixed effects logistic regression model.

As this study is conducted at 2 health systems, we will also conduct health system-specific analyses. Any differences in effectiveness will be contextualized with differences in intervention implementation.

Reporting. Results will be reported in accordance with the 2023 update to the CONSORT extension for factorial trials.[7] For all study outcomes, we will report the denominator and numerator counts in the month prior to the intervention beginning and during the last month of the active intervention, as well as a rate per 10,000 patients per year for these measures in each study arm (unadjusted data). For model-based effect estimation, we will margin over the clinician random intercepts and clinic-level covariates to recover arm-specific marginal predicted risks over time. The primary DiD estimate will be displayed graphically as marginal predicted risk trajectories by arm over time, with 95% confidence bands. A table of arm-specific monthly rates of change, pairwise DiD contrasts, and interaction estimates will accompany the primary results.

Secondary analysis: Health-system trends in polypharmacy. To understand how the patient population eligible for our study changes over time, we will conduct an interrupted time series analysis. The time period will be October 2022 – April 2025 for NM and November 2023 – May 2026 for UPMC. For each participating primary care clinic in month *m*, we will calculate the number of in-person or telehealth encounters with individuals aged 65 years or older that occurred between months *m-12* and month *m* (denominator). Of individuals who satisfy denominator criteria, we will determine whether the number of active medications in their chart as of the 1st day of month *m* equals or exceeds 5 (numerator). The outcome variable will be clinic-specific monthly observations of numerator / denominator – when multiplied by 100 this is the percentage of patients at a clinic in a given year who have polypharmacy. We will fit a piecewise linear model in the clinic-monthly data:

$$E[Y_{km}|m, X_k] = b_0 + b_1 m + b_2 D + b_3 P + b_4 X_k.$$

Here, m is a continuous variable which indicates months passed from the beginning of the site-specific observational period, D is an indicator variable for whether the observation was collected before or after the intervention was implemented, and P is a continuous variable indicating months since the intervention was implemented (prior to implementation, this variable takes the value of 0). We will also adjust for clinic-specific characteristics, X_k, used in the constrained randomization procedure, as well as health system. We will standardize our estimates (coefficient b_3 will display effect size – difference in the slopes comparing pre- to during intervention periods) to the observed distribution of clinic-level covariates. Robust Huber-White variance will be estimated and used to create 95% confidence intervals and do hypothesis testing.

In a sensitivity analysis, we will study trends in *hyperpolypharmacy*, altering the numerator criteria to determine whether the number of active medications in their chart as of the 1st day of month *m* equals or exceeds 10. Analysis will be conducted as described above.

Secondary analysis: Fixed-cohort analysis of time to high-risk polypharmacy (HRPP) resolution. We plan to identify a cohort of individuals who meet study eligibility criteria at baseline and have the composite high-risk polypharmacy outcome (meet at least one of the 7 components, **Table**

1). This can be thought of mimicking identification of the individuals who would be eligible to participate in a patient-randomized trial (in contrast to this pragmatic cluster-randomized trial). Among this cohort, which we anticipate to contain roughly 30-40% of individuals meeting eligibility criteria at baseline, we will study their time to high-risk polypharmacy resolution. Individuals will be followed from baseline until the earliest of (1) resolution of all high-risk polypharmacy numerator criteria (the earliest month when patients do not meet any of the 7 numerator components of HRPP, **Table 1**), 18 months after their most recent in-person or telehealth appointment, or end of the study period (18 months after baseline).

We will describe the characteristics of the cohort, using means and SDs for continuous variables and counts and percentages for categorical variables. Outcome estimation will start by estimating the intervention arm-specific probabilities of HRPP resolution over time (months), taking 1 - Kaplan-Meier-based estimates of the survival curve for having at least 1 HRPP component. Log-rank tests will be used to determine if any of the intervention arms are different from control. Then, we will use pooled logistic regression to calculate the risk difference at 18 months adjusted for clinic characteristics included in the constrained randomization algorithm, comparing the 3 intervention arms to control. The model will allow the association between time since baseline and the risk of the outcome to vary non-linearly, using restricted cubic splines with knots at 3, 9, and 15 months to model time. The risk difference will be standardized to the empirical distribution of the clinic characteristics. Sensitivity analysis will include multiple imputation for missing outcome data, assessing the robustness of our findings to differential loss to follow-up across intervention arms.

This analysis will be repeated for 7 outcomes: each of the individual components that make up the HRPP composite outcome (**Table 1**). Subgroup analysis, for the HRPP composite outcome only, will be conducted in the pre-specified subgroups listed above.

7. ANALYTIC DATASET

Analyses will be conducted in an intention-to-treat dataset, whereby all clinics with data at baseline will be included in analyses according to the study arm to which they were randomized, regardless of adherence to the study protocol (e.g. regardless of whether the clinical decision support fired appropriately). Clinicians will be attributed to the clinic at which they spent the most time during the 365 days prior to the site-specific randomization data pull (March 1, 2023 at NM; March 1, 2024 at UPMC; this assignment is fixed for the duration of the study. Patients will be attributed to clinician with whom they had the plurality of their encounters in the 12 months prior to the measurement period. There are no planned per protocol analyses.

We do not anticipate missing data, as clinics can only leave the study if 1) leadership withdraws consent; 2) the clinic closes or leaves the health system altogether; or 3) all clinicians attributed to a given clinic at baseline leave the health system after the baseline assessment period. Clinicians who leave the health system before the end of the study period (and their attributed patients) will contribute data in measurement periods when they were working at the health system. Clinicians who join a participating clinic at the health system after the historical baseline period (and their attributed patients) will not contribute data to the study.

8. POWER AND SAMPLE SIZE CONSIDERATIONS

As the two health systems included in this study, NM and UPMC, have a fixed number of clinics participating (and therefore clinicians and their patients), power calculations focused on our

power to detect a range of clinically meaningful risk reductions in the primary outcome for a given number of clinics/clinicians/patients.

For illustrative purposes, we present power calculations for the primary outcome: the rate of the composite HRPP outcome among patients aged 65 years or greater taking 5 or more medications. NM includes 59 primary care clinics with 390 clinicians who empanel >62,000 adult patients aged 65 years or greater who are taking at least 5 medications. The UPMC system looks similar, with an estimated 65 primary care clinics with approximately 414 clinicians who empanel >42,000 older adult patients expected to meet eligibility criteria. Thus, our power calculation was designed with 31 clinics in each arm, each encompassing 26,000 unique patients. Data available at the time of randomization indicated that 25.9 (SD 5.4) and 33.7 (5.5) patients met criteria for the composite HRPP outcome at NM and UPMC, respectively, per 100 patients aged 65 years or older taking at least 5 medications who attended at least one face-to-face visit with their primary care physician in the prior 365 days. We therefore assumed that the baseline event rate would vary between 26% and 33%. Assuming 100% clinician participation at both sites, clinic-level ICC of 0.017 (empirically derived from Northwestern data), and Bonferroni multiple-comparison corrections for 3 comparisons (each active treatment arm to control), we calculated the following levels of power for a coefficient from a hierarchical mixed-effects logistic regression model at the 0.05 level of significance for 2-, 1.5-, and 1-absolute percentage point reductions in the primary outcome (**Table 3**).

Table 3. Statistical Power for Bonferroni-Corrected (α=0.017) Comparisons of Event Risk Between Trial Study Arms

	Absolute risk reduction of 2 percentage points			Absolute risk reduction of 1.5 percentage points			Absolute risk reduction of 1 percentage points		
Baseline event rate	26%	30%	33%	26%	30%	33%	26%	30%	33%
Corresponding relative risk	0.923	0.933	0.939	0.942	0.950	0.955	0.962	0.967	0.970
Power for 3 comparisons	>99%	>99%	>99%	94%	92%	90%	60%	55%	53%

9. TECHNICAL DETAILS

This SAP is subject to version control, and we anticipate modifications to analytic plans. Any changes will be documented herein. These changes may be due to assumption violations, logistical issues, unexpected empirical distributions of study outcomes, or a combination thereof. In these cases, the SAP will be updated accordingly. All analyses will be performed via SAS version 9.4 or higher (The SAS Institute; Cary, NC) or R version 4.3.0 or higher (The R Foundation for Statistical Computing platform). Table and figure formatting and style may be dictated by mode of dissemination or specific target journal(s) for results dissemination.

9.1 Summary of changes between v1.0 and v2.0

- Study timeline at UPMC included
- Names of variables in database were updated (Section 2)
- Patient characteristics updated to include comorbidities (detailed in Section 3.2)
- Finalized definition of underlying statistical model and refined all statistical analysis plans included in Section 6

- Updated power calculations to reflect randomization event rate and number of participating clinics (Section 8)
- Grammatical changes throughout

10. TIMELINE FOR ANALYSES

As this is a low-risk clinical trial, the SAP does not include any formal interim statistical analyses involving hypothesis testing or any pre-specified stopping criteria for effectiveness or futility on primary or secondary outcomes. Interim reports to the study team and data and safety monitoring board (DSMB) will consist of process measures such as missing values (clinicians leaving the health system) and simple descriptive statistics on primary and safety outcomes of interest. Quality control and assurance for primary, secondary, and safety outcomes will be conducted on a quarterly basis.

We foresee the DSMB requiring specific data listings or summarizations, but these will be specified at the time of the relevant DSMB meeting(s). Open sessions of DSMB meetings will review data NOT stratified by study arm to preserve blindedness by most study investigators. At DSMB meetings, study outcomes stratified by study arm may be presented in a closed session to the DSMB members, study statistician (Dr. Petito), and study statistical analyst (Ms. Lee); the DSMB members will be un-blinded upon request.

To preserve the integrity of the study, no formal final statistical analyses will occur until the REDCap database has been locked and all queries/discrepancies resolved; the date of database lock will be documented.

11. APPENDIX: HRPP COMPOSITE MEASURE SPECIFICATIONS

Table 4. Detailed definitions of criteria for monthly measurements of each component of the primary HRPP composite outcome

Measure	Denominator	Numerator
Fall condition-drug interaction	All patients aged ≥65 years on the year prior to measurement date AND had at least one qualifying face-to-face or telehealth office visit (s) during one year prior to measurement date AND Alive AND total active medications[1] ≥5 AND have ≥1 fall marker[2] within the past 2 years	≥1 active eligible fall risk medication(s) (Opioids are not included as a criterion for fall condition-drug interaction)
Fall drug-drug interaction	all patients aged ≥65 years on the year prior to measurement date AND had at least one qualifying face-to-face or telehealth office visit(s) during one year prior to measurement date AND Alive AND total active medications[1] ≥ 5	≥ 3 active eligible fall risk mediations
HF-thiazolidinedione interaction	all patients aged ≥ 65 years on the year prior to measurement date AND had at least one qualifying face-to-face or telehealth office visit(s) during one year prior to measurement date AND Alive AND total active medications[1] ≥ 5 AND have ≥ 1 HF marker[3] within the past 2 years	≥1 active eligible thiazolidinedione medication(s)
HF-NSAID interaction	all patients aged ≥ 65 years on the year prior to measurement date AND had at least one qualifying face-to-face or telehealth office visit(s) during one year prior to measurement date AND Alive AND total active medications[1] ≥ 5 AND have ≥ 1 HF marker[3] within the past 2 years	≥ 1 active eligible NSAID medication(s)
HF-non-dihydropyridine calcium channel blocker interaction	all patients aged ≥ 65 years on the year prior to measurement date AND had at least one qualifying face-to-face or telehealth office visit(s) during one year prior to measurement date AND Alive AND total active medications[1] ≥ 5 AND have ≥ 1 HFrEF marker[4] within the past 2 years	≥ 1 active eligible CCB medication(s)
	all patients aged ≥ 65 years on the year prior to measurement date AND	≥ 1 active eligible glyburide or glimepiride-containing medication(s)

CKD-glyburide/glimepiride interaction	had at least one qualifying face-to-face or telehealth office visit(s) during one year prior to measurement date AND Alive AND total active medications[1] ≥ 5 AND have most-recent eGFR[5] < 60 within the past 2 years	
CKD-NSAID interaction	all patients aged ≥ 65 years on the year prior to measurement date AND had at least one qualifying face-to-face or telehealth office visit(s) during one year prior to measurement date AND Alive AND total active medications[1] ≥ 5 AND have most-recent eGFR[5] < 30 within the past 2 years	≥ 1 active eligible NSAID medication(s)
Primary Composite Measure: Experiencing any of the 7 component measures	Met at least one of the above criteria	Experienced at least one of the above criteria

Footnotes

1. Logic to count total active medications
   - Include only medications with ROUTE = (oral, buccal, sublingual, or transdermal)
   - de-duplicate the resulting current medication list using the field “simple_generic_name” as the basis of deduplication. Discard any entries that do not have a “simple_generic_name” value

2. Fall marker

Variable	Values
ICD-10*	W00-W19 and all sub-codes [all categories of falls] S72.0 and all sub-codes [fracture of head and neck of femur] S72.1 and all sub-codes [pertrochanteric fracture] S72.2 and all sub-codes [subtrochanteric fracture of femur]
Fall risk screening marker	Health system variable to indicate a patient responded affirmatively to a fall risk screening questionnaire.

3. Heart Failure marker

Variable	Values
ICD-10	I25.5 and all sub-codes [ischemic cardiomyopathy] I42 and all sub-codes [cardiomyopathy] I43 and all sub-codes [cardiomyopathy in other diseases] I50 and all sub-codes except I50.8 (and all subcodes) [heart failure, excluding right heart failure]

4. Heart Failure low ejection fraction marker

Variable	Values
ICD-10	I50 and all sub-codes **except I50.3 (and all subcodes) and I50.8 (and all subcodes)** [heart failure, excluding CHF with preserved EF and right heart failure] I42 and all sub-codes [cardiomyopathy] I43 and all sub-codes [cardiomyopathy in other diseases] I25.5 and all sub-codes [ischemic cardiomyopathy]

5. Dementia or mild cognitive impairment

Variable	Values				
ICD-10	F03.90	G30.0	G31.09	G31.83	F01.51
	F02.80	G30.1	G31.1	G31.84	
	F02.81	G30.8	G31.2	G31.85	
	F03.90	G30.9	G31.81	G31.89	
	F03.91	G31.01	G31.82	F01.50	

6. eGFR (estimated glomerular filtration rate): mL/min/1.73m^2 from the race-free CKD-EPI Creatinine Equation[8]

$$eGFR = 142 \times min\left(\frac{S_{cr}}{K}, 1\right)^{\alpha} \times max\left(\frac{S_{cr}}{K}, 1\right)^{-1.2} \times 0.9938^{Age} \times \beta,$$

where:

S_{cr} (Standardized serum creatinine) = mg/dL

K = 0.7 if female or 0.9 if male

α = -0.241 if female or -0.302 if male

β = 1.012 if female or 1 if male

12. REFERENCES

1. Persell SD, Brown T, Doctor JN, et al. Development of High-Risk Geriatric Polypharmacy Electronic Clinical Quality Measures and a Pilot Test of EHR Nudges Based on These Measures. *J Gen Intern Med*. Aug 2022;37(11):2777-2785. doi:10.1007/s11606-021-07296-1
2. Moore BJ, White S, Washington R, Coenen N, Elixhauser A. Identifying Increased Risk of Readmission and In-hospital Mortality Using Hospital Administrative Data: The AHRQ Elixhauser Comorbidity Index. *Med Care*. Jul 2017;55(7):698-705. doi:10.1097/MLR.0000000000000735
3. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. Apr 2009;42(2):377-81. doi:10.1016/j.jbi.2008.08.010
4. Siddique J, Li Z, O'Brien MJ. Covariate-constrained randomization in cluster randomized 2 x 2 factorial trials: application to a diabetes prevention study. *Trials*. Sep 6 2024;25(1):593. doi:10.1186/s13063-024-08415-z
5. Brown T, Rowe TA, Lee JY, et al. Design of Behavioral Economic Applications to Geriatrics Leveraging Electronic Health Records (BEAGLE): A pragmatic cluster randomized controlled trial. *Contemp Clin Trials*. Jan 2022;112:106649. doi:10.1016/j.cct.2021.106649
6. Ciolino JD, Jackson KL, Liss DT, et al. Design of healthy hearts in the heartland (H3): A practice-randomized, comparative effectiveness study. *Contemp Clin Trials*. Aug 2018;71:47-54. doi:10.1016/j.cct.2018.06.004
7. Kahan BC, Hall SS, Beller EM, et al. Reporting of Factorial Randomized Trials: Extension of the CONSORT 2010 Statement. *JAMA*. Dec 5 2023;330(21):2106-2114. doi:10.1001/jama.2023.19793
8. Inker LA, Eneanya ND, Coresh J, et al. New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. *N Engl J Med*. Nov 4 2021;385(19):1737-1749. doi:10.1056/NEJMoa2102953