CTN-0074

Primary Care Opioid Use Disorders Treatment (PROUD) Phase 2

Statistical Analysis Plan (SAP)

Version 2.0

NCT03407638

December 21, 2020

Table of Contents

Listing of Tables

Listing of Figures

LIST OF ABBREVIATIONS

BMC	Boston Medical Center
BUP	Buprenorphine
CONSORT	Consolidated Standards of Reporting Trials
CTN	Clinical Trials Network
DEA	Drug Enforcement Agency
DD	Daily Dose
ED	Emergency Department
EHR	Electronic Health Record
EMR	Electronic Medical Record
FDA	Food and Drug Administration
FY	Fiscal Year
GLMM	Generalized Linear Mixed-effect Models
HCPCS	Healthcare Common Procedure Coding System Codes
HCS	Health Care System
IRB	Institutional Review Board
ITT	Intent-To-Treat
MA Model	Massachusetts Model
NCM	Clinic-based Nurse Care Manager
NDC	National Drug Code
NRUPC	Non-randomized usual primary care clinics
NTX	Naltrexone
OUD	Opioid Use Disorder
PC	Primary Care
PI	Principal Investigator
PPD	Pills Per Day
PROUD	Primary Care Opioid Use Disorders Treatment
QIT	Quality Improvement Team
SAP	Statistical Analysis Plan
TA	Technical Assistance
UPC	Usual Primary Care (Control Condition)
XR	Extended release

OVERVIEW OF CHANGES FROM PRIOR VERSION OF SAP

Version 1.0 of the SAP described the analysis related to the primary Objective 1 outcome of the clinic-level number of days of OUD treatment. The current version of the SAP updates the power calculations for Objective 1 addressing an error in the power calculations in alignment with the Protocol modification. The current version of the SAP also includes descriptions of the following:

- Descriptive analyses
- Augmentation of the intervention
- Additional pre-planned secondary analyses of the primary Objective 1 outcome
- Analytic plans for the main Objective 2 effectiveness outcome of acute care utilization
- Analytic plans for secondary outcome measures for Objectives 1 and 2
- Shell tables for the main Objective 1 and Objective 2 papers

1.0 SUMMARY OF STUDY DESIGN AND PROCESSES

1.1 Study Objectives

This implementation trial is a Hybrid type III, blending implementation and effectiveness objectives in a single trial, but with an emphasis on implementation objective.

The **implementation objective** of the PROUD trial (Objective 1; primary aim) is to evaluate whether the PROUD intervention designed to implement the Massachusetts (MA) Model[1] of collaborative care for management of opioid use disorders (OUDs) in primary care (PC; the "PROUD intervention") increases OUD treatment with buprenorphine or extended release (XR) injectable naltrexone (XR-NTX), documented in the electronic health records (EHRs) of PC patients, over a 2-year follow-up, as compared to usual PC (UPC).

The **effectiveness objective** (Objective 2; powered secondary aim) is to test our hypothesis that PC patients with OUDs in the 3 years prior to randomization who receive care in PROUD intervention clinics, compared to those who receive care in UPC clinics, will have fewer days of acute care utilization (including urgent care, emergency department [ED] and hospital care) in the 2 years after randomization. This "effectiveness" objective assesses whether implementation of the MA Model improves patient outcomes.

Secondary objectives explore assumptions in choice of outcomes and analyses, as well as providing descriptive, explanatory, and exploratory secondary outcomes and analyses.

Observational study aims include comparing PROUD trial (randomized) clinics to non-randomized usual primary care (NRUPC) clinics and comparing outcomes of PROUD intervention clinics to non-randomized "exemplar" clinics selected by the health systems as providing quality care for patients with OUD. These observational analyses are described in separate statistical analysis plans outside of this main trial SAP.

1.2 Study Design and Intervention

1.2.1 Study Design and Randomization

The PROUD trial is a hybrid type III pragmatic, cluster-randomized, quality improvement trial. Hybrid type III trials are mixed effectiveness and implementation trials, with greater emphasis on implementation.[2] The trial is conducted in six health systems across the United States. Randomization is stratified by health system. Each health system has recruited 2 PC clinics (or a cluster of smaller clinics) willing to implement collaborative care for patients with OUDs using a model developed at the Boston Medical Center (BMC) in Massachusetts and spread across federally qualified health clinics in that state (the "MA Model" hereafter). One of the two recruited PC clinics in each health system is randomized to implement the MA Model, while the other continues with UPC.

All quantitative data for sample identification and outcome measures are derived solely from existing electronic health records (EHRs), which include but are not limited to electronic administrative data, patients' electronic medical records (EMRs), and/or electronic data on health insurance claims.

1.2.2 The PROUD Intervention

Intervention: Implementation of the MA Model of Collaborative Care for OUDs. The PROUD trial provides financial support to cover the salary of a Nurse Care Manager (NCM) and technical assistance (TA) for the duration of the study, but the health system—not investigators—implement the MA Model program as part of quality improvement, and the health system and its clinicians provide all clinical care. One PC clinic or cluster of smaller clinics ("*PROUD intervention clinics*" hereafter) is randomized to the PROUD intervention in each health system and implements the

MA Model after randomization. Specifically, the PROUD intervention includes 3 strategies used to implement the Model in Massachusetts.

(1) Clinic leadership receives funding for a 1.0 full time equivalent NCM for 2 years after randomization and technical support for recruiting and hiring the NCM. Once hired for the study, the NCM will receive TA from experts in Massachusetts supported by PROUD, but NCMs will be employed and supervised by the health system.

(2) Experts at BMC who originally developed and disseminated the MA Model will: provide intervention clinics with a MA Model Manual; train PROUD NCMs at BMC for 1.5-2 days; and provide the ongoing TA for 2 years after randomization.

(3) At least three PC providers in the PROUD intervention clinic agreed to obtain DEA waivers to prescribe buprenorphine for OUDs (if not already waivered) and work closely with the NCM to offer high quality PC care for OUD (e.g., medication treatment with buprenorphine or injectable naltrexone with close follow-up to maximize retention in treatment), if randomized to the PROUD intervention.

Augmentation of the intervention to increase support to NCMs. PROUD included pre-planned, ongoing formative evaluation by an Implementation Monitoring Team to assess whether the 3 implementation strategies were adequate, or whether they needed to be adapted or augmented. As of early July 2019, none of the sites were consistently meeting the goal of seeing 1-2 new patients with OUD a week. A modification to the intervention was therefore approved by the IRB (7/31/2019) to add a new implementation strategy: recommending weekly interdisciplinary quality improvement meetings in the intervention clinics. Holding weekly quality improvement meetings is an evidence-based approach to quality improvement in primary care, and it was an affordable option within the PROUD budget. Health system lead investigators and project managers were asked to arrange quality improvement meetings with the intervention clinics with the goal of increasing the number of patients with OUD being treated by the clinic. The recommended elements of the quality improvement meetings were as follows. (1) Site lead investigator(s) and project managers work with clinic leaders to identify an interdisciplinary quality improvement team (QIT) consisting of champions from the intervention clinic (medical assistants, primary care waivered prescribers, front desk staff, etc.), the NCM, other clinical leaders, and themselves. (2) The QIT holds weekly meetings, ideally for an hour but at a minimum of 30 minutes, to review quality improvement activities from the past week and plan quality improvement activities for the next week using the plan-do-check-adjust (PDCA) cycle approach. (3) A note taker sends an email summarizing actions taken, results, and next steps/action items to the Boston TA Team after each meeting (cc'ing the PROUD study email box). The TA team may—as time allows at their weekly TA meetings with NCMs—ask the NCMs to share their experiences with the weekly QIT meetings, highlighting successes and lessons learned. All sites agreed to bring together an interdisciplinary group of stakeholders to help support the NCMs in regular quality improvement meetings, and some sites sent action items to the Boston TA team.

Comparison: Usual Primary Care (UPC). Clinics randomized to UPC do not receive any resources or support from the study but are free to improve OUD care in any way they choose, but they are asked not to use the OBAT manual from Boston Medical Center to replicate the PROUD intervention in the Usual PC clinic. UPC is the appropriate comparison to evaluate the impact of implementation of the MA Model on access to and quality of OUD care because most PC clinics do not currently offer treatment for OUDs, but that could change over the course of the trial.

1.3 Sample and Sample Size

The sample for the trial consists of patients who have visited the PROUD trial PC clinics in the six participating health systems. Health systems were selected for the PROUD trial based on 1) leadership support for participating in the trial, 2) elements of clinic eligibility such as adequate size and having at least 3 PC providers willing to prescribe buprenorphine in each of the PC clinics, and 3) a demonstrated ability to obtain the secondary data necessary for the PROUD trial measures—specifically days of OUD treatment with buprenorphine and injectable naltrexone and days of acute care utilization. Smaller clinics were eligible if a group of clinics near each other included adequate numbers of patients (target ~10,000 unique patients with visits in a year), and were willing to participate as a single clinic for purposes of this trial, that is, if selected to implement the MA Model, a nurse care manager (NCM) would be shared between 2 intervention clinics.

PC patients 16-90 years old with at least 1 visit to the participating clinics from 3 years before to 2 years after randomization (the 5-year study period) will be included in the trial. The total sample of PC patients in the trial is anticipated to be over ~170,000 patients across the 12 clinics, since over 14,000 patients were seen, on average, in each clinic in 2016. The implementation objective is addressed in the total trial sample, while the secondary objective is evaluated in the subsample of trial patients who have EHR documentation of an OUD during the pre-randomization period.

2.0 GENERAL PROCEDURES AND DEFINITIONS

2.1 Intent-to-treat (ITT) Analysis

Unless otherwise specified, all analyses will follow an intent-to-treat principle whereby clinics (and patients therein) will be analyzed according to the treatment arm to which they were randomized regardless of the subsequent sequence of events.

2.2 Study Day 1

The randomization date is defined as study day 0 and study day 1 is defined as the day after randomization.

2.3 Pre-randomization Period

The pre-randomization period is defined as the period 3 years prior to randomization through study day 0 (randomization date), except at one health system that changed EHR systems where the pre-randomization period is limited to 2.8 years. For simplicity, throughout the SAP this pre-randomization period is referred to as 3 years. The pre-randomization period is used to define the Objective 2 primary sample. Including the randomization date in the pre-randomization period is appropriate because health systems were notified of their randomization status the day after the randomization date.

2.3.1 Baseline Period

Because all clinics have data 2 years prior to randomization through study day 0, many covariates for descriptive analyses and regression adjustment are defined over the 2 years prior to randomization through study day 0 (since all clinics have data from this period), referred to as "baseline" measures.

2.4 Follow-up or "Post-randomization" Period

Unless otherwise specified, the follow-up period is defined as the period from study day 1 to 2 years after the randomization date (or 1.5 years for one health system that randomized 6 months late). For simplicity, throughout the SAP this follow-up ("post-randomization") period is referred to as 2 years.

2.5 Study Period

The study period is defined as up to 3 years pre-randomization (3 years at 5 health systems and 2.8 years at 1 health system) through up to 2 years following randomization (2 years at 5 health systems and 1.5 years at 1 health system). The full study period of up to 5 years—defined as the start of the pre-randomization period to the end of the follow-up period for each health system—is used to define the Objective 1 sample.

2.6 Data Collection Period

For eligible patients (defined in Section "Study Population"), data are collected from 3 years prior to randomization through 2 years post randomization (Figure 1). Limited datasets are extracted from the EMR and insurance claims data during four interim time points ("Data Pulls" 1-4) during the post-randomization period for reports to the Data and Safety Monitoring Board and refinement of data specifications and measures (Table 1). Objective 1 and 2 analyses use data from Data Pull 5 collected after the end of follow-up.

Figure 1: Data Collection Period

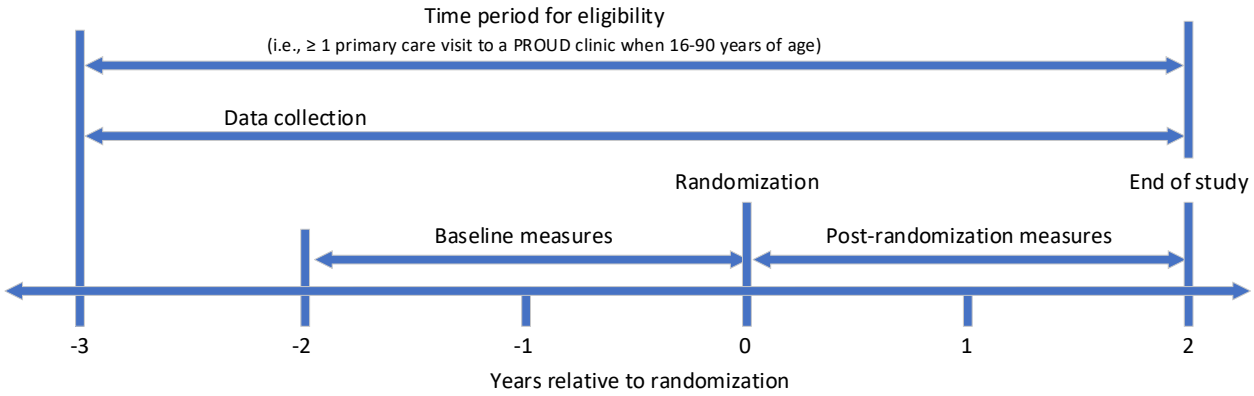


Table 1: Data Collection Timeline

Data Pull Number	Cohort Eligibility*	Data Collection*	When data are transferred to Lead Node (Post randomization)
1	T-3 to T0	T-3 to T0	6 mo
2	T-3 to T0.5	T-3 to T0.5	12 mo
3	T-3 to T1	T-3 to T1	18 mo
4	T-3 to T1.5	T-3 to T1.5	24 mo
5	T-3 to T2	T-3 to T2	33 mo

*T0 is randomization date.

3.0 STUDY POPULATION

Patients are eligible for inclusion in the trial if:

1) they had a PC visit at a PROUD trial PC clinic in the pre-randomization and/or follow-up period (defined above), and
2) age is 16 to 90 years at the time of a PC visit during the study period.

Eligible PC visits refer to outpatient clinic visits (in which the patient was age 16 to 90 on the day of the visit) to the departments of adolescent and pediatric medicine (including teen clinics), family practice, geriatric, general practice, or internal medicine that are provided by a physician, physician assistant, nurse practitioner, resident or fellow with one of these specialties.

3.1 Assignment of Patients to Clinics

Patients will be assigned to clinics based on the number of eligible PC visits. In the rare case that a patient had PC visits to both PROUD Intervention clinics and UPC clinics, they will be assigned the clinic which they visited the most pre-randomization (if they visited during that period), and if they are tied, the clinic visited nearest to and preceding the time of randomization will be considered the patient's PC clinic. If a patient only had PC visits to a trial clinic post-randomization but not pre-randomization, and they visit both PROUD and UPC clinics, they will be assigned to the clinic they visited the most, and if tied they will be assigned to the clinic they visited last.

3.2 Objective 1 Study Sample

The primary implementation (Objective 1) outcome, along with secondary implementation measures, will be analyzed among the full sample of eligible patients defined above.

3.3 Objective 2 Primary Study Sample ("Pre-randomization Sample")

The primary effectiveness (Objective 2) outcome, along with secondary effectiveness measures, will be analyzed in the subset of patients from the full sample of eligible patients who had (1) an eligible PC visit at a PROUD trial PC clinic during the pre-randomization period, (2) a documented OUD diagnosis during the pre-randomization period, and (3) who were age ≥16 as of the start of the baseline period (two years pre-randomization). This latter criterion ensures that the patient has at least 2 years pre-randomization in which to have an eligible PC visit (in which they are ≥16 years of age). Clinic assignment for the Objective 2 primary sample will be the same as above. Because study sample is defined using pre-randomization data, this sample is also referred to as the Objective 2 pre-randomization sample.

3.4 Objective 2 Secondary Study Sample

The Objective 2 primary sample as defined in the preceding section, which restricts analyses to patients seen in PROUD trial clinics with documented OUD pre-randomization, was selected for primary analyses to avoid the potential for *identification bias*[3] (type of selection bias that can occur when the analytic sample is defined using post-randomization data). However, an important limitation of analyses within the Objective 2 primary sample is that they miss any impact of the PROUD intervention among patients with a new OUD diagnosis post-randomization, including patients who initiate treatment in PROUD intervention clinics who are diagnosed with OUDs after randomization, and patients who are newly attracted to the clinic specifically because of the PROUD intervention. To address this limitation, secondary analyses are conducted in a broader sample that includes patients new to the PROUD trial clinics or with newly recognized OUDs post-randomization (see Figure Shell 1, Objective 2 paper, Section 10.2).

Specifically, the effectiveness (Objective 2) secondary study sample is defined as the subset of patients from the main trial sample who had a documented OUD diagnosis during either the pre-

randomization or follow-up periods and who were age ≥16 as of the start of the baseline period. Clinic assignment for the Objective 2 secondary sample will be the same as above.

3.4.1 Objective 2 "Post-randomization Sample"

We note that the Objective 2 secondary study sample is a broader sample that includes the Objective 2 primary "pre-randomization" study sample as a subset (Figure Shell 1, Objective 2 paper, Section 10.2). We refer to patients who are included in the secondary study sample but who are not included in the primary pre-randomization sample as having "newly recognized OUD in trial clinics post-randomization," and we refer to this sample of patients as the "post-randomization" Objective 2 sample. This post-randomization sample consists of patients who were new to the trial clinics or with newly documented OUD post-randomization (see Supplemental Figure Shell 1**,** Objective 2 paper, Section 10.2, which illustrates the different ways patients can enter the post-randomization sample).

4.0 OUTCOME MEASURES

4.1 Definition of Primary Implementation Outcome Measure

The primary outcome (clinic-level measure) is the number of patient days of OUD medication treatment documented in the EHR in each clinic post-randomization. To account for varying clinic sizes, the clinic-level outcome is divided by the number of patients seen in the clinic post-randomization and then multiplied by an appropriate scaling factor in order to report the results (e.g., multiplying by 10,000 to calculate the number of patient days of OUD treatment per 10,000 patients), and reported as patient-years of treatment provided by a clinic (calculated by simply dividing days of OUD treatment by 365).

"OUD medication treatment" includes medications for OUD that can be prescribed in PC and documented in EHRs—buprenorphine formulations indicated for OUD (oral, implants, sustained release injection with or without an OUD diagnosis) or injectable extended release (XR) naltrexone with a diagnosis of OUD. An OUD diagnosis is required for injectable XR naltrexone because it is often used for alcohol use disorders (AUD). We do not require an OUD diagnosis for buprenorphine because there are buprenorphine formulations specific for OUD and PROUD Phase 1 analyses revealed OUD diagnoses are often missing, consistent with the literature.[4] Further, an OUD diagnosis is likely to be documented by PROUD clinic providers based on the PROUD NCM manual, so that requiring an OUD diagnosis could bias findings toward favoring the intervention clinics.

4.1.1 Algorithm for Calculating the Primary Implementation Outcome Measure

Two FDA-approved treatments for OUDs that can be provided in medical settings and documented in EHRs are included as OUD treatment. Buprenorphine and injectable XR naltrexone use will be determined from medication orders (EHR data) and procedures (EHR for all health systems and claims data for 2 health systems). Text string searches on generic and brand medication name will be used to ascertain buprenorphine formulations indicated for OUD treatment (sublingual, tablet, film, subdermal implants, and subcutaneous) and injectable XR naltrexone use from medication orders. A clinical co-investigator then reviews all hits on the text string search as part of quality control. Procedure codes will also be used to identify injectable formulations, implants, and oral formulations given in the office setting. Pharmacy dispensings are the gold standard for outpatient non-injectable medication use but they are not comprehensively available for 5 of the 6 health systems (whereas medication orders are routinely available in all 6 health systems). These 4 health systems are not insurers and therefore do not receive claims from outside pharmacies. Site (health system) principal investigators state that patients frequently obtain medications from pharmacies outside of the health system.

A single medication order or procedure code for buprenorphine formulations indicated for OUD treatment or injectable XR naltrexone (the latter with an OUD diagnosis) will be considered OUD treatment, though sensitivity analyses will examine this assumption (Section 5.2.4)

The algorithm for calculating the primary outcome measure of days on OUD medication (buprenorphine or injectable XR naltrexone) is as follows with details in Appendix A.

Ascertained from the EHR and claims as specified above, all buprenorphine and naltrexone XR injections will be quality checked, combined, cleaned and values imputed (as necessary), and collapsed into an episode or episodes (if gaps in use) of OUD treatment.

Episodes of OUD treatment will then be summed to calculate the total patient days covered with OUD treatment in the time-period of interest. Details on estimating episodes of treatment and summing episodes are in Appendix A but briefly they will be estimated as follows from variables

commonly available in the EHR – medication orders (i.e., medication name, date ordered, form, quantity, strength, strength unit, number of refills, and directions for use [SIG]) and procedures (medication name, form, strength, strength unit, and date administered).

Specifically, we will do the following:

1) Perform quality checks for outlier/implausible values for variables required to calculate the outcome and impute outliers/implausible values and missing values. See Appendix A for further details.
2) Estimate days' supply for orders of buprenorphine film and sublingual tablets.
    a. Translate directions for use (SIG) into pills per day (PPD). For example, "Take 1 pill twice a day" is a PPD=2.
    b. Divide the quantity field by PPD to estimate days' supply. Add any additional day's supply from refills on the original order to estimate the total days' supply of the order. For example, an order with quantity=60, PPD=2, and refills=2 would have a day's supply=90.
3) Assign naltrexone injection and buprenorphine subcutaneous Sublocade a day's supply=28.
4) Assign Probuphine (implant) a day's supply=180. Assign Brixadi a day's supply of 7 or 28 depending on the product.
5) Estimate runout dates (date when the medication supply provided runs out) as order/procedure date plus days' supply. An order (or dispensing in secondary measures for sensitivity analyses) on 1/1/2018 with a day's supply = 15 is estimated to runout on 1/15/2018.
6) Prior to dropping injectable XR naltrexone orders or procedures, we will describe patients with both OUD/OD and AUD diagnosis codes and adjudicate if needed
    a. For patients with 1+ injectable naltrexone orders or procedures and both 2+ OUD/OD and 1+ AUD diagnoses, provide a table of the numbers of OUD/OD and AUD diagnoses in 2 years pre-randomization and a similar table for post-randomization users of naltrexone.
    b. Adjudicate subjects with both 2+ OUD/OD and 1+ AUD diagnosis codes to decide whether to include or exclude as OUD treatment, using the number of times a code was used from 7a.
    c. Drop remaining injectable XR naltrexone if OUD diagnosis criteria are not met:
        i. Drop injectable XR naltrexone in the pre randomization period if there are not 2+ visits with an OUD/opioid overdose (OD) diagnosis (can be 1 OUD and 1 OD code) in the pre randomization period.
        ii. Drop naltrexone injections in the post randomization period if there are not 2+ visits with an OUD/OD diagnoses (can be 1 OUD and 1 OD code) in the pre or post randomization period.
7) In the pre and post-randomization periods, create continuous use episodes defined as OUD treatments with gaps ≤7 days between the runout of one order/procedure and the start date of the subsequent (other cut points to be evaluated in sensitivity analyses after summarizing the distribution of gaps). For example, two OUD treatment episodes with treatment episode 1 start 1/1/2018 and runout 1/15/208 and treatment episode 2 starting 1/17/2018 and runout 2/5/2018 are rolled into one episode of continuous treatment (start

1/1/2018 and end 2/5/2018). In general, this gap applies to same medication orders/procedures and different medications.

   a. One exception is that gaps of ≤14 days between the runout of a buprenorphine order and subsequent naltrexone XR injections are allowed in defining continuous use episodes to account for required washout periods prior to naltrexone XR injection.

8) Pre and post randomization episodes are not allowed to overlap. They will be left and right censored accordingly. For example, a treatment episode of 90 days with 15 days in the post randomization period (includes randomization date) will be split into two episodes with one episode contributing 75 days of treatment pre-randomization and a second episode contributing 15 days of treatment post randomization.

9) Add the number of days of each continuous use episode to arrive at total days of OUD treatment in the distinct periods of pre-randomization and post-randomization. The algorithm does not allow for double counting of any overlapping orders or procedures.

4.2 Definition of Objective 2 Effectiveness Outcome Measure (Powered Secondary Outcome)

The primary effectiveness (Objective 2) endpoint is a person-level count measure of the number of days of acute care utilization in the follow-up period, among patients with an OUD diagnosis pre-randomization. This measure includes visits to urgent care clinics or emergency departments (EDs), as well as days hospitalized. Acute care utilization will be determined from the EMR and insurance claims data when available.

For hospitalizations, the number of inpatient days will be the number of days from admission to discharge, inclusive. For urgent care or ED (referred to collectively as “emergency care”), each unique date with a visit to an urgent care or ED will be counted as 1 day (even if the patient stays overnight), except in the uncommon event that an ED visit spanned 3 or more days (< 0.5% of ED records collected in Data Pull 4, including both randomized and non-recruited UPC clinics), in which case the encounter will be classified as a hospitalization (as patients may be boarded in an ED until a hospital bed becomes available and some are subsequently discharged from the ED if no hospital bed ever became available). If a patient is admitted from urgent care or an ED, the ED or urgent care day is not (double) counted when counting the number of days of acute care.

4.2.1 Addressing Data Anomalies and Outliers

In examining the hospitalization data during interim Data Pulls, a few data anomalies were discovered (e.g., extremely long hospitalizations). Records with an anomaly will first be manually reviewed by health system study staff and corrected (where possible) or handled as follows:

- Discharge dates before the admission date: We assume the dates were incorrectly reversed and switch them back (< 50 records in Data Pull 4 out of over 300,000), unless the resulting length of stay is an outlier (identified by comparing to the distribution in the full sample), in which case a basic imputation approach will be applied (defined below).
- Missing discharge date: most hospitalization records with a missing discharge date were embedded within another hospital record; after removing these and collapsing across any overlapping records, < 0.5% of the remaining ~275,000 hospitalization records (in Data Pull 4) had a missing discharge date. For these records length of stay will be imputed (details below).

In addition, hospital records with long inpatient stays (≥180 days) will be manually reviewed by the health system to determine whether there may have been a data error.

Imputation approach: For the rare scenarios with data anomalies, we will impute the median length of stay based on inpatient episodes among patients within the same health care system (HCS), stratified by whether the patient had any OUD diagnosis during the study period.

4.3 Definitions of Secondary Measures

Additional secondary measures will be analyzed to describe outcome measures reflecting processes of care, implementation, and effectiveness. The following table describes each of the measures that will be examined either descriptively, or through formal modeling, as described in the statistical analysis section below. Measures will be calculated broadly in the full study sample to facilitate analyses across different study samples. For example, we will calculate the number of days of OUD medication treatment for each patient, which will be summed across patients as part of the primary Objective 1 outcome and will also be analyzed at a patient-level among patients with an OUD diagnosis pre-randomization. Other implementation measures (e.g., number of patients treated) will be analyzed in both the full study sample of all patients with a visit, as well as in the subsample of these patients who had an OUD diagnosis pre-randomization.

Table 2: Primary, Secondary, and Other Outcomes Measured During Post-Randomization Period Based on EHR Data (*measures from clinicaltrials.gov)

Outcome Measures
***Objective 1. Patient-days of OUD medication treatment (primary outcome).** Clinic-level number of patient-days of OUD treatment with buprenorphine and XR-NTX documented in the EHR during the period from randomization until two years after, reported per 10,000 PC patients in the clinic in the two years post-randomization.
***Objective 2. Acute care utilization (secondary outcome).** Patient-level number of days of acute care utilization during the period from randomization until two years after, among patients with an OUD diagnosis documented in the EHR in the three years prior to randomization.
Other Outcome Measures of Implementation
***Newly diagnosed OUD (Implementation Reach).** Clinic-level number of patients with a new International Classification of Disease (ICD) code for OUD documented in the EHR during the period from randomization until two years after who did not have an OUD diagnosis documented in the EHR in the three years prior to randomization, reported per 10,000 patients in the PC clinic in the two years post-randomization. This measure will also be estimated for any OUD diagnosis (new and previous diagnoses).
***Initiation[§] of OUD treatment (Implementation Reach).** Clinic-level number of patients who initiate: (1) buprenorphine or (2) XR-NTX with an indication of OUD as documented in the EHR during the period from randomization until two years after, reported per 10,000 PC patients in the clinic in the two years post-randomization. This measure will also be estimated for any OUD treatment (initiation and on-going treatment). [§]*Initiation of buprenorphine and* XR-NTX *in the context of PROUD trial outcome measures refers to the first order for OUD medication treatment post-randomization with no treatment with these medications in the prior 365 days (including pre-randomization). See Section 4.1.1. for operationalization of OUD in this context.*

Buprenorphine daily dose of 16 mg or more (Implementation Reach). Clinic-level number of patients on ≥ 16 mg per day of buprenorphine at any time during the period from randomization until two years after randomization as documented in the EHR, reported per 10,000 PC patients in the clinic in the two years post-randomization.

Retention measures of OUD treatment (Implementation Fidelity).

***Retention in OUD treatment**. Clinic-level number of patients initiating§ OUD treatment during the period from randomization until two years after randomization as documented in the EHR, who also receive OUD treatment on 80% of days available after initiation, reported per 10,000 PC patients in the clinic in the two years post-randomization [5].

Retention in OUD treatment for ≥ 6-months. Clinic-level number of patients initiating§ OUD treatment during the period from randomization until two years after randomization as documented in the EHR, who remain on treatment for ≥ 6 months after initiation, reported per 10,000 PC patients in the clinic in the two years post-randomization.***

Discontinuation of OUD treatment. Clinic-level number of patients initiating§ OUD treatment during the period from randomization until two years after randomization as documented in the EHR, who discontinue treatment (defined as a gap of 60+ days), reported per 10,000 PC patients in the clinic in the two years post-randomization.

§See definition of initiation above;

****Excludes patients whose only eligible PC visit occurred in the last 6 months of the post-randomization period or who began OUD treatment in the last 6-months of the post-randomization period. Measures of retention above will also be estimated for all subjects regardless of new or on-going treatment and for subjects with on-going treatment.*

Number of buprenorphine prescribers (Implementation Fidelity). Clinic-level number of buprenorphine prescribers† during the period from randomization until two years after randomization as documented in the EHR, who prescribe buprenorphine, reported per 10,000 PC patients in the clinic in the two years post-randomization.

Number of buprenorphine prescribers will also be reported per X total prescribers in the clinic in the two years post-randomization.

†Prescribers determined from medication orders in the electronic health records. Providers assigned to clinics based on number of visits with patients in the clinic pre-randomization.

***Naloxone prescribing (Implementation Fidelity).** Patient-level number of prescriptions of naloxone for overdose management in the period from randomization until two years after, among patients with an OUD diagnosis in the three years prior to randomization.

OUD treatment duration (Implementation Fidelity). Patient-level number of days of OUD treatment with buprenorphine and XR-NTX documented in the EHR during the period from randomization until two years after, among patients with an OUD diagnosis in the three years prior to randomization. This measure will be modeled as a categorical variable (0 days, 1-30 days, 31-90 days, 91-180 days, ≥180 days).

Other Outcome Measures of Effectiveness

Urgent care or ED use***.** Patient-level number of visits to urgent care or EDs during the period from randomization until two years after, among patients with an OUD diagnosis documented in the EHR in the 3 years prior to randomization. Urgent care and ED are combined into a single outcome to represent “emergency care”.

This measure will secondarily be modeled as a categorical variable (with cut-points based on the empirical distribution).

*** ED or Urgent Care visits that lead to hospitalization are classified as inpatient.*

***Inpatient days hospitalized**[**]**.** Patient-level number of days hospitalized during the period from randomization until two years after, among patients with an OUD diagnosis documented in the EHR in the three years prior to randomization.
Any acute care. Patient-level binary indicator for whether the patient had any acute care utilization during the period from randomization until two years after, among patients with an OUD diagnosis documented in the EHR in the three years prior to randomization.
Number of hospitalizations[**]**.** Patient-level number of hospitalizations during the period from randomization until two years after, among patients with an OUD diagnosis documented in the EHR in the three years prior to randomization. This measure will be modeled as a categorical variable (with cut-points based on the empirical distribution).

5.0 ANALYSES OF OUTCOME MEASURES

5.1 Primary Analytic Method for Primary Implementation Outcome Measure

Since randomization and the intervention occur at the clinic-level, the unit of analysis will be clinic and not patient for the primary outcome. The mixed effects model[6] evaluating the effect of the PROUD intervention is:

$$y_{ij} = \alpha + \beta * trt_{ij} + \gamma * z_{ij} + \theta_j + \epsilon_{ij} \quad (1)$$

where

1) y_{ij} is the observed value of the primary outcome measure for clinic i at health care system (HCS) j
2) trt_{ij} is the treatment indicator (PROUD intervention) for clinic i at HCS j
3) z_{ij} is the observed value of the primary outcome measure for the two years prior to randomization (hereafter baseline) for clinic i at HCS j
4) θ_j is the random effect for HCS j and is distributed $N(0, \sigma_\theta^2)$
5) ϵ_{ij} is the error term for clinic i at HCS j and is distributed $N(0, \sigma_\epsilon^2)$

Our primary hypothesis is that there will be a significant increase in the number of patient days of medication treatment for OUDs during the follow-up period in clinics randomized to the PROUD intervention as compared to clinics randomized to UPC. To evaluate this hypothesis, analyses test whether β (PROUD intervention effect) significantly exceeds zero using a one-sided hypothesis test at the 0.05 level. This is appropriate because our primary aim is to test superiority of implementation of the MA model relative to Usual PC in order to inform health systems' decisions as to whether to implement this model of OUD care.

Example SAS code for the analysis is as follows:

```
proc mixed data = primout;
  class arm;
  model scaletotdays = arm basevalue / solution;
  random intercept / subject = site;
  estimate "Treatment Effect" arm 1 -1;
run;
```

where "arm" is the treatment assignment (trt_{ij}), "scaletotdays" is the value of the primary outcome measure (y_{ij}), "basevalue" is the value of the primary outcome measure for the two years prior to randomization (z_{ij}) and "site" indexes the different HCSs (j).

This model allows for clinics to be correlated within a HCS, in addition to allowing for an association of the scaled days of OUD treatment prior to randomization with the post-randomization outcome.

5.2 Secondary Analyses of Primary Objective 1 Outcome Measure

5.2.1 Adjustment for Covariates

A secondary analysis of the primary Objective 1 outcome measure will adjust for additional covariates that are associated with the primary outcome. This secondary analysis will be the same as the primary analysis (i.e., it will still adjust for the baseline value of the outcome and include HCS-specific random intercepts), except that it will include additional covariates. The primary goal of this secondary analysis is as a sensitivity analysis, in which we will investigate the degree to which the treatment effect estimate (β) from the primary analysis changes after covariate adjustment; thus, we plan focus on the magnitude of the change in the estimate rather than conduct inference on the treatment effect estimated under this secondary analysis.

The following approach will be utilized to identify the covariates for potential inclusion in the secondary analytic model of the primary outcome measure. First, we will consider a candidate set of clinic-level covariates (described below). Because of our small sample size (n=12) for the clinic-level primary outcome, we are limited in the number of covariates we can include in the outcome model to avoid overfitting, which can lead to unstable coefficient estimates. Consequently, we plan to conduct baseline analyses (detailed below) to identify a subset of these predictors to adjust for. The goal of these baseline analyses is to identify which covariates are associated with the clinic-level number of patient years of OUD treatment during a 1-year period pre-randomization, after adjusting for this measure in prior baseline years (to mimic the primary analysis but using only baseline data). The planned baseline analyses will proceed as follows.

1) Consider the pre-randomization data from the two years prior to randomization (Period 1 = two years prior to randomization; Period 2 = year prior to randomization). Denote B1 as the "baseline" value of the outcome measure in Period 1 and B2 as the "baseline" value in Period 2. Obtain values of each of the clinic-level covariates being considered during the earlier baseline period (Period 1), such as the proportion of patients seen in the clinic during Period 1 who were female.

2) Conduct a regularized (lasso) regression of the B2 values on the B1 values as well as all of the candidate clinic-level covariates for Period 1 (earlier baseline period). Lasso regression avoids overfitting by penalizing the magnitude of the coefficients such that some of the coefficients may be shrunk to exactly zero.[7,8] The B1 value of the outcome will be "forced" into the model (i.e., the corresponding coefficient will not be penalized). We will use cross-validation initially to select the best-fitting value of the tuning parameter λ (which controls the number of non-zero coefficients[7]). Specifically, we will perform a "leave-one-cluster-out" cross-validation as follows. For each candidate value of the tuning parameter λ, we predict the B2 values of each HCS using a regularized mixed-effect model fitted with the data from all other HCSs. We then compute the mean squared error (MSE) of the prediction, and take the average of these MSEs across all HCSs ("the cross-validated MSE"), as the measure of the performance of the candidate value of the tuning parameter λ. We select the candidate value of the tuning parameter λ with the smallest cross-validated MSE.

3) If the selected value of the tuning parameter λ is such that there are more than 2 non-zero coefficients, we plan to vary the value of λ such that no more than 2 covariates have non-zero coefficients. These 2 covariates will be the ones that we include in the covariate-adjusted analysis (except that the time window for the covariates included in the outcome model will be over the full 2-year baseline period, rather than just over Period 1).

With 12 clinics and 4 variables (intervention indicator, baseline value of the outcome, plus 2 covariates), this proposed covariate-adjusted analysis would have 3 observations (clinics) per variable. Although there is no firm rule on the maximal number of variables able to be included in the model, recent literature has suggested that having as few as two subjects per variable in a standard linear regression model did not adversely impact parameter estimation.[9]

Candidate set of covariates: We identified a broad set of covariates based on known and hypothesized predictors of initiation in and retention to medication treatment for OUDs from the literature and discussions with study co-investigators. Each covariate will be defined at a patient level, and then aggregated into Period-specific (i.e., Period 1 and Period 2) clinic-level measures (e.g., mean age of patients seen in a clinic in Period 1 or 2, proportion of patients with ≥1 depression diagnosis) to be considered for the covariate-adjusted analysis of the clinic-level outcome.

1) Age[1,10-15]
    a. Continuous: mean age
    b. Categorical: proportion in each group (<25, 25-44, 45-64, ≥65)
    c. Binary: proportion <45
2) Gender:[11,15] proportion female
3) Race/ethnicity:[1,11,13,15,16] proportion in each of the following groups:
    a. Hispanic
    b. Non-Hispanic Asian
    c. Non-Hispanic Black or African American,
    d. Non-Hispanic White,
    e. Non-Hispanic Other race (any of the following, as well as each of the following groups separately):
        i. American Indian / Alaska Native
        ii. Native Hawaiian or Pacific Islander
        iii. Multiple race
        iv. Other race
    f. Missing/Unknown
4) Number of days with an OUD diagnosis: proportion of patients with 0, ≥1, ≥2, ≥5, ≥10
5) Number of days with a diagnosis of each of the following conditions (using clinic-level proportions as for OUD above):[15]
    a. Depression
    b. Anxiety
    c. PTSD
    d. ADHD
    e. Other mental health conditions[17]
    f. Any mental health condition (a-e above)
    g. Alcohol use disorder
    h. Cannabis use disorder
    i. Stimulant use disorder
    j. Other (non-opioid) substance use disorder
    k. Any substance use disorder (g-j above)
        i. Also including OUD
        ii. Excluding OUD
    l. HCV
6) Number of days with an ICD code for housing instability / homelessness (using clinic-level proportions as for OUD above)
7) Neighborhood-level SES measures, deciles across patients of:
    a. median household income[18]
    b. percent below the federal poverty line
    c. percent unemployed[1,15,19]
8) Buprenorphine prescribers
    a. number in the clinic
    b. as a percentage of all prescribers in the PC clinic
9) Rurality of residence (rural urban commuting codes of zip code): proportion in each group (urban, large rural city/town, small and isolated small rural town)
10) Quadratic term of baseline outcome

Covariates will be centered and standardized to have mean 0 and standard deviation (SD) of 1 so that they have the same scale.[20] We note that many of the above measures are expected to be highly correlated; however, the proposed lasso approach is able to select from among correlated measures, based on the predictive ability.

5.2.2 Modified Analyses Accounting for Success of Implementation

Our Objective 1 analysis considers days of OUD treatment over the entire follow-up period since randomization. However, it took time for HCSs to hire the NCM, and for the NCM to start seeing patients once hired, and it is possible some may have never succeeded in implementing the model. We therefore plan to conduct two modified analyses, which will apply the same statistical model as the primary analysis but will either modify the definition of the follow-up period over which the Objective 1 outcome measure will be calculated or will restrict which HCSs are included.

5.2.2.a. Analysis During the Period in Which the NCM was Seeing Patients

First, we will modify the time period over which the main outcome is calculated to only include periods after which the NCM has engaged a patient in OUD treatment in the intervention clinic. For example, if the intervention clinic of a HCS hired the NCM 16 weeks after randomization and s/he was trained 8 weeks later and s/he engaged the first patient 1 week after training, for that HCS we would define the time period from 23 weeks post randomization until the end of the follow-up period, for both the clinic randomized to the PROUD intervention and the clinic randomized to UPC.

5.2.2.b. Analysis Limited to Clinics that Successfully Implemented the MA Model

Second, we will further restrict the set of clinics analyzed by excluding HCS from the analysis if the clinic randomized to the intervention in that HCS did not successfully implement the MA model. Successful implementation is defined as the NCM at the intervention clinic at that HCS seeing at least 30 patients. The choice of 30 patients was selected *a posteriori* as a level that several HCS had achieved and which was hypothesized as a tipping point for the clinic to be known as providing OUD treatment within the HCS. If more than 2 HCS are excluded, then we may not be able to fit the mixed-effect model; if that is the case then we will instead describe the outcome across intervention arms in these sets of clinics.

5.2.3 Sensitivity Analysis Addressing the Algorithm for Assigning Patients to Clinics

The primary algorithm for clinic assignment (see Section 3) prioritizes the clinic that the patient visited pre-randomization. If a patient visits the UPC clinic pre-randomization but then starts going to the PROUD clinic to receive care from the NCM, under this approach for clinic assignment any medication days of treatment will be “counted” toward the UPC clinic, which could bias the effect of PROUD toward the null. To address this issue, we plan to describe the number of patients with OUD (diagnosed any time during the study period) treated in the intervention clinic during follow-up who were assigned to the UPC clinic (see Section 3 above). If we observe that *ever* occurs at a health system, we will apply an alternate algorithm as a sensitivity analysis, in which we will assign patients to clinics based on the number of visits to the clinic during the post-randomization period.

5.2.4 Sensitivity Analyses Focused on Assumptions Made in Estimating Patient-Years of OUD Medication Treatment (Objective 1: Main Implementation Outcome)

While medication dispensings are only a proxy for ingestion of medications, they are the gold standard for ascertaining medication use with EHR and claims data. When evaluating patient days of OUD treatment, only medication orders are available at most PROUD health systems.

Pharmacy dispensings are complete at only one health system. Medication orders suffer from the same limitation of no assurance the medication was ingested plus medication orders could have been written and never dispensed. We expect that this is uncommon as preliminary data from Phase I at the health system with complete dispensings indicated 98% of patients with a buprenorphine order have a dispensing within 60 days of the order). However, medication dispensing data includes dispensed refills, but orders only include an indication that there can be a refill. Preliminary data from Phase I at this health system indicate approximately 14% of buprenorphine orders contain at least 1 refill on the order.

To address limitations of medication orders, we plan to conduct sensitivity analyses where we replicate the objective 1 main outcome analysis using different assumptions for calculating days of OUD medication treatment to see if there are changes in results and conclusion. Specifically, the sensitivity analyses below include more restrictive objective 1 outcome analyses with fewer assumptions about medication use and refills. In addition, we propose sensitivity analyses to address assumptions made in smoothing gaps between medication orders and/or procedures, a sensitivity analysis to add methadone maintenance therapy from the 1-2 health systems that have such data, and an analysis using dispensing data where available for at least some patients in both trial clinics within a health system.

5.2.4.a. Omitting Last Order in an Oral Buprenorphine Treatment Episode

To address the potential limitations of medication orders, we will conduct sensitivity analyses with the objective 1 main outcome restricted to episodes of treatment with at least two oral buprenorphine orders if the patient has only oral buprenorphine orders (majority of patients), with the last order in the episode omitted in case it was not picked up and taken by the patient. This sensitivity analysis also addresses any short-term buprenorphine therapy to assist with opioid tapers. This does not impact oral buprenorphine captured by procedure codes when given in the office or any injectable buprenorphine, buprenorphine implants, or injectable XR naltrexone. We will also describe the number of treatment days that are excluded in this restrictive analysis by intervention arm and health system. When the last order is omitted, all refills after the last order are also omitted.

5.2.4.b. Vary the Allowable 7-day Gap to Define Continuous Use Episodes

To address the limitation that we smooth small gaps of ≤ 7 day between medication orders and/or procedures to arrive at continuous episodes of OUD medication use, we will vary the allowable ≤ 7 day gap to smaller and larger values (e.g., 2 days, 5 days, 10 days based on descriptive data) when estimating the objective 1 main outcome in this sensitivity analysis. For example, the main analysis treats a 10 day order for buprenorphine on 1/1/2019 and a 10 day order for buprenorphine on 1/15/2019 as one continuous episode lasting from 1/1/2019 to 1/24/2019 (24 days) because the gap between when the first order ran out and the second order began is < 7 days. Altering the allowable gap to only 2 days would result in two different episodes (1/1/2019 – 1/10/2019 and 1/15/2019-1/24/2019) that contribute 20 days of treatment.

5.2.4.c. Incorporating Dispensings Data to Define Days of OUD Treatment

For the health systems with outpatient pharmacies in both intervention and usual care clinics (including one health system that has complete dispensing data), we will estimate days of OUD treatment in this sensitivity analysis using a combination of medication orders and dispensing data. For each dispensing, if its dispense date (rxdate) is in “the range” of an order date or a refill date, then it is considered to be linked to that order or refill and dropped in calculations; otherwise, this dispensing is considered as not linked to any order and added to the calculation of days covered. The range is defined as dispensed within 30 days after the order/refill’s start date (an

order's start date is the order's date; a refill's start date is derived from the original order's date and days supply).

To characterize the potential for incomplete ascertainment of OUD medication treatment, we will describe 1) what proportion of patients have a buprenorphine order but not a dispensing for buprenorphine within 30 days of the order and 2) what proportion have a dispensing for buprenorphine but do not have a buprenorphine order 30 days prior to the dispensing. We will vary the 30-day window between order and dispensing (e.g., 60 days), and report by period (e.g., baseline and post-randomization) and by trial arm.

5.2.4.d. Incorporating Limited Data on Methadone Maintenance Therapy (MMT)

We will describe the availability of MMT by health system. For health systems with methadone maintenance outpatient treatment programs (OTPs) data available (one HCS and possibly a second), we will incorporate procedure codes for Methadone Maintenance OTPs when estimating the objective 1 main implementation outcome in this analysis. Based on consultation with addiction medicine specialists, we will require at least 2 procedure codes to be included as using Methadone Maintenance from an OTP and assume that treatment lasts for the time period between codes for a maximum of 365 days per code (e.g., 2 codes 365 days apart). The end date of MMT (i.e., last claim) will estimated at half the average days between previous MMT codes.

5.2.4.e. Combined Sensitivity Analysis Using All Optimal Data

We will conduct a sensitivity analysis of the objective 1 main implementation outcome using both MMT data (1-2 health systems) and dispensing data as they are detailed above. This will be done across the 1-2 health systems that have both MMT and dispensing data.

5.2.5 Timing of OUD Diagnosis

Our primary analysis includes all days of OUD treatment for buprenorphine and injectable naltrexone, including among patients with and without an OUD diagnosis prior to randomization, and among patients who were seen in the clinic previously as well as patients who were new to the clinic or to the health system entirely. If we find that the PROUD intervention increases the provision of OUD treatment based on our primary analysis, we will further explore which of the following mechanisms may have contributed to this increase: (1) by increasing the number of days treated among individuals in the clinic pre-randomization who also had a documented OUD diagnosis pre-randomization, (2) by increasing the number of days treated among individuals in the clinic pre-randomization but who did not have a documented OUD diagnosis pre-randomization, or (3) by attracting patients post-randomization not previously seen in the clinic who have OUD (from within or outside the HCS; Supplemental Figure Shell 1, Objective 2, Section 10.2). To address each of these possibilities, we plan to conduct secondary analyses of the Objective 1 outcome among each of these 3 subsets overall by arm (as well as conducting descriptive analyses by health system). To facilitate comparison of days treated across these different study populations, we will apply the same scaling factor as in the definition of the primary analysis of the outcome (i.e., the number of patient-days of OUD treatment will be divided by the total number of patients seen in the clinic over the follow-up period).

5.2.6 Secondary Analyses Among Patients Treated for OUD

Our primary outcome is a scaled clinic-level measure of OUD treatment among all patients assigned to the clinic, scaled by the number of patients seen in the clinic, over the 2-year post-randomization period. In secondary analyses, we plan to repeat the primary analysis where we calculate the outcome measure among patients with documented OUD treatment during the 2 years post-randomization (and scale by the number of patients with OUD; see Table Shell 6, Objective 1, Section 10.1). However, we note that because the intervention could impact both the

denominator (number of patients diagnosed with OUD) as well as the numerator (initiation and duration of treatment among treated patients), interpretation of these secondary analyses should be interpreted cautiously.

5.3 Primary Analytic Method for Objective 2 Primary Effectiveness Outcome Measure

Our primary outcome for the main Objective 2 (effectiveness) outcome is a patient-level measure of the number of days of acute care utilization over the follow-up period, which is a 2-year period for 5 of the 6 HCSs, and a 1.5-year period for the 6th HCS.

The primary analysis will be among individuals with an eligible PC visit pre-randomization who had an OUD diagnosis during the pre-randomization period (see definition of "Objective 2 primary study sample" in Section 3). The patient-level analysis will follow ITT principles, with patients analyzed according to the randomization group of the clinic to which they were assigned pre-randomization, regardless of the degree to which the clinic actually implemented the intervention, and regardless of whether the patient was actually treated by the NCM or seen post randomization.

We hypothesize that, among patients who had an eligible PC visit and were identified as having an OUD diagnosis pre-randomization (documented in their EHRs up to 3 years prior to randomization), individuals from a PROUD intervention clinic will have decreased acute care utilization after randomization as compared to individuals from a UPC clinic. We plan to fit a mixed-effect Poisson regression model (with log link) at the patient level to the number of days of acute care utilization. The model will account for clustering of patients within a clinic by including clinic-specific random intercepts. Specifically, the regression model will be of the following form:

$$log\left[E\left(y_{ijk}\right)\right] = \log\left(\text{days}_\text{j}\right) + \alpha + \beta * trt_{ij} + \gamma * z_{ijk} + \theta_{ij}$$

where

- y_{ijk} is the observed number of days of acute care utilization of patient k in clinic i of HCS j over the follow-up period
- trt_{ij} is the treatment indicator (PROUD intervention) for clinic i in HCS j
- z_{ijk} is the baseline value of the outcome during the two years prior to randomization for patient k in clinic i of HCS j (γ is the corresponding coefficient)
- θ_{ij} is the random effect for clinic i in HCS j

Because one of the HCSs was randomized at a later date than the others, not all HCSs have the same amount of follow-up time. To account for this difference, we will include in the model an offset term for the number of days of potential follow-up time (e.g., 2 years for 5 of the HCSs and 1.5 years for the 6th HCS); in the above model, $\log\left(\text{days}_\text{j}\right)$ denotes the offset term where days_j denotes the days of follow up in HCS j.

We will evaluate our primary Objective 2 hypothesis by testing the null hypothesis $H_0: \beta = 0$ versus the two-sided alternative hypothesis that β is non-zero with a type 1 error rate of 0.05. In cluster-randomized trials with a small number of clusters, a small-sample correction is often necessary to obtain correct type 1 error rates.[21,22] Although small-sample degree of freedom (DF) correction methods have been evaluated for continuous and binary outcomes under generalized linear mixed-effect models (GLMMs), there has not been prior guidance of which approach to use in our setting with a count outcome and covariate adjustment. Given this lack of knowledge, members of the PROUD statistical team are conducting simulations to evaluate alternate small-sample DF methods in this setting. Our results so far suggest that the optimal choice of testing procedure varies depending on the data-generating scenario in terms of the intraclass correlation

coefficient (ICC), number of clusters, and sample size within a cluster; however, using the likelihood ratio test with the Between-Within (BW) DF correction method (also referred to as "inner-outer" approach) appears to perform generally well in the scenario that aligned most closely with PROUD Phase 1 data used in the power simulation (see Appendix C).

Inclusion of a HCS-specific random effect. The above model accounts for clustering of patients within clinics but does not account for the possibility of additional correlation of outcomes from patients within the same HCS beyond any within-clinic correlation. The reason for this is because analyses of Phase 1 data suggested that within-clinic correlation was considerably larger than within-HCS correlation (the random-effect variance was 0.00016 for HCS versus 0.055 for clinic from a model including random effects for both). Although including a random-effect term that is not truly needed should not impact parameter estimates, including a HCS-level random effect leads to inferential challenges. In particular, as above there has not been guidance on applying small-sample DF correction methods when there are random-effects at two levels (here 12 clinics nested within 6 HCSs). We therefore chose the clustering level that explained most of the correlation for primary Objective 2 analyses. As a sensitivity analysis, we plan to examine the impact of additionally including a HCS-level random effect on the treatment effect parameter and standard error estimate.

5.4 Secondary Analyses of Objective 2 Primary Effectiveness Outcome Measure within the Primary Objective 2 Sample

5.4.1 Objective 2 Covariate-adjusted Model

We will apply a sensitivity analysis that includes additional covariate adjustment beyond the number of days of acute care at baseline. We plan to additionally adjust for the following pre-specified covariates, identified based on known and hypothesized predictors of acute care utilization among patients with OUDs identified from the literature and from discussions with study co-investigators:

1) age (at randomization, including linear and quadratic terms)
2) gender (F/M)
3) race/ethnicity (Hispanic, Asian, Black or African American, White, Other, Unknown)
4) Neighborhood-level measures capturing socioeconomic status (obtained from census data linked via zip code [using most recent zip code available pre-randomization]):
    a. median Neighborhood household income
    b. percent Neighborhood below the federal poverty line
    c. percent Neighborhood unemployed
5) Insurance status (binary indicators for the patient's type of insurance [Medicaid, Medicare, other insurance, uninsured], using the most recent available known value pre-randomization)[23]
6) Days of OUD medication treatment (in the two years pre-randomization)[24]
7) Number of days with OUD documented
8) Comorbidity (yes/no, in the two years pre-randomization)
    a. Alcohol use disorders[25]
    b. Other (non-opioid) substance use disorder
    c. Schizophrenia and other psychotic disorders
    d. Weighted summary score of other comorbidities (Elixhauser index[26], pulling out a-c above).[25]
9. Housing instability, including homelessness (indicator for having a V or Z code in the two years pre-randomization)[23]

Note that some of these covariates (e.g., race/ethnicity; insurance status) have some missing values; the approach to handling missing data in adjustment variables is described below in Section 8.1.

5.4.2 Objective 2 Sensitivity Analysis Among "Active" Patients

Because we have a visit-based sample, we do not know when a patient is no longer observable (e.g., switched clinics, moved, went elsewhere for medical or OUD treatment, or died), except among enrolled patients in HCSs with insurance claims (discussed in the following section). In particular, we cannot distinguish between a patient who has no acute care utilization over the follow-up period and a patient who has acute care utilization outside of the health system (that is not captured across all HCSs in our data). To address this, we plan to repeat the primary analysis among the subset of patients who are "active in the HCS", defined as having any evidence that they are still observable (e.g., any visit (PC and non-PC), or diagnosis anywhere in the health system during follow-up). In addition, we plan to conduct a second sensitivity analysis among patients "active in PC", defined as having a PC visit to their "assigned" trial clinic during the follow-up period.

5.4.3 Objective 2 Sensitivity Analysis Among Patients Enrolled in the HCS Health Plan

Two of the HCSs are integrated health systems that insure a subset of the patients who receive clinical care in the health system. We will examine the proportion of patients with an OUD diagnosis at baseline who are in the insured sample and who were enrolled in the health plan at the time of randomization. Descriptively, we will compare the estimated outcome rate separately within each of these two HCS in a model that accounts for enrollment as compared to the main analysis approach that does not account for enrollment (e.g., by comparing the mean number of days of acute care utilization per year enrolled vs. the mean number of days of acute care utilization per year of the 2-year HCS follow-up period).

In addition, depending on the size of the subsample, we will consider repeating the primary Obj. 2 analysis among this subsample. The offset term will be slightly modified from the primary analysis: rather than use the amount of possible follow-up from the randomization date, we will use more precise information on the number of days the patient was continuously enrolled in the health plan post-randomization during the available follow-up period (allowing gaps of up to 90 days). Although this proposed analysis is only available on a subset of the Objective 2 primary sample in the 2 HCSs, by restricting to enrolled patients and accounting for follow-up time in which the patient was enrolled (such that their outcome data is known to be observable), analyses within this sample could mitigate issues whereby outcomes of patients included in the primary Objective 2 visit-based sample may not be observable (e.g., if they left the health system or accessed acute care outside the system). Understanding differences in this analysis and HCS-level estimates for these sites will aid in interpretation of findings.

5.4.4 Objective 2 Sensitivity Analysis Addressing Clinic Assignment/crossover

We will follow a similar approach as the above Objective 1 sensitivity analysis (see Section 5.2.3 above). Specifically, if we observe that any patients assigned to a UPC clinic is ever treated in the intervention clinic at that HCS during the follow-up period, we will apply an alternate algorithm in which patients are assigned to the clinic where they have the most PC visits post-randomization.

5.4.5 Objective 2 Secondary Modified Analyses Accounting for Success of Implementation

We will conduct secondary modified analyses accounting for success of implementation following a similar approach as for Objective 1 (see Section 5.2.2 “Modified analyses accounting for success of implementation”), including (1) analyses restricted to the follow-up period in which the NCM was seeing patients (for which we will modify the offset term), and (2) analyses limited to HCS in which the clinic randomly assigned to the PROUD intervention successfully implemented the MA model. We also plan to repeat analysis (1) restricted to patients with an active PC visit post-randomization (Section 5.4.2).

5.4.6 Objective 2 Sensitivity Analysis Applying Stricter Eligibility Criteria

Our primary analysis requires just a single OUD diagnosis pre-randomization for a patient to be included in the sample, as well as just a PC visit at any time pre-randomization (up to 3 years). To account for the possibility that having only one diagnosis could reflect an error in identifying OUD and that patients with a PC visit early in the pre-randomization period may have left the clinic, we will conduct a sensitivity analysis among the sub-sample of our primary sample that restricts to patients with both ≥2 days with an OUD diagnosis during the pre-randomization period and a PC visit in the year prior to randomization.

5.5 Secondary Analyses of Primary Objective 2 Outcome Measure within the Objective 2 Secondary Sample

Here we describe secondary analyses of the primary Objective 2 outcome measure within the larger secondary sample (see Section 3.3) that includes patients in the primary Objective 2 sample (i.e., the pre-randomization sample with a PC visit and OUD diagnosis pre-randomization), as well as the post-randomization sample of patients with newly recognized OUD in the trial clinics post-randomization.

5.5.1 Rationale for Conducting Analyses Within a Secondary Sample

It is expected that some of the patients who initiate treatment in PROUD intervention clinics will have been diagnosed with OUDs after randomization, potentially due to the PROUD intervention. It is also possible that patients may be newly attracted to the clinic (or to the health system entirely) potentially because of the PROUD intervention (based on the fact that at least 77% of patients treated in MA were new to the clinic after implementation).[27] The primary analysis of Objective 2 described above would miss any impact of the PROUD intervention on both of these subsamples of patients, because they were not identified pre-randomization (because they were not previously diagnosed with OUDs or because they did not visit the clinic in the pre-randomization period). These secondary analyses within the secondary Objective 2 study sample, which includes patients seen in the trial clinics with an OUD diagnosis post-randomization only, are designed to capture these additional patients who may be affected by the PROUD intervention.

On the other hand, these secondary analyses must account for the fact that patients with newly recognized OUD in the PROUD intervention clinics post-randomization are likely to differ markedly from patients with newly recognized OUD in the UPC clinics post-randomization. Further, it is likely that these patients could differ in ways that may be associated with acute care utilization (e.g., patients could be referred for ongoing buprenorphine treatment from an ED or hospital that started treatment). To address this, analyses will adjust for covariates known or hypothesized to be associated with acute care utilization (based on prior literature), as well as secondarily for any additional covariate observed to differ across patients with newly recognized OUD post-randomization in the UPC versus PROUD intervention clinics.

5.5.2 Defining Covariates for the Objective 2 Secondary Sample That Includes Patients Not Seen in the Health System Pre-randomization

Defining covariates using pre-randomization data is preferred for randomized studies, because this ensures that the intervention does not affect the covariate values. However, this is not possible for patients who were not seen in the HCS pre-randomization. For these patients we therefore plan to use the post-randomization value of time-varying covariates that is available closest to randomization. We note that using this approach means that analyses including patients new to the HCS post-randomization cannot adjust for "baseline" acute care utilization (since patients new to the HCS post-randomization do not have baseline data observed). We considered applying a multiple imputation approach to address missing baseline data among these patients; however, we did not think the missing at random (MAR) assumption was reasonable (i.e., we did not think using data from patients in the HCS pre-randomization was sufficient to impute data from patients new to the HCS). Therefore, sensitivity analyses will be applied to examine the subpopulation excluding these patients (see Section 5.5.4).

Decisions regarding the set of covariates to be included in analyses within this secondary sample (listed below) were made in conjunction with the Investigator team to determine whether individual covariates measured post-randomization are likely to be impacted by the intervention. We wanted to avoid including covariates that are likely to be impacted by the intervention (as they may be on the causal pathway). Although documentation of certain comorbidities were thought to be most likely to be impacted by the intervention, it was hypothesized that documentation of any comorbidities could be increased via the intervention due to increased care in general; insurance status was also plausibly thought to be related to intervention status (as poor OUD-related outcomes could affect insurance status).

Covariate adjustment: We plan to adjust for the following subset of covariates as in the secondary analysis within the primary Objective 2 sample (described above in Section 5.4.1):

1) age (at randomization, including linear and quadratic terms)
2) gender (F/M)
3) race/ethnicity (Hispanic, Asian, Black or African American, White, Other, Unknown)
4) Neighborhood-level measures capturing socioeconomic status (obtained from census data linked via zip code [using zip code available closest to randomization, prioritizing pre-randomization if available]):
    a. median Neighborhood household income
    b. percent Neighborhood below the federal poverty line
    c. percent Neighborhood unemployed

In a sensitivity analysis, we will further adjust for additional covariates (including insurance status, comorbidities, and housing instability measures from Section 5.4.1) found to differ between individuals with newly recognized OUD post-randomization in the PROUD intervention clinics as compared to individuals with newly recognized OUD post-randomization in the UPC clinics (e.g., with a standardized mean difference of <0.10). However, because differences in these covariates (e.g., comorbidities) could be affected by the intervention, results from this sensitivity analysis will be interpreted with caution.

5.5.3 Primary Objective 2 Outcome Analysis within the Secondary Sample

We plan to fit a similar mixed-effect Poisson regression model as in the primary Objective 2 analysis but that (1) includes additional covariates and (2) that allows the treatment effect comparing the PROUD intervention clinics to UPC clinics to differ among patients with newly

recognized OUD in the clinics post-randomization vs. those with visits to the trial clinics and OUD documented pre-randomization. Specifically, the model will be of the following form:

$$log\left[E\left(y_{ijk}\right)\right] = \log\left(\text{days}_\text{j}\right) + \alpha_0 + \alpha_1 \text{period}_{ijk} + (\beta_0 + \beta_1 * \text{period}_{ijk}) * trt_{ij} + \gamma * z_{ijk} + \theta_{ij}$$

where *period* is an indicator for the period when the patient was first recognized with OUD in the trial clinics (i.e., whether the patient is in the post-randomization sample or pre-randomization sample; see Section 3.3), and the other terms are defined as in the primary analysis, except for the covariate vector (*z*), which includes the set of covariates listed above in Section 5.5.2.

We will evaluate our secondary Objective 2 hypothesis by testing the composite null hypothesis $H_0: \beta_0 = \beta_1 = 0$ versus the alternative hypothesis that at least one of β_0 or β_1 is non-zero by conducting a likelihood ratio test. Additionally, we will estimate the treatment effect separately among patients identified in the pre-randomization period (β_0) and patients identified in the post-randomization period ($\beta_0 + \beta_1$). The coefficient β_1 is the difference in the treatment effect comparing patients identified post-randomization to those identified in the pre-randomization period. This could either reflect a true difference in the treatment effect, or, more likely, it could reflect unmeasured confounders (not included in *z*) that differ between patients newly recognized with OUD post-randomization in the intervention versus UPC trial clinics. We do not have a specific hypothesis regarding β_1 because new patients may be attracted to the clinic to receive the PROUD intervention, as seen in Labelle,[1,27] and these patients may be sicker (or healthier) than patients identified pre-randomization, or more motivated for treatment, which could increase (or decrease) acute care utilization during the period after randomization. We are not powered to test for the difference in intervention effects between these two groups (β_1); rather, the analyses proposed here are exploratory analyses that will generate hypotheses for testing in future studies.

5.5.4 Sensitivity Analyses of Primary Objective 2 Outcome Measure within the Objective 2 Secondary Sample

5.5.4.a. Restricting to Patients in the HCS Pre-randomization

As discussed in Section 5.5.2, analyses in the full secondary sample that includes patients new to the HCS post-randomization are challenging due to the inability to adjust for baseline characteristics (including the baseline value of the outcome). To address this limitation, we plan to repeat the Objective 2 analysis in the secondary sample restricted to patients in the HCS pre-randomization (see Supplemental Figure Shell 1, Objective 2, Section 10.2) and using the full set of covariates as in the secondary analysis within the Objective 2 primary sample (described above in Section 5.4.1). Because differences in estimates in this sample could be due either to using a fuller set of covariate adjustment or using a different sample, we will additionally conduct this analysis using the same set of covariates as in Section 5.5.2 (age, gender, race/ethnicity, and neighborhood-level SES).

5.5.4.b. Secondary Analyses Accounting for Implementation and Other Sensitivity Analyses

We also plan to apply the same set of sensitivity analyses in the Objective 2 secondary sample as in the Objective 2 primary sample, including among “active” patients (in the HCS, in PC; Section 5.4.2), among enrolled patients (in 2 HCSs; Section 5.4.3), under the alternate clinic assignment algorithm (Section 5.4.4), under modified analyses accounting for success of implementation (Section 5.4.5), and among patients with ≥2 OUD diagnoses (Section 5.4.6) over the full study period.

5.6 Analyses of Secondary Outcome Measures

Analyses of pre-specified secondary clinic-level outcomes (Table 2, Section 4.3) will use the same analytic approach as for the primary Objective 1 outcome, and analyses of pre-specified

secondary patient-level outcomes will use the same analytic approach as for the primary Objective 2 outcome, using the appropriate link function (e.g. logit for binary measures). In addition, patient-level versions of the clinic-level implementation outcomes (e.g., number of days of OUD medication treatment) will be analyzed at a patient level among the Objective 2 sample (Objective 2 sample, see Shell Table 2; Section 10.2).

Note that unless otherwise specified, analyses will be adjusted for the baseline measure of the respective outcome being considered; exceptions are that binary or categorical outcomes derived from continuous measures will adjust for the continuous version of the measure (e.g., the binary variable of "Any acute care") and that the patient-level binary versions of the initiation and retention measures described in in Table 2 (Section 4.3) will adjust for the baseline number of days of OUD treatment.

5.7 Effect Modification and Subgroup Analyses

5.7.1 Objective 1 Primary Implementation Outcome

Given the NIH requirement to perform subgroup analyses of the primary (Objective 1) outcome on the basis of sex, race and ethnicity, and the importance of understanding how the MA Model performs in individuals < 26 years, we plan to conduct analyses of subgroups based on: age (< 26 vs older), where age is calculated at the eligible visit closest to randomization, prioritizing any pre-randomization visits; sex; and race/ethnicity (Hispanic, Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian, Other). We note that race and ethnicity are combined into a single variable, because in some HCS race is not documented if the patient reports ethnicity as Hispanic. An interaction term between demographic subgroup and treatment assignment can be used to evaluate whether the demographic factor moderates the implementation effect. Any such comparisons will likely be underpowered and must be interpreted with caution. The original Massachusetts studies observed that patients who were male or Black or Hispanic were less likely to engage in PC treatment of OUD with the MA Model, compared to female and white patients, respectively,[1,15] but no differences were observed across age groups. As a result, PROUD investigators hypothesize that the intervention will result in smaller increases in OUD medication treatment in patients who are male or non-Hispanic Black or Hispanic but hypothesize no differences across age groups.

For each subgroup of interest, we will define the Objective 1 outcome measure among individuals within that subgroup as the number of patient-days of OUD treatment among individuals in that subgroup who were seen in the clinic during the 2 years after randomization, scaled by the total number of patients seen in the clinic (within that subgroup) during that time period. We denote the outcome for clinic i at HCS j among subgroup g as y_{ijg}.

As a concrete example, here we write down the model for the effect modification analysis for age group. Let g be an indicator variable that takes the value 1 if a person is < 26 and takes the value 0 if a person is 26 years or older. We plan to adapt the model for the primary analysis as follows:

$$y_{ijg} = \alpha_0 + \alpha_1 g + \beta_0 trt_{ij} + \beta_1 trt_{ij} g + \gamma * z_{ijg} + \theta_j + \epsilon_{ijg}$$

where

- y_{ijg} is the age-group specific Objective 1 outcome measure for clinic i at HCS j (defined above),
- trt_{ij} is the treatment indicator (PROUD intervention) for clinic i at HCS j,
- z_{ijg} is the age-group specific Objective 1 outcome measure for the two years prior to randomization (hereafter baseline) for clinic i at HCS j,
- θ_j is the random effect for HCS j and is distributed $N(0, \sigma_\theta^2)$, and

- ϵ_{ijg} is the error term for clinic i at HCS j and is distributed $N(0, \sigma_{\epsilon}^2)$.

We will test whether the effect of the PROUD intervention differs across age groups by testing whether the term β_1 differs from zero using an F test. We will also estimate age-group specific intervention effects (e.g., β_0 for individuals of age 26 or older and $\beta_0 + \beta_1$ for individuals < 26). Subgroup-specific intervention effects will be tested by using a one-tailed test, consistent with the primary (i.e., un-stratified) analysis of the Objective 1 outcome. The same modeling approach will be applied for analyses of gender. If race/ethnicity is a categorical variable with more than 2 categories, we will evaluate whether the intervention effect differs across any of the racial/ethnic groups (i.e., if effect modification is present) with an omnibus (overall) test. In addition, per our hypotheses above, we will test for a difference in outcomes between non-Hispanic Black patients and Hispanic patients, each as compared to white patients. Again, as discussed above such comparisons will likely be underpowered and should be interpreted with caution.

5.7.2 Objective 2 Primary Effectiveness Outcome

We plan to conduct analyses of subgroups based on: age (< 26 vs older) at randomization; sex; race and ethnicity (Hispanic, Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian, Other), using similar subgroups as Objective 1. If one of the subgroups is small (fewer than 5% of patients in the sample), we will consider excluding that category or combining with the “Other” category. To examine whether the treatment effect may differ across groups, we plan to apply the same analytic approach as for the primary analysis of the Objective 2 outcome, in which we will additionally include an interaction term between the categorical subgroup variable of interest and the treatment effect term. As above, such comparisons are expected to be underpowered and will be interpreted cautiously. To aid in the interpretation of potential subgroup-specific effects, we plan to additionally conduct analyses of these subgroups for the patient-level treatment outcome of number of days of OUD medication treatment.

6.0 POWER CONSIDERATIONS

6.1 Power Simulations for Primary Objective 1 Outcome

The power calculations focus on the six HCS determined to be eligible during Phase 1 of CTN-0074, with each HCS contributing two clinics. One clinic from each HCS will be randomized to implement the PROUD intervention, while the other will continue with usual primary care (UPC). Simulations were conducted to calculate the power associated with various values of the treatment effect, which is parameterized as the mean of the primary outcome measure in the intervention clinics divided by the mean in the UPC clinics. Of ultimate interest in the calculations presented here is whether, with the health systems selected in Phase 1, there will be sufficient power (>80%) to detect at least a 5-fold increase in the number of patient days of OUD treatment (per 10,000 participants) associated with implementation of the PROUD intervention as compared to UPC. To accomplish this, we considered various values of the treatment effect and calculated the corresponding statistical power via simulation.

6.1.1 Data-generation and Analysis Model

In our power simulation, there were 6 HCSs and 2 clinics within each HCS (one clinic assigned to the intervention and one to UPC), making 12 clinics in all.

The outcome variable for each clinic is *Treated Days per Patient Seen,* which with actual data, we would calculate by dividing the total number of treated days at the clinic during 2 years of follow-up by the total number of unique patients seen by the clinic during that same time period. To evaluate power under an analysis approach that adjusts for the baseline value of the outcome, we used Phase 1 data to calculate an approximation of the magnitude of the association between the number of OUD-treated days per patient seen pre- and post-randomization. Using the analytic model specified in Equation (1) would require four years of data for this estimation, however only three years of data are available from Phase 1 (FYs 2014, 2015 and 2016). Thus, data were simulated in a recursive fashion using an autoregressive lag-1 approach.

Step 1: Estimate via regression the relationship between the number of OUD-treated days per patient seen in the last two years of Phase 1 data (2015-2016) and the same measure for the first two years of the Phase 1 data (2014-2015). Letting *j* index HCS and *i* index clinics within an HCS, this corresponds to regression of the outcome for 2015-2016 on 2014-2015 (see Section 6.1.2). That is, model

$$y_{ij\langle 2015-2016\rangle} \sim \mu + \rho\, y_{ij\langle 2014-2015\rangle}. \qquad (2)$$

For this model, a random effect capturing the correlation of clinics from the same health care system was *not* included (see Section 6.1.2). The parameter ρ captures the association between the number of treated days per patient seen across consecutive years, and the parameter μ captures the mean number of treated days per patient seen if no individuals were treated for OUD in the previous year. Let $\hat{\mu}$ and $\hat{\rho}$ denote the estimates from the above model.

Step 2: Using this estimated regression model $(\hat{\mu}, \hat{\rho})$, simulate the outcome measure for the next two-year period (2016-2017) for the UPC clinics (i.e., i=1) from

$$\hat{y}_{1j\langle 2016-2017\rangle} = \varphi + \omega\, \tilde{y}_{1j\langle 2015-2016\rangle} + \varepsilon_{1j}$$

where ε_{1j} are normally distributed with mean zero and variance σ_ε^2. The parameters $\varphi$and ω are analogous to μ and ρ from Equation (2). Note that the $\tilde{y}_{1j\langle 2015-2016\rangle}$ are generated from a normal distribution with

$$E\left(\tilde{y}_{1j\langle 2015-2016\rangle}\right) = {}^{\varphi}/_{(1-\omega)}$$

$$Var\left(\tilde{y}_{1j\langle 2015-2016\rangle}\right) = {\sigma_\epsilon^2}\Big/{(1-\omega^2)}$$

See Appendix B for the justification of these initial values.

Step 3: Repeat the simulation for the UPC clinics in the next overlapping two-year period (i.e., 2017-2018) from

$$\hat{y}_{1j\langle 2017-2018\rangle} = \varphi + \omega\, \tilde{y}_{1j\langle 2016-2017\rangle} + \varepsilon_{1j}.$$

Step 4: Then repeat the simulation for the UPC clinics a final time for the post-randomization period corresponding roughly to 2018-2019 from

$$\hat{y}_{1j\langle 2018-2019\rangle} = \varphi + \omega\, \tilde{y}_{1j\langle 2017-2018\rangle} + \varepsilon_{1j}.$$

Then simulate the 2018-2019 outcome data for the MA Model clinics (i.e., *i*=2) from

$$\tilde{y}_{2j\langle 2018-2019\rangle} = \tau + \tilde{y}_{1j\langle 2018-2019\rangle} + \pi\, \vartheta_j$$

where j are normally distributed with mean zero and variance σ_ϑ^2. Note that $\vartheta_j$induces a correlation between the two clinics from HCS *j*.

6.1.2 Parameters Used to Generate the Simulated Data

Table 3 provides the number of OUD-treated days during a two-year period per patient seen in the clinic. These values were used to estimate the parameters for Steps 1 and 2 in the algorithm summarized in Section 6.1.1.

Table 3: Number of Days Treated for OUDs (with buprenorphine or injectable naltrexone) per Patient Seen at 2 Phase 1 Clinics in Each of 6 Health Care Systems (HCS)

HCS	Fiscal Years 2014-2015		Fiscal Years 2015-2016	
	Clinic 1	Clinic 2	Clinic 1	Clinic 2
1	0.031	0.008	0.085	0.019
2	0.339	0.215	0.354	0.173
3	0.001	0.022	0.001	0.076
4	0.015	0.089	0.021	0.431
5	0.007	0.002	0.004	0.002
6	1.385	0.715	1.361	0.703

From Table 3 we fit a random effects model with a fixed intercept for the number of OUD-treated days per patient seen in FYs 2015-2016 as a function of FYs 2014-2015 where the random effect captures the correlation between clinics from the same HCS. The estimated intercept was 0.05, the coefficient for FYs 2014-2015 was 0.94 and the variance of the random effect was not significantly different from zero (in fact, the actual estimated variance was zero). Thus, the predictive model used to generate the simulated data did not include a random effect for HCS (see Section 6.1.1).

The parameters used for this set of simulations are summarized in Table 4 below.

Table 4: Parameters Explored in the Power Simulation

Parameter	Value
σ_ϵ	0.10
σ_ϑ	0
ω	0.94
φ	0.05

6.1.3 Results of Simulations

Table 5 presents power results (detectable effect size) for the 0.05-level one-tailed test, based on 10,000 iterations per table cell for two models: one without adjustment for baseline, and one with baseline included as a covariate. The column "Inclusion of Baseline as a Covariate" adjusts for the pre-randomization/baseline value of the primary outcome measure.

Table 5: Power Results for a 0.05-level One-Tailed Test, Based on 10,000 Iterations Per Cell

	Model	
k-fold Increase in Primary Outcome (Treatment Effect)	No Adjustment for Baseline	Inclusion of Baseline as a Covariate
1.00	5%	5%
1.06	9%	13%
1.12	15%	29%
1.18	21%	49%
1.24	30%	68%
1.30	39%	84%
1.36	50%	93%
1.42	60%	98%

Based on Table 5, there is at least 80% power to detect a 30% increase in the number of OUD-treated days per patient seen. Thus, with two clinics in each of six HCSs, the study is sufficiently powered to detect the targeted 5-fold increase in the primary outcome measure. As anticipated, there is a substantial gain in power when the baseline value is included as a covariate in the primary outcome model.

6.1.4 Potential Exclusion of One Health Care System

At the time power calculations were originally conducted, it was thought possible that one of the HCSs might not be able to participate due to issues with ceding to the single Institutional Review Board (IRB), though this did not end up occurring (instead, this site randomized 6 months late; see Section 2.5). To address this possibility, the power calculations were repeated but using information from only the five other HCSs and the data simulated arise from only five HCSs and ten clinics.

As with the original power calculations, we fit a random effects model to the remaining health systems' data in Table 3 with a fixed intercept for the number of OUD-treated days per patient seen in FYs 2015-2016 as a function FYs 2014-2015 where the random effect captures the correlation between clinics arising from the same HCS. The parameter estimates from this model did not change substantially. The estimated intercept was 0.05, the coefficient for FYs 2014-2015 was 0.94 and the variance of the random effect was not significantly different from zero. Thus, the predictive model used to generate the simulated data did not include a random effect for HCS. Table 6 summarizes the results of these additional simulations.

Table 6: Power Results for a 0.05-level One-Tailed Test, Based on 10,000 Iterations Per Cell with *Five* Health Care Systems

	Model	
k-fold Increase in Primary Outcome (Treatment Effect)	**No Adjustment for Baseline**	**Inclusion of Baseline as a Covariate**
1.00	5%	5%
1.03	8%	12%
1.06	12%	25%
1.09	16%	40%
1.12	22%	59%
1.18	35%	87%
1.24	49%	98%

With only five HCSs, the study will still be powered sufficiently to detect at least 80% power to detect an 18% increase in the number of OUD-treated days per patient seen. Specifically, the power to detect a 1.18-fold increase in the per patient primary outcome measure is 87%. Thus, with two clinics in each of five HCSs, the study is sufficiently powered to detect the targeted 5-fold increase in the primary outcome measure.

6.2 Power Simulations for Primary Objective 2 Outcome

6.2.1 Description of Power Simulation

We investigated the power of the primary Objective 2 analysis via Monte Carlo simulation. We assumed the following sample sizes for the number of patients with a prior OUD diagnosis over a 3-year period from the Phase 1 data, reflecting the 3-year baseline period of PROUD during which patients with an OUD diagnosis will be identified:

site_id	clin_num	Clin	nOUD
A	1	A1	9
A	2	A2	12
C	1	C1	63
C	2	C2	39
E	1	E1	58
E	2	E2	200
I	1	I1	100
I	2	I2	49
J	1	J1	27
J	2	J2	10
K	1	K1	388
K	2	K2	290

We generated individual-level outcome data within each of the 12 clinics as follows. First, we randomly assigned one of the two clinics within a HCS to receive the PROUD intervention and the other to the Usual Primary Care (UPC) group. We then generated outcome data from a Poisson distribution with mean number of acute care days over a two-year period (time-frame of PROUD outcome ascertainment) using the following mean model,

$$\log E(y_{ijk}) = \alpha + \beta * trt_{ij} + \theta_{ij}$$

where

1) y_{ijk} is the number of days of acute care utilization of patient k in clinic i of HCS j
2) trt_{ij} is the treatment indicator (PROUD intervention versus UPC) for clinic i in HCS j
3) $\theta_{ij} \sim N(0, \tau^2)$ is the random effect for clinic i in HCS j (τ is the standard deviation of the clinic-level random effect)

For the parameter α we assumed that the baseline rate of acute care utilization over a two-year period for patients assigned to a UPC clinic was equal to the average number of acute care visits among patients with a prior OUD diagnosis obtained from Phase 1 data (=4.0 visits) multiplied by the average number of days per acute care visit. The average number of days per acute care visit was based on data from one health system on the average length of stay among all patients (since length of stay data was not available from all health systems at Phase 1), which was 2.04 days per visit. That is, we assumed $\alpha = \log(4 * 2.04) = 2.1$. We considered a range of values for the intent-to-treat relative risk parameter $RR_{ITT} = \exp(\beta)$ governing the association between being assigned to the PROUD intervention and acute care utilization over the follow-up period. Finally, we considered three different values for the standard deviation τ of the clinic-level random effect θ_{ij}. Specifically, we estimated a value for τ using Phase 1 data of $\tau = 0.23$, and also considered two values as a sensitivity analysis: one that was 50% smaller ($\tau = 0.12$) as well as one that was 50% larger ($\tau = 0.35$). We estimated power and type 1 error based on the standard Wald test, as well as the Wald F test that used a denominator degree of freedom based on the Between-Within (BW) small sample degree of freedom correction. For testing the coefficient β from the above model, the BW method uses as denominator degree of freedom (10 = 12 clinics - 2 fixed effect parameters being estimated). Results are based off of 1,000 simulation repetitions.

In addition to presenting power across values of RR_{ITT}, we also provide additional context in light of the fact that not all individuals in the intervention clinic with an EHR documented OUD diagnosis pre-randomization will visit the PROUD NCM and receive sustained treatment with buprenorphine or injectable naltrexone (hereafter “treated for OUDs”), which is hypothesized to meaningfully reduce acute care utilization.[17] Specifically, for different values of RR_{ITT}, we report the proportion of patients who would need to be treated for OUDs (denoted by p_{treat}), if the relative risk of acute care utilization comparing patients who are treated for OUDs versus patients who are not treated for OUDs ($RR_{treated}$) is 0.1 or 0.2. A value of 0.2 corresponds approximately to the observed RR of acute care visits comparing those without OUD to those with OUD; a value of 0.1 corresponds to the assumption that those who are treated for OUDs will have a 50% decrease in the average visit length compared to those with OUD who are not treated. These calculations assume that patients with OUD in a PROUD intervention clinic who are not treated for OUDs have the same rate of acute care utilization as patients with OUD in UPC clinics.

6.2.2 Results of Power Simulations

The following table shows the type 1 error rates for each of the 3 values of clinic-level random effect SD (τ) using the naïve Wald test (“Wald” below), as well as the Wald F test based on the BW degree of freedom correction (BW below):

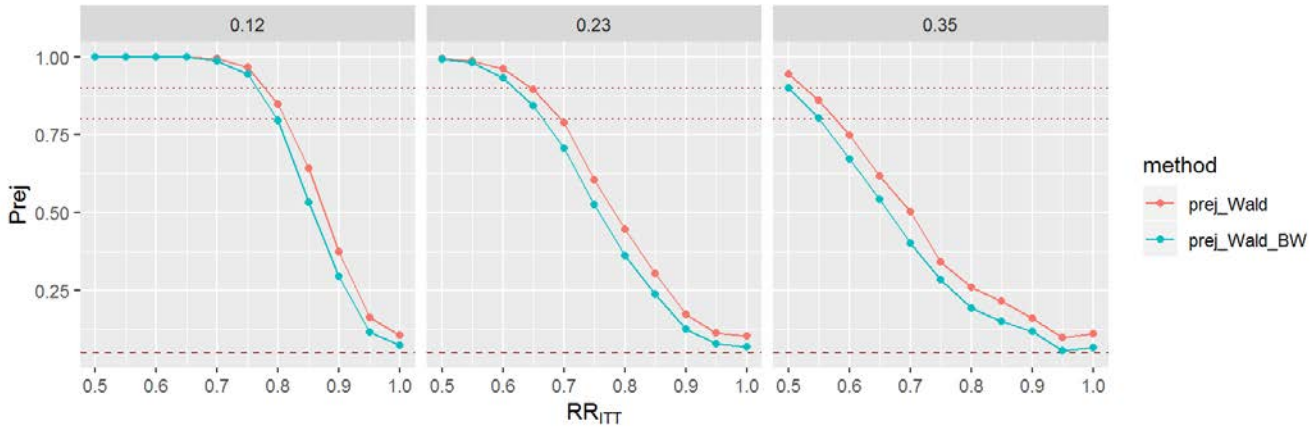
	τ	**Wald**	**BW**
Sensitivity	0.12	0.106	0.075
Primary	0.23	0.104	0.070

	τ	Wald	BW
Sensitivity	0.35	0.111	0.067

Although the type 1 error rates using the BW method (0.067-0.075) are still slightly elevated over the nominal 0.05 level, they are much closer to the correct level as compared to the standard Wald test (all > 0.1). We will continue to explore whether this can be improved further as the SAP continues to be developed.

The power (or type 1 error rates for $RR_{ITT} = 1$) across different effect sizes (parameterized by RR_{ITT}) and different values of the random-effect SD τ (0.12, 0.23, and 0.35) is given in Figure 2:

Figure 2: Objective 2 Power Across Different Effect Sizes



Here 'Prej' denotes the proportion of Monte Carlo iterations for which the null hypothesis was rejected; each panel corresponds to a different value of the random-effect SD (τ). Based on using the BW degree of freedom correction approach, we estimated that we have >80% power to detect values of RR_{ITT} < 0.65 when $\tau = 0.23$, corresponding to a 35% reduction in the acute care utilization rate among patients in a PROUD versus UPC clinic. Similarly, we have >80% power to detect values of RR_{ITT} < 0.80 when $\tau = 0.12$ and of RR_{ITT} < 0.55 when $\tau = 0.35$.

We next provide additional context on the corresponding proportion of patients in PROUD intervention clinics who would need to be treated for OUDs (p_{treat}) in order to detect the above values of RR_{ITT} when the underlying relative risk of acute care utilization comparing patients treated for OUDs versus those who are not treated ($RR_{treated}$) is 0.1 or 0.2.

	τ	RR_{ITT}	Proportion of patients needing to be treated for OUD (p_{treat})	
			RR among treated ($RR_{treated}$)=	
			0.1	0.2
Sensitivity	0.12	< 0.80	>22%	>25%
Primary	0.23	< 0.65	>39%	>44%
Sensitivity	0.35	< 0.55	>50%	>56%

Thus, under our primary assumption for the random-effect SD (τ) if at least 39-44% of patients with OUD in the PROUD intervention arm at baseline are treated for OUD by the NCM, then we will have over 80% power to detect at least a 35% decrease in acute care utilization (RR_{ITT} < 0.65)

comparing patients with OUD in the PROUD intervention arm versus UPC, when the true RR comparing treated to untreated patients with OUD ($RR_{treated}$) is 0.1-0.2. If fewer patients are treated, then our power would be less than 80% under these same assumptions.

7.0 DESCRIPTIVE ANALYSES

7.1 Analyses of Demographic and Baseline Data

The demographic variables for this study include age, sex, race/ethnicity, zip code-based characteristics (e.g., census variables), and type of insurance. The baseline clinical characteristics include diagnoses of medical conditions, mental health disorders, substance use disorders, and co-morbidity indices.

Descriptive statistics for baseline and demographic variables will be presented for the randomized clinics and for participants assigned to the clinics, overall and separately for each of the treatment arms (as well as secondary analyses within each HCS). Descriptive statistics will include N, mean, standard deviation, median, 25th and 75th percentiles, minimum and maximum for continuous variables and proportions and percentages for categorical variables. Since randomization is expected to produce balance at baseline between the two arms of the trial, statistical comparisons of treatment groups with respect to baseline characteristics will not be conducted. The updated CONSORT statement for parallel group-randomized trials no longer recommends formal testing of statistical significance of differences between baseline characteristics.[28]

7.2 Crossover

For all measures described below, patients will be assigned to a primary care clinic based on the algorithm stipulated above in the SAP (Section 3.0).

7.2.1 Main measures of Crossover of Patients with OUDs and Treated for OUDs

We will describe cross-over in terms of the following key measure:

Cross-over between clinic arms*	The number of patients with OUDs assigned to each clinic (PROUD and UPC) in the pre-randomization period who are seen in the other clinic post-randomization.

For this measure, we will evaluate crossover in both directions, each separately (from PROUD to UPC and vice versa). However, the main analysis of interest will be the proportion of UPC patients with OUD diagnosed prior to randomization who are treated for OUDs in a PROUD clinic post-randomization.

7.2.2 Other General Measures of Crossover

In addition to evaluating crossover among patients *with an OUD diagnosis* in the 3 years prior to randomization, we will also describe crossover more generally, among *all patients seen* in the primary care clinic. Specifically, crossover will be defined as occurring if a patient assigned to a clinic in one arm of the trial (PROUD or UPC) during the 3 years prior to randomization is seen in a clinic in the other arm post-randomization. As above, we will evaluate crossover in both directions (separately), with the main analysis of interest the proportion of UPC patients who are treated for OUDs in a PROUD clinic post-randomization.

7.2.3 Crossover of Providers

If we observe that patients assigned to the UPC clinic (with or without an OUD diagnosis) are treated for OUDs post-randomization in the PROUD intervention clinic or vice versa, we will examine crossover of PC providers who prescribe buprenorphine. Specifically, we will examine whether PC providers move between randomized PROUD intervention and UPC clinics.

7.3 Implementation

We will describe, in each of the PROUD intervention clinics, the time from randomization until the NCM was (1) hired, (2) trained in Boston, (3) engaged the first patient in care in the PROUD intervention clinic, (4) left the study; (5) second NCM hired; and (6) engaged the first 10 patients in care in the PROUD intervention clinic. We will also summarize other relevant implementation milestones, such as when the Boston Training and Technical Assistance team conducted a clinic visit as well as those captured by the weekly NCM reports:

For each week
- # of New starts
- # of Re-engagements
- # of Transfers (patients new to NCM who were already on treatment)
- # of Discharges

Cumulative
- Total Ever Treated on Meds
- Total Currently Treated on Meds
- # of Starts for injectable BUP, Oral BUP, Oral NTX, Injectable NTX
- # of Re-engagements
- # of Transfers (patients new to NCM who were already on treatment)

7.4 Characteristics of Treated Patients

Key components of the PROUD intervention are to increase documentation and treatment of OUD among patients already in the clinic, and to attract new patients to receive care in the clinic who have not been seen previously in the clinic or who may be new to the HCS entirely. We therefore plan to describe, overall and by intervention arm, the proportion of patients with medication treatment for OUDs after randomization who were

1) In the clinic pre-randomization and
    a. had an EHR-documented OUD diagnosis (“documented OUD”) pre-randomization
    b. did not have documented OUD pre-randomization
2) Not in the clinic pre-randomization and
    a. were seen pre-randomization in the clinic of the HCS that was randomized to the other arm of the trial
        i. With documented OUD pre-randomization
        ii. Without documented OUD pre-randomization
    b. were seen pre-randomization in a different clinic of the HCS that was not randomized
        i. With documented OUD pre-randomization
        ii. Without documented OUD pre-randomization
    c. were not seen pre-randomization in the HCS

We will also describe these proportions among patients who engaged with the NCM.

7.5 Descriptive Analysis of Outcomes Among Treated Patients

Objective 2 effectiveness analyses (described above) provide an estimate of the intervention effect among patients with an OUD diagnosis who were in an intervention versus control clinic, regardless of whether the patient was actually treated by the NCM. For example, among patients with an OUD diagnosis at baseline, some may have left the HCS, some may not be interested in receiving medication treatment for OUD, and some may be seeking treatment externally (e.g., via methadone maintenance). We therefore plan to additionally describe patient-level outcomes (e.g., acute care utilization, days of OUD medication treatment) among patients with an OUD diagnosis pre-randomization across levels of engagement with the NCM (e.g. seen by the NCM vs not).

We also plan to conduct descriptive analyses to characterize whether OUD treatment is being provided elsewhere within HCS before the patient sees the NCM, and is transferred to the PROUD intervention clinic, by describing the number of days of OUD treatment (pre- and/or post-randomization) a patient received prior to their first visit with the NCM using the most inclusive secondary measure (e.g., including dispensed buprenorphine; see Section 5.2.4).

7.6 Descriptive Information on Data Sources

We will describe the proportion of patients within each HCS with claims data available, which is expected to include most of the sample for one of the HCS and 10-30% enrolled patients from a second HCS. We will look at this overall and by clinic (intervention vs. control).

7.7 Health Care System-specific Descriptive Analyses of Outcome Measures

Because some HCS may have had different degrees of success in implementation, we plan to describe primary and secondary study outcome measures by study arm within each health care systems.

8.0 OTHER CONSIDERATIONS

8.1 Missing Data

Given that the primary and secondary outcomes rely on the EHR data or insurance claims data, if there is no evidence of a particular event, such as provision of buprenorphine or a visit to the ED, we will assume that the event did not occur. These same assumptions apply to the only covariate in the primary analyses of the primary Objective 1 and Objective 2 measures (baseline value of the respective measure in the 2 years prior to randomization).

Our primary Objective 1 outcome uses medication orders to capture days of OUD medication treatment as orders were available for all health systems and dispensings are considered incomplete at 5 of the 6 health systems. However, because dispensing data and not medication orders are considered the gold standard, we will conduct sensitivity analyses in which we estimate the primary outcome at each health system using only medication dispensing data in the 3 health systems that have some dispensing data and pharmacies in both the usual care and intervention clinics (Section 5.2.4).

For Objective 2, given our visit-based sample we are not able to identify when a patient is no longer observable. To address this, we will conduct a sensitivity analysis for our Objective 2 outcome among the two health systems that are health insurance plans in which we use enrolled samples (Section 5.4.3). We will also conduct sensitivity analyses that restrict to patients with evidence of being active in the health system post-randomization (as described above in Section 5.4.2).

Secondary analyses of the primary Objective 1 outcome and Objective 2 outcomes adjust for additional covariates as described above. Based on explorations of covariate distributions in prior data pulls (e.g., data pull 4), we anticipate that fewer than 6% of patients will have a missing value for any one of the individual covariates being considered (See Table 7). For Objective 1, in defining clinic-level covariates (e.g., proportion of patients with commercial insurance), we plan to exclude patients with missing covariate values when calculating clinic-level measures. For example, we will estimate the proportion of a clinic's patients who are commercially insured based on those with non-missing insurance status. Similarly, we will calculate the average neighborhood-level SES measures for a clinic based on patients with non-missing values (i.e., who have a zip code that was able to be linked to the census data). For Objective 2, for patients with missing covariate values, we plan to apply either mean imputation (for continuous covariates) or to use the indicator method (for categorical variables;[29] i.e., by including a "missing" category).

Table 7: Missingness in Covariates among Primary Care Patients with a Visit to one of the 12 Randomized Clinics During the Pre-randomization Period (N = 291,136 patients), Based on Data Pull 4

Covariate*	N (%) of patients with missing/unknown
Gender	4 (0.00%)
Race/ethnicity (combined version**)	16,825 (5.78%)
Insurance status (no insurance record, or recorded value listed as "unknown")	11,082 (3.81%)
Neighborhood-level SES	
No zip code available to link to census data	4,422 (1.52%)
No zip code or zip code does not link to valid value	5,906 (2.03%)

* Only covariates with any missingness are shown; diagnosis-based measures (e.g., comorbidity flags) are assumed to not be present if a diagnosis is not documented

** HCSs ask about race/ethnicity in different ways: in some health systems, race is not documented if the patient reports ethnicity as Hispanic; thus, the combined race/ethnicity is used for covariate adjustment

9.0 SAFETY AND INTERIM ANALYSES

Due to the nature of this study—testing an implementation intervention, with all care provided by the health systems and using only secondary data—there are no formal interim analyses of safety performed. Further, all clinical care—and therefore responsibility for the quality of care—in this cluster-randomized pragmatic quality improvement trial is provided by the health systems. Therefore, this study monitored diverse measures of interest for the DSMB, including deaths and overdoses, but was not be able to intervene on the basis of any data and had no formal interim analyses linked to stopping rules. Since all care is provided by the health system, not the study, it would not be appropriate to intervene at the patient or provider level for any safety issue.

10.0 Table Shells

10.1 Objective 1 Paper

Table 1a. Characteristics of clinics' patient population in the pre-randomization period.		
	6 intervention clinics (No. patients =)[a]	**6 usual care clinics (No. patients =)[a]**
	Clinic Mean (SD)	
Staffing and Size of Clinics		
Number of providers (MD, PA, ARNP) in clinic[b]		
Number of buprenorphine prescribers in clinic[c]		
Number of patients seen in clinic		
Proportion of Clinics' Patient Population		
Age, years		
16-17		
18-24		
25–44		
45–64		
65-74		
75+		
Female		
Race/ethnicity		
Hispanic ethnicity		
Non-Hispanic		
Asian		
Black or African American		
American Indian / Alaska Native		
Native Hawaiian or Pacific Islander		
White		
Multiple race		
Other race		
Missing race and ethnicity data		
Insurance status closest to randomization		
Medicare		
Medicaid		
Otherwise insured (e.g., commercial, private)		
Uninsured		
Unknown		
Patients' neighborhood[e]		
Median household income		
% unemployed		
% below federal poverty level		
Rurality-urbanicity[f]		
Urban		
Suburban		
Rural		

Table 1a. Characteristics of clinics' patient population in the pre-randomization period.		
	6 intervention clinics (No. patients =)[a]	**6 usual care clinics (No. patients =)[a]**
	Clinic Mean (SD)	
Housing instability[g]		
Any mental health diagnoses [g]		
Depression		
Anxiety		
ADHD		
PTSD		
Schizophrenia/psychoses		
Other mental health conditions		
Any non-opioid SUD diagnoses[g]		
Tobacco		
Alcohol		
Cannabis		
Stimulant		
Other		
Other diagnoses[g]		
HCV infection		
HIV infection		
Non-cancer pain		
Elixhauser Comorbidity Index [h]		
0		
1		
2+		

[a] All patients with an eligible PC visit to one of the PROUD trial clinics during baseline (prior to randomization)
[b] Number and type of providers determined from encounter data in the electronic health record
[c] Prescribers determined from medication orders in the electronic health records. Providers assigned to clinics based on number of visits with patients in the clinic pre-randomization.
[d] At eligible visit closest to and prior to randomization
[e] Using zip code closest to randomization date
[f] based on rural urban commuting codes (RUCA)
g based on International Classification of Disease codes 2 years prior to randomization
[h] It is standard to calculate the Elixhauser using 1 year of data and thus was calculated using data in the year prior to randomization
Abbreviations: ADHD – attention deficit hyperactivity disorder; OUD – opioid use disorder; PTSD – post-traumatic stress disorder; SD – standard deviation; SUD – substance use disorder

Table 1b. Opioid use disorder (OUD) and OUD treatment related characteristics of clinics in the PROUD trial in the 2 years prior to randomization, N=12		
	6 intervention clinics (No. patients =)[a]	**6 usual care clinics (No. patients =)[a]**
	Mean (SD) across clinics per 10,000 patients seen in the clinic pre-randomization	
Patient years of OUD treatment with buprenorphine or XR-NTX, pre-randomization [a]		
Proportion of clinics' patient population pre-randomization with:		
OUD diagnosis		
Opioid overdose		
Other drug overdoses		
OUD treatment [a]		
Buprenorphine for OUD		
XR-NTX for OUD		
80% of days covered by OUD treatment		
≥ 6-months retention in OUD treatment[b]		
Discontinuation of OUD treatment[c]		
Buprenorphine daily dose ≥16 mg		
Naloxone prescribed		

[a] Defined as having a medication order or procedure code for buprenorphine formulations that are indicated for OUD or having a medication order or procedure code for XR-NTX and 2+ visits with diagnosis code for OUD or opioid overdose during the pre-randomization period.
[b] Restricted to subjects who entered the sample (i.e., had a PC visit) in the 6 months-2.0 years prior to randomization to allow for at least 6 months of follow-up
[c] Defined as a gap in OUD treatment of 60 days
[d] at any time during pre-randomization treatment
Abbreviations: OUD – opioid use disorder; SD – standard deviation; XR-NTX – extended release injectable naltrexone

Table 2. Clinic level primary and secondary implementation outcomes of the Primary Care Opioid Use Disorders Treatment (PROUD) trial, during 2 years post randomization.

Implementation measures	6 intervention clinics	6 usual care clinics	P value[a]
	Mean (SD) across clinics per 10,000 patients seen in clinic post randomization		
Patient years of OUD treatment (primary outcome[b])			
Proportion of clinics' patient population post-randomization with:			
Any OUD diagnosis post-randomization			
Any OUD treatment[b]			
80% of days covered by OUD treatment			
≥ 6-months retention in OUD treatment[b,c]			
Discontinuation of OUD treatment[b,d]			
Buprenorphine daily dose ≥16 mg[e]			
Naloxone prescribed			
Number of buprenorphine prescribers[f]			

[a] Random effects model adjusted for the outcome measure at baseline (two years prior to randomization)
[b] Treatment defined as having a medication order or procedure code for buprenorphine formulations that are indicated for OUD or having a medication order or procedure code for XR-NTX and 2+ visits with diagnosis code for OUD or opioid overdose (diagnosis codes can be during the 2-years pre or 2-years post randomization).
[c] Restricted to subjects who entered the sample (i.e., had a PC visit) in the 1 day to 1.5 years post randomization (1 day to 12 months for the site randomizing 6-months late) to enable at least 6 months of follow-up
[d] Discontinuation defined as a gap in OUD treatment of 60 days
[e] At any time during OUD treatment post randomization
[f] Prescribers determined from medication orders in the electronic health records. Providers assigned to clinics based on number of visits with patients in the clinic post-randomization.
Abbreviations: NCM – nurse care manager; ITT- intent to treat; OUD: opioid use disorder; SD – standard deviation; XR-NTX – extended release injectable naltrexone

Supplemental Table 1a is the main Table 1a above by health system
Supplemental Table 1b is the main Table 1b above by health system
Supplemental Table 2 is the main Table 1a above but using data for the post randomization period and among patients with an eligible visit post randomization (closest to study end as anchor for variables)
Supplemental Table 3 is the main Table 2 above by health system
Supplemental Table 4 is the main Table 2 above with OUD treatment related outcomes stratified by newly initiated treatment[a] versus on-going treatment[b], and the OUD diagnosis outcome stratified by newly diagnosed OUD[c] versus prevalent OUD[d]. [a] OUD treatment in the 2 years post randomization and no OUD treatment in the prior 365 days [b] OUD treatment post-randomization that does not meet criteria for initiated above [c] 1+ OUD diagnosis anywhere in the health system in the 2 years post randomization and no OUD diagnosis in the 2 years pre randomization. [d] 1+ OUD diagnoses in both the 2 years pre randomization and 2-years post randomization

Supplementary Table 5. Sensitivity analyses focused on assumptions made in estimating patient-years of OUD medication treatment, clinic-level results in the 2 years post randomization

	Patient years treatment /10,000 patients seen in clinic post-randomization[a]		
	6 intervention clinics	**6 usual care clinics**	**Mean difference (95% CI)[b]**
Primary outcome			---
Limited to the time period in which the NCM started seeing patients[d]			
Limited to health systems that successfully implement the MA model [e]			
Covariate adjusted[c]			
Using a combination of medication orders and pharmacy dispensings			
Adding Methadone Maintenance Therapy data from 1-2 sites, per 10,000 patients seen in the 1-2 clinics post randomization			
Adding pharmacy dispensings and MMT data at the 1-2 sites with these sources of data			
Requiring 2+ orders/procedure codes to define treatment and omitting the last order			
Altering allowable gap to define continuous episodes of OUD treatment (main analysis allowable gap=7 days)			
2-day allowable gap			
5-day allowable gap			
10-day allowable gap			
Seen in the clinic pre-randomization			
OUD pre-randomization			
No OUD pre-randomization			
Patients new to the clinic post randomization			
Altering assignment of clinic to be the clinic with the most visits post randomization			

[a] Defined as having a medication order or procedure code for buprenorphine formulations that are indicated for OUD or having a medication order or procedure code for injectable naltrexone and 2+ visits with an OUD or OD diagnosis code (diagnosis codes can be during the 2-years pre or 2-years post randomization)

[b] Mean difference in the primary outcome comparing clinics randomized to the PROUD intervention vs. clinics randomized to usual care, estimated from a random effects model adjusted for the primary outcome measure at baseline (two years prior to randomization)

[c] Adjusted for {*2 covariates determined from lasso regression}* and baseline OUD treatment in random effects model

Supplementary Table 6. Implementation of OUD medication treatment at the clinic-level in the 2 years post randomization stratified by patient characteristics, Primary Care Opioid Use Disorders Treatment (PROUD) trial

	6 intervention clinics	**6 usual care clinics**	
	Mean (SD) patient years of treatment across clinics per 10,000 patients in the subgroup		**P value**[b]
Age at eligible visit closest to randomization[a]			
<26			
≥ 26			
Sex			
Female			
Male			
Race/Ethnicity			
Hispanic			
Non-Hispanic			
White			
Black			
Asian			
Other			

[a] Eligible primary care visits closest to and prior to randomization. If no eligible visits prior to randomization, post-randomization visit closest to randomization date was chosen.

[b] P values are presented for omnibus tests evaluating whether there is any difference in the intervention effect across subgroups, as well as for differences in subgroup-specific intervention effects, obtained from a linear mixed model with interaction terms between the intervention group and the subgroup

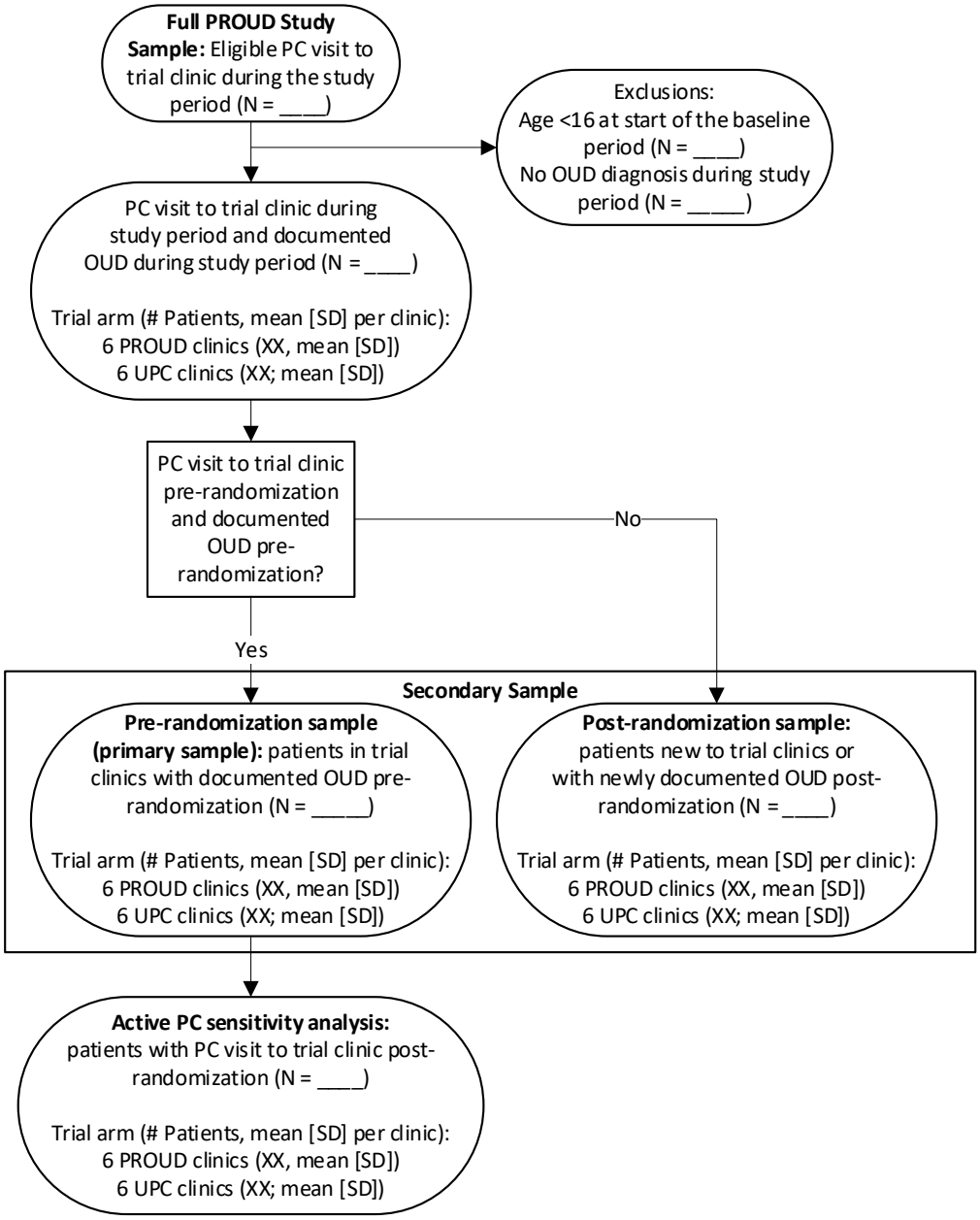
Table 7. Clinic level primary and secondary implementation outcomes of the Primary Care Opioid Use Disorders Treatment (PROUD) trial among patients receiving OUD treatment during the 2 years post randomization.

Scaled implementation measure among treated patients	6 intervention clinics	6 usual care clinics	P value[a]
	Mean (SD) across clinics per 100 patients treated in the clinic post randomization		
Patient years of OUD treatment[b]			
Primary outcome (ITT analysis)[c]			
Limited to the time period after the NCM started seeing patients[d]			
Limited to health systems that successfully implement the MA model [e]			
Number of patients with:			
Newly initiated treatment [a, f]			
On-going treatment[a,g]			
80% of days covered by OUD treatment			
Newly initiated treatment [a, f]			
On-going treatment[a,g]			
≥ 6-months retention in OUD treatment [h]			
Newly initiated treatment[a,f]			
On-going treatment[a,g]			
Discontinuation of OUD treatment[i]			
Newly initiated treatment[a,f]			
On-going treatment[a,g]			
Buprenorphine average daily dose in mg			
Buprenorphine daily dose ≥16 mg[j]			
Naloxone prescription			

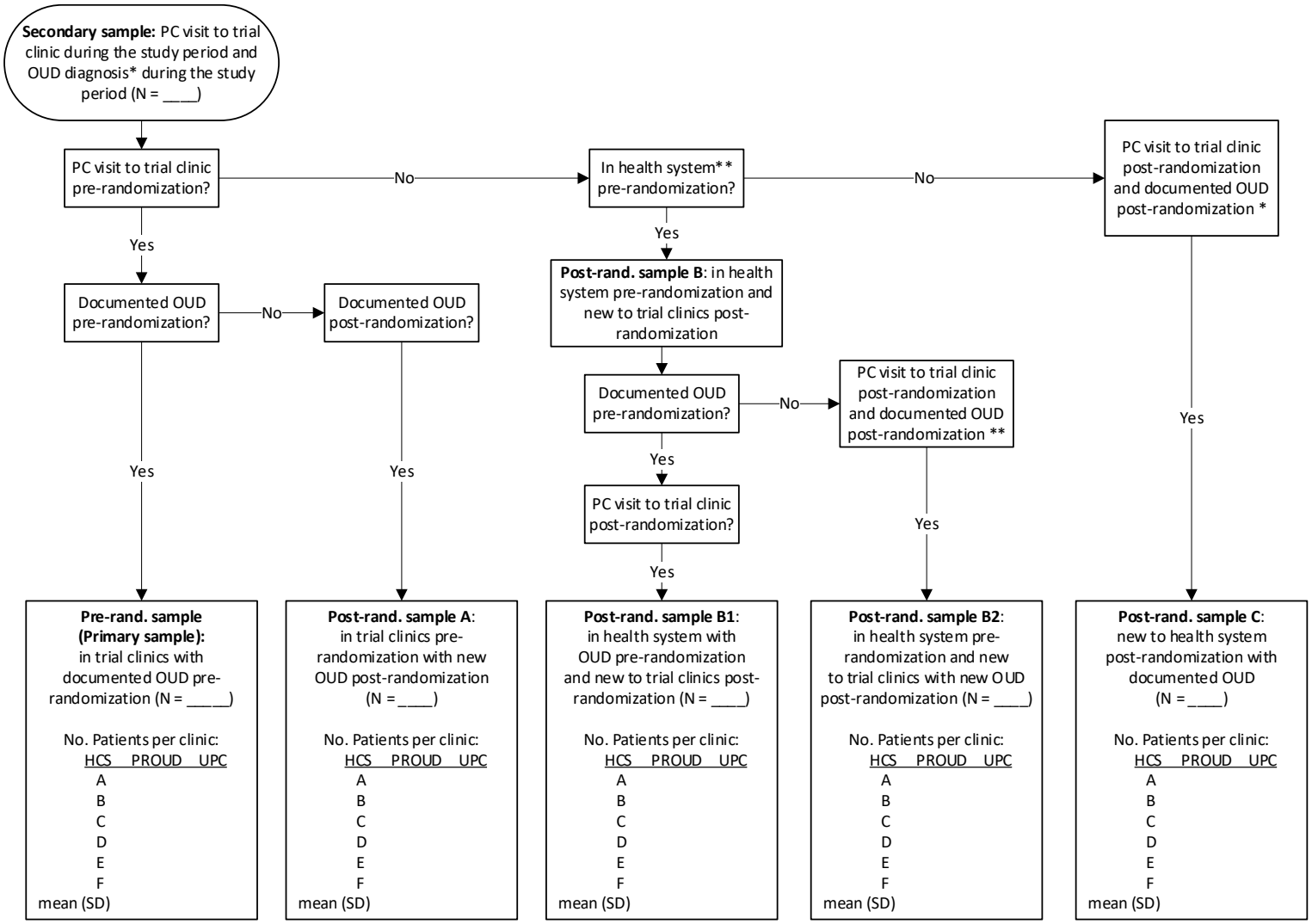
[a] Random effects model adjusted for the primary outcome measure at baseline (two years prior to randomization)
[b] Defined as having a medication order or procedure code for buprenorphine formulations that are indicated for OUD or having a medication order or procedure code for XR-NTX and 2+ visits with diagnosis code for OUD or opioid overdose (diagnosis codes can be during the 2-years pre or 2-years post randomization)
[c] Includes all clinics and all follow-up time post randomization
[d] Time period during which NCM seeing patients: X days for clinic A, X days for clinic B, X days for clinic C, X days for clinic D, X days for clinic E, and X days for clinic F
[e] Successful implementation defined as the nurse care manager seeing at least 30 patients. Analysis includes X of 6 health systems.
[f] OUD treatment in the 2 years post randomization and no OUD treatment in the prior 365 days
[g] OUD treatment in post-randomization, but not meeting criteria for initiated
[h] Restricted to subjects who entered the sample (i.e., had a PC visit) in the 1 day to 1.5 years post randomization (1 day to 12 months for the site randomizing 6-months late) to enable at least 6 months of follow-up
[i] defined as a gap in OUD treatment of 60 days
[j] 16 mg or more of buprenorphine/day at any time post randomization
Abbreviations: NCM – nurse care manager; ITT- intent to treat; OUD: opioid use disorder; SD – standard deviation; XR-NTX – extended release injectable naltrexone

10.2 Objective 2 Paper

Figure 1. Cohort identification of patients for the primary and larger secondary sample of the PROUD Trial effectiveness analyses



Supplemental Figure 1. Samples of patients in the primary and secondary Objective 2 analysis



HCS = health care system
rand. = randomization
* OUD diagnosis anywhere in the health system
** Defined by having any visit, diagnosis, or procedure

Table 1. Characteristics of primary care patients with documented OUD who visited PROUD trial clinics (Effectiveness Objective Samples)

	Pre-randomization Sample: Patients in Trial Clinics with Documented OUD Pre-randomization		**Post-randomization Sample**[a, b]**:** Patients New to Trial Clinics or with Newly Documented OUD Post-randomization	
Characteristic	**Intervention** (N =)	**Usual Care** (N =)	**Intervention** (N =)	**Usual Care** (N =)
	%			
Age at randomization, years				
16-17				
18–24				
25–44				
45–64				
65–74				
75+				
Female				
Race/ethnicity				
Hispanic ethnicity				
Non-Hispanic				
Asian				
Black or African American				
American Indian / Alaska Native				
Native Hawaiian or Pacific Islander				
White				
Multiple race				
Other race				
Missing race and ethnicity data				

Table 1. Characteristics of primary care patients with documented OUD who visited PROUD trial clinics (Effectiveness Objective Samples)

	Pre-randomization Sample: Patients in Trial Clinics with Documented OUD Pre-randomization		**Post-randomization Sample** [a, b]**:** Patients New to Trial Clinics or with Newly Documented OUD Post-randomization	
Characteristic	**Intervention** (N =)	**Usual Care** (N =)	**Intervention** (N =)	**Usual Care** (N =)
Insurance status closest to randomization [b]				
Medicare				
Medicaid				
Otherwise insured (e.g., commercial, private)				
Uninsured				
Unknown				
Unstable housing including homelessness [b]				
Socioeconomic status variables using zip code closest to randomization, median (IQR)				
Median household income				
% unemployed				
% below federal poverty level				
Comorbidity in the two years pre-randomization [b]				
Alcohol use disorder				
Other (non-opioid) substance use disorder				
Schizophrenia and other psychotic disorders				

Table 1. Characteristics of primary care patients with documented OUD who visited PROUD trial clinics (Effectiveness Objective Samples)

	Pre-randomization Sample: Patients in Trial Clinics with Documented OUD Pre-randomization		**Post-randomization Sample** [a, b]**:** Patients New to Trial Clinics or with Newly Documented OUD Post-randomization	
	Intervention	**Usual Care**	**Intervention**	**Usual Care**
Characteristic	(N =)	(N =)	(N =)	(N =)
Baseline OUD treatment [b]				
Days of treatment per year, mean (IQR)				
OUD treatment duration				
0 days				
1-30 days				
31-90 days				
91-180 days				
180+ days (sustained OUD treatment)				
Baseline acute care utilization [b]				
Days of acute care utilization per year, mean (IQR)				
Days hospitalized [c]				
Days of emergency care [d]				

Table 1. Characteristics of primary care patients with documented OUD who visited PROUD trial clinics (Effectiveness Objective Samples)

	Pre-randomization Sample: Patients in Trial Clinics with Documented OUD Pre-randomization		**Post-randomization Sample** [a, b]**:** Patients New to Trial Clinics or with Newly Documented OUD Post-randomization	
Characteristic	**Intervention** (N =)	**Usual Care** (N =)	**Intervention** (N =)	**Usual Care** (N =)
Proportion of patients with:				
Any acute care				
Hospitalization [c]				
Never				
Once				
2-3 times				
4+ times				
Emergency care visit [d]				
Never				
Once				
2-3 times				
4+ times				

HCS = health care system

[a] Includes patients in the trial clinics without an OUD diagnosis pre-randomization who had an OUD diagnosis post-randomization, patients in the HCS (but not in the trial clinics) with an OUD diagnosis pre-randomization or post-randomization, and patients who were new to the HCS post-randomization (see Supplemental Figure)

[b] Patients in the HCS pre-randomization use baseline values of time-varying covariates; patients new to HCS post-randomization do not have baseline values of time-varying covariates (insurance status and zip code use post-baseline measure available closest to randomization date; clinical characteristics use two years post-randomization; baseline treatment and outcomes not available)

[c] Days hospitalized also includes emergency department or urgent care visits that resulted in a hospitalization. [Report % of hospitalizations that started in (or immediately preceded) an emergency visit.]

[d] Emergency care includes visits to an emergency department or urgent care facility that did not result in hospitalization

Table 2. Effect of PROUD Intervention on OUD treatment among patients with documented OUD pre-randomization

Measure of OUD treatment in 2 years post-randomization	**Intervention** (n=xxx)	**Usual Care** (n=xxx)	**Treatment effect** [a]		**P value**
Main treatment measure: days of OUD treatment [c]	mean (IQR) days per year [b]		RR	95% CI	
Primary sample with OUD pre-randomization (intention to treat)					
Sub-samples and/or follow-up periods restricted to:					
a. Clinics that successfully implemented MA model [d]					
b. Time period in which the NCM was seeing patients					
c. Patients with PC visit to assigned clinic post-randomization					
d. Both b and c					
Secondary treatment measures	N (%)[b]		OR	95% CI	
OUD treatment duration					
0 days					
1-30 days					
31-90 days					
91-180 days					
180+ days (sustained OUD treatment)					
Initiation of new episode of OUD treatment [e]					
Initiated and 80% adherence to treatment					
Initiated and sustained OUD treatment					

[a] Estimated from a mixed-effect model with clinic-specific random intercept, adjusted for days of OUD treatment during the baseline period

[b] Unadjusted estimates. We note that if we observe a statistically significant difference from the adjusted model, but the unadjusted estimates appear similar then we plan to also calculate the adjusted estimates

[c] OUD treatment defined as having a medication order or procedure code for buprenorphine formulations that are indicated for OUD or having a medication order or procedure code for XR-NTX and 2+ visits with diagnosis code for OUD or opioid overdose during the pre-randomization period.

[d] Successful implementation is defined as the NCM at the intervention clinic at that HCS seeing at least 30 patients. We note that estimates from the secondary analyses of the main treatment measure may instead be separated into a different paper or included in a supplemental table.

[e] OUD treatment following a period of 365 days with no OUD treatment

Table 3. Effect of PROUD Intervention on acute care utilization among primary care patients with documented OUD pre-randomization

Measure of Acute Care Utilization in 2 years post-randomization	**Intervention** (n=xxx)	**Usual Care** (n=xxx)	Treatment Effect [a]	95% CI	P value
	mean (IQR) days per year [b]		RR		
Primary outcome: days of acute care utilization (intention to treat)					
Sub-samples and/or follow-up periods restricted to:					
a. Clinics that successfully implemented MA model [c]					
b. Time period in which the NCM was seeing patients					
c. Patients with PC visit to assigned clinic post-randomization					
d. Both b and c					
Days of hospitalization [d] (intention to treat)					
Sub-samples and/or follow-up periods restricted to:					
a. Clinics that successfully implemented MA model [c]					
b. Time period in which the NCM was seeing patients					
c. Patients with PC visit to assigned clinic post-randomization					
d. Both b and c					
Days of emergency care [e] (intention to treat)					
Sub-samples and/or follow-up periods restricted to:					
a. Clinics that successfully implemented MA model [c]					
b. Time period in which the NCM was seeing patients					
c. Patients with PC visit to assigned clinic post-randomization					
d. Both b and c					

Table 3. Effect of PROUD Intervention on acute care utilization among primary care patients with documented OUD pre-randomization

Measure of Acute Care Utilization in 2 years post-randomization	**Intervention** (n=xxx)	**Usual Care** (n=xxx)	Treatment Effect [a]	95% CI	P value
	mean (IQR) days per year [b]		RR		
Proportion of patients with:	N (%)[b]		OR		
Any acute care					
Hospitalization [d]					
Never					
Once					
2-3 times					
4+ times					
Emergency care visit [e]					
Never					
Once					
2-3 times					
4+ times					

[a] Estimated from a mixed-effect model with clinic-specific random intercept, adjusted for the baseline value of the outcome (binary/categorical measures adjust for the continuous value of the measure)

[b] Unadjusted estimates. [We note that if we observe a statistically significant difference from the adjusted model, but the unadjusted estimates appear similar then we plan to also calculate the adjusted estimates.]

[c] Successful implementation is defined as the NCM at the intervention clinic at that HCS seeing at least 30 patients. [We note that estimates from the secondary analyses of the main treatment measure may instead be separated into a different paper or included in a supplemental table.]

[d] Days hospitalized also includes emergency department or urgent care visits that resulted in a hospitalization. [Report % of hospitalizations that started in (or immediately preceded) an emergency visit.] Specific cut-points for categorization to be determined based on empirical distribution at baseline

[e] Emergency care includes visits to an emergency department or urgent care facility that did not result in hospitalization; specific cut points for categorization to be determined based on empirical distribution at baseline

Table 4. Observational analysis of effect of PROUD Intervention on OUD treatment in the larger secondary sample that includes patients with newly recognized OUD in trial clinics post-randomization

	Pre-randomization Sample: Patients in Trial Clinics with Documented OUD Pre-randomization [a]		**Post-randomization Sample:** Patients New to Trial Clinics or with Newly Documented OUD Post-randomization				**Secondary Sample** (Overall)
Measure	Treatment effect [b]	95% CI	**Intervention**	**Usual Care**	Treatment effect [b]	95% CI	Omnibus P value [c]
Same measures as Table 2							

[a] Estimating the same thing as Table 2, but adjusted for demographics and not adjusting for baseline (which is not measured among patients new to the health system post-randomization); likely will be omitted from main paper (include in supplement instead)

[b] Estimated from a mixed-effect model with clinic-specific random intercept and an interaction term between the intervention effect and the timing of when the patient was first seen in the trial clinic with diagnosed OUD (pre- or post-randomization) and adjusted for demographics

[c] Testing for whether there is a difference between intervention and usual care clinics for either subset of patients (primary sample or secondary sub-sample added to the primary sample)

Table 5. Observational analysis of intervention effect on acute care utilization in the larger secondary sample including patients with newly recognized OUD in trial clinics post-randomization

	Pre-randomization Sample: Patients in Trial Clinics with Documented OUD Pre-randomization [a]		**Post-randomization Sample:** Patients New to Trial Clinics or with Newly Documented OUD Post-randomization				**Secondary Sample** (Overall)
Measure	RR [b]	95% CI	**Intervention**	**Usual Care**	Treatment Effect [b]	95% CI	Omnibus P value [c]
Same measures as Table 3							

[a] Estimating the same thing as Table 3, but adjusted for demographics and not adjusting for baseline (which is not measured among patients new to the health system post-randomization); likely will be omitted from main paper

[b] Estimated from a mixed-effect model with clinic-specific random intercept and an interaction term between the intervention effect and the timing of when the patient was first seen in the trial clinic with diagnosed OUD (pre- or post-randomization) and adjusted for demographics

[c] Testing for whether there is a difference between intervention and usual care clinics for either subset of patients (primary sample or secondary sub-sample added to the primary sample)

11.0 REFERENCES

1. Alford DP, LaBelle CT, Kretsch N, Bergeron A, Winter M, Botticelli M, Samet JH. Collaborative care of opioid-addicted patients in primary care using buprenorphine: five-year experience. Arch Intern Med 2011;171:425-31.
2. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012;50:217-26.
3. Bobb JF, Qiu H, Matthews AG, McCormack J, Bradley KA. Addressing identification bias in the design and analysis of cluster-randomized pragmatic trials: a case study. Trials 2020;21:289.
4. Gordon AJ, Lo-Ciganic WH, Cochran G, Gellad WF, Cathers T, Kelley D, Donohue JM. Patterns and Quality of Buprenorphine Opioid Agonist Treatment in a Large Medicaid Program. J Addict Med 2015;9:470-7.
5. Bryson CL, Au DH, Sun H, Williams EC, Kivlahan DR, Bradley KA. Alcohol screening scores and medication nonadherence. Ann Intern Med 2008;149:795-804.
6. Hedeker D, Gibbons RD, Flay BR. Random-effects regression models for clustered data with an example from smoking prevention research. J Consult Clin Psychol 1994;62:757-65.
7. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological). 1996 Jan 1:267-88.
8. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33:1-22.
9. Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. J Clin Epidemiol 2015;68:627-36.
10. Bhatraju EP, Grossman E, Tofighi B, McNeely J, DiRocco D, Flannery M, Garment A, Goldfeld K, Gourevitch MN, Lee JD. Public sector low threshold office-based buprenorphine treatment: outcomes at year 7. Addict Sci Clin Pract 2017;12:7.
11. Burns L, Gisev N, Larney S, Dobbins T, Gibson A, Kimber J, Larance B, Mattick RP, Butler T, Degenhardt L. A longitudinal comparison of retention in buprenorphine and methadone treatment for opioid dependence in New South Wales, Australia. Addiction 2015;110:646-55.
12. Gryczynski J, Mitchell SG, Jaffe JH, O'Grady KE, Olsen YK, Schwartz RP. Leaving buprenorphine treatment: patients' reasons for cessation of care. J Subst Abuse Treat 2014;46:356-61.
13. Hser YI, Saxon AJ, Huang D, Hasson A, Thomas C, Hillhouse M, Jacobs P, Teruya C, McLaughlin P, Wiest K, Cohen A, Ling W. Treatment retention among patients randomized to buprenorphine/naloxone compared to methadone in a multi-site trial. Addiction 2014;109:79-87.
14. Marcovitz DE, McHugh RK, Volpe J, Votaw V, Connery HS. Predictors of early dropout in outpatient buprenorphine/naloxone treatment. Am J Addict 2016;25:472-7.
15. Weinstein ZM, Kim HW, Cheng DM, Quinn E, Hui D, Labelle CT, Drainoni ML, Bachman SS, Samet JH. Long-term retention in Office Based Opioid Treatment with buprenorphine. J Subst Abuse Treat 2017;74:65-70.
16. Lagisetty PA, Ross R, Bohnert A, Clay M, Maust DT. Buprenorphine Treatment Divide by Race/Ethnicity and Payment. JAMA Psychiatry 2019;76:979-81.
17. Lo-Ciganic WH, Gellad WF, Gordon AJ, Cochran G, Zemaitis MA, Cathers T, Kelley D, Donohue JM. Association between trajectories of buprenorphine treatment and emergency department and in-patient utilization. Addiction 2016;111:892-902.
18. Hansen H, Siegel C, Wanderling J, DiRocco D. Buprenorphine and methadone treatment for opioid dependence by income, ethnicity and race of neighborhoods in New York City. Drug Alcohol Depend 2016;164:14-21.

19. Stein MD, Cioe P, Friedmann PD. Buprenorphine retention in primary care. J Gen Intern Med 2005;20:1038-41.
20. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997;16:385-95.
21. Kahan BC, Forbes G, Ali Y, Jairath V, Bremner S, Harhay MO, Hooper R, Wright N, Eldridge SM, Leyrat C. Increased risk of type I errors in cluster randomised trials with small or medium numbers of clusters: a review, reanalysis, and simulation study. Trials 2016;17:438.
22. Leyrat C, Morgan KE, Leurent B, Kahan BC. Cluster randomized trials with a small number of clusters: which analyses should be used? Int J Epidemiol 2018;47:321-31.
23. Bell J, Turbow S, George M, Ali MK. Factors associated with high-utilization in a safety net setting. BMC Health Serv Res 2017;17:273.
24. Mohlman MK, Tanzman B, Finison K, Pinette M, Jones C. Impact of medication-assisted treatment for opioid addiction on Medicaid expenditures and health services utilization rates in Vermont. J Subst Abuse Treat 2016;67:9-14.
25. Robinson RL, Grabner M, Palli SR, Faries D, Stephenson JJ. Covariates of depression and high utilizers of healthcare: Impact on resource use and costs. J Psychosom Res 2016;85:35-43.
26. Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S. A combined comorbidity score predicted mortality in elderly patients better than existing scores. J Clin Epidemiol 2011;64:749-59.
27. LaBelle CT, Han SC, Bergeron A, Samet JH. Office-Based Opioid Treatment with Buprenorphine (OBOT-B): Statewide Implementation of the Massachusetts Collaborative Care Model in Community Health Centers. J Subst Abuse Treat 2016;60:6-13.
28. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med. 2010 Mar 24;8:18.
29. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. Stat Med 2005;24:993-1007.

Appendix A. Algorithms for Days of OUDs Treatment

A1. Operationalization to Estimate Number of Patient-days of Treatment for Opioid Use Disorders from Electronic Health Records and Claims Data

The data used to estimate patient-days of treatment for OUD can include medication orders from the EHR, and procedure codes from the EHR. Medication dispensings from pharmacy claims and the EHR will be considered in sensitivity analyses only. Medication orders (and dispensings) commonly include variables for the drug name (generic and/or brand); date dispensed/ordered; quantity dispensed or ordered; intended days' supply (dispensings only); directions for use (used to estimate days' supply for orders); strength per unit; and prescriber. Procedure data includes the drug name, strength, date administered, and provider. Using the data noted above, we estimate runout dates (date when the OUD treatment days provided by that particular dispensing/order/procedure ends) for each order/procedure (and dispensings in sensitivity analyses) of OUD treatment (buprenorphine [Table A1] and naltrexone) by adding the intended or estimated days' supply to the start date (date of dispensing, order, or procedure). Sublocade and naltrexone injections are assumed to provide 28 days of OUD treatment (i.e., days' supply=28). Probuphine is assumed to provide 6 months of treatment. Each unique order, and procedure (and dispensings when included), are then combined into an episode or episodes (if gaps) of OUD treatment. Episodes are then summed to calculate the total patient days covered with OUD treatment in a given time period. See Section A3 for details on how episodes are estimated from the unique orders and procedures (and dispensings when included).

Often data obtained from clinical and/or administrative systems confronts us with missing data fields (i.e., quantity dispensed/ordered or intended days' supply) or situations that would be very unlikely or impossible, such as an unrealistic daily dose of buprenorphine or a day's supply greater than 6 months for a controlled substance. Therefore, as part of Phase 1 of CTN-0074, we developed data cleaning rules to address these situations. Section A2 outlines some of the key situational assumptions.

Table A1: Buprenorphine Formulations Included for OUD Treatment

Medication Name(s)	Route of Administration (form)	Strength in Milligrams (mg)
Buprenorphine (Brands: Subutex®)	Sublingual tablet and film	2 mg, 8 mg
Buprenorphine/naloxone (Brands: Suboxone®, Cassipa®)	Sublingual tablet and film	2 mg/0.5 mg, 4mg/1mg, 8 mg/2 mg, 12mg/3mg, 16mg/4mg
Bunavail® (buprenorphine/naloxone)	Buccal film	2.1 mg/0.3 mg, 4.2 mg/0.7 mg, 6.3 mg/1 mg
Buprenorphine/naloxone (Brand: Zubsolv®	Sublingual tablet	1.4 mg/0.36 mg, 2.9 mg/0.71 mg, 5.7 mg/1.4 mg, 8.6 mg/2.1 mg, 11.4 mg/2.9 mg
Buprenorphine (Brand: Probuphine®)	Subdermal implants	4 single rods each 74.2 mg
Buprenorphine (Brand: Sublocade)	Subcutaneous injection	100 , 300 mg

A2. Data Cleaning and Quality Checks for Individual Oral Buprenorphine Orders (and dispensings for sensitivity analyses)

Variables	Definition
Sid	Unique study id
rxname	Drug name (e.g., buprenorphine, naltrexone, buprenorphine and naloxone)
rxamt	Quantity ordered (e.g., 30 tablets)
rxsup	Supply of the order or how long the order should last (e.g., 30 days) This is determined for orders by 1) converting SIG (directions for use) into pills per day (PPD). For example, “take 1 tablet twice a day” is 2 PPD; 2) dividing the rxamt by ppd. For example, an order for rxamt=30 that has PPD=2 would have a rxsup=30/2 or 15 days. Days supply is then rounded to nearest whole number.
strength	Mg in one dose (e.g., 8 mg)
daily dose (DD)	=PPD * strength, for example, 2 PPD of 8 mg bup is 16 mg/day.
date	Date medication was ordered or dispensed

1) Summarize duplicates for orders. Duplicates defined as orders for the same **patient (sid)** with same rxname, date, and strength.
    - Number of total orders =
    - Number of unique subjects =
    - Number of subjects with at least 1 duplicate pair:
    - Percent of subjects with at least 1 duplicate pair:
    - Number of total orders that are considered a potential duplicate =
        - Example: 6 orders for the same patient where orders 1&2 have the same rxname, order date, and strength, and orders 3-5 also have the same rxname, date, and strength. This is a total of 5 duplicates. The percent of duplicate pairs across all this patient’s orders is 5 divided by 6 (total number of orders).
    - Percent of duplicate pairs across total orders:
2) Provide potential duplicates to health systems and ask that they chart review and de-dupe where needed (data anomaly) or provide revised order “date” if they are not duplicates
3) **De-dupe remaining duplicates**. If rxsup is different or 1+ is missing, keep the one with the largest rxsup. If rxsup is the same but quantity differs, keep the one with the largest quantity. If rxsup and quantity are the same across dupes, randomly keep one order.
    - Number of total orders =
    - Number of unique subjects (should not change):
4) Derive a pills per day (PPD) variable =Rxamt / Rxsup for each bup rx where rxamt and rxsup are known and reasonable (not missing, not zero, not <1).

5) Calculate daily dose (DD) variable = PPD x strength (where strength is known). Round DD to an integer for each bup rx.
6) Set missing refills to = 0
7) Provide frequencies in table 1 and table 2 below for buprenorphine. This is post de-dupe. (The days supply in health system C are calculated as quantity/ppd before calculating the frequencies. They are originally missing.)

Table 1a. Proportion of orders and unique individuals with data anomalies		
	No. bup orders=	**No. subjects=**
	n (%)	
Missing		
Strength		
DD		
Rxsup		
Rxamt		
Both rxsup & rxamt		
Missing strength, rxsup, & rxamt		
Zero value		
Rxsup		
Rxamt		
Both Rxsup & rxamt		
<0 as a value		
Rxsup		
Rxamt		
Both Rxsup & rxamt		
Specific cut points		
DD>40		
DD>40 in 3+ consecutive Rx		
DD>40 in < 3 consecutive Rx		
Rxsup<1		
Rxsup>30		
Rxsup >60		
Rxsup>90		
Rxsup>180		
Rxsup=1 & DD>8 mg		
Refills ≥1		
Refills >6		

Table 1a. Proportion of orders and unique individuals with data anomalies		
	No. bup orders=	**No. subjects=**
	n (%)	
Rxamt<1		
Rxamt=1		
Rxamt >30		
Rxamt >60		
Rxamt >90		

Table 2a. Distribution of values for raw variables of interest						
	Min	**25th**	**50th**	**75th**	**Max**	**% above 3 times 75th percentile**
DD						
Strength						
Rxamt						
Rxsup						
Refills						

8) Output all bup orders for subjects with outliers (defined as above 3 x the 75th percentile) for rxsup and/or rxamt.

9) Output all bup orders for subjects with rxup and/or rxamt less than 1.

10) Co-I reviews all bup orders identified in steps 7 and 8 to flag which outliers are likely legit and should not be imputed. Co-I to determine if and what to impute for rxsup or rxamt <1.

11) For rxsup and rxamt outliers that are to be imputed and for rxsup and rxamt values <1 that are to be imputed, set to missing.

12) Impute all missing rxsup and rxamt as follows below. We do not need to impute refills, strength, ppd, or DD.

- Use the most common value from the same patient's bup rx w/in +/- 100 days of missing field Take the value closest to missing if there is a tie in the # of occurrences. If tie in both, use the smallest value.
- If the patient has no other bup rx w/in +/- 100 days of the missing, impute the median value from all bup rx at that health system. If all bup rx are missing the field at that health system, impute the median value from all bup rx at all health systems.

13) Populate Table 1b and 2b below (post imputation of bup)

Table 1b. Proportion of buprenorphine orders and unique individuals with data anomalies		
	No. bup orders=	**No. subjects=**
	n (%)	
Specific cut points		
DD>40		
DD>40 in 3+ consecutive Rx		
DD>40 in < 3 consecutive Rx		
Rxsup<1		
Rxsup>30		
Rxsup >60		
Rxsup>90		
Rxsup>180		
Rxsup=1 & DD>8 mg		
Refills ≥1		
Refills >6		
Rxamt<1		
Rxamt=1		
Rxamt >30		
Rxamt >60		
Rxamt >90		

Table 2b. Distribution of values for raw variables of interest that were imputed					
	Min	**25th**	**50th**	**75th**	**Max**
Rxamt					
Rxsup					

14) Pull in buprenorphine (J0570, G2070, G2072, Q9991, Q9992, G2068, G2079, G2069, J0571, J0572, J0573, J0574, J057), injectable naltrexone (J2315, G2073, HZ84ZZZ, HZ94ZZZ), and methadone maintenance therapy (HZ91ZZZ, HZ81ZZZ, H0020, G2067, G2078) procedures from procedures data tables and injectable naltrexone (defined as brand= Vivitrol or route=intramuscular) from medication order tables.

Procedure data	Number of occurrences	Unique number of patients
Implant probuphine - J0570, G2070, G2072		
Sublocade - Q9991, Q9992, or G2069		
Oral bup - J0571, J0572, J0573, J0574, or J0575, G2068, G2079		
Oral bup procedure codes after adding up ones billed on same day (see Step 14)		
XR injectable naltrexone procedures (J2315, G2073, HZ84ZZZ, HZ94ZZZ)		
XR injectable naltrexone from medication orders		
Methadone maintenance therapy (HZ91ZZZ, HZ81ZZZ, H0020, G2067, G2078)		

15) Create a start and end date for each bup rx and procedure

- Start date=rxdate (orders) or adate (procedures)
- For bup orders: End date= (rxdate + rxsup) -1
    - If refills, multiply rxsup by number of refills +1 to calculate the end date. For example, an order with one refill and rxsup=30 would be 30*2 = 60 days. An order with 2 refills and rxsup=30 would be 30*3=90 days.
- For oral bup procedures (J0571, J0572, J0573, J0574, J0575), set rxsup=1 and enddate=(adate + rxsup)-1. Essentially, each procedure code for oral med covers 1 day. If duplicates (same day for any of these 5 codes), add them together and then keep 1. For example, a subject with 3 oral bup codes (any not distinct codes) on 1/1/2020 would have a start date of 1/1/2020 and rxsup=3. Keep just one of these records with rxsup=3 and enddate=(1/1/2020 + 3) -1 or 1/3/2020
- For oral bup procedures (G2068 or G2079), set rxsup=7 and enddate=adate + rxsup -1
- For procedures J0570, G2070, G2072: (buprenorphine rods; brand is probuphine): enddate=adate+179 days
- For procedure Q9991, Q9992, or G2069 (Sub Q buprenorphine; brand is sublocade): enddate=adate + 27 days
- For injectable naltrexone (from procedure codes J2315, G2073 or naltrexone XR injectable orders (brand=Vivitrol or route=intramuscular): enddate=adate (procedure) or rxdate (order) + 27 days

- Define the list of OUD ICD-9 and ICD-10 diagnosis codes

Opioid abuse	F11.1
Opioid abuse, uncomplicated	F11.10
Opioid abuse, in remission	F11.11

Opioid abuse with intoxication	F11.12
Opioid abuse with intoxication, uncomplicated	F11.120
Opioid abuse with intoxication delirium	F11.121
Opioid abuse with intoxication with perceptual disturbance	F11.122
Opioid abuse with intoxication, unspecified	F11.129
Opioid abuse with opioid-induced mood disorder	F11.14
Opioid abuse with opioid-induced psychotic disorder	F11.15
Opioid abuse with opioid-induced psychotic disorder with delusions	F11.150
Opioid abuse with opioid-induced psychotic disorder with hallucinations	F11.151
Opioid abuse with opioid-induced psychotic disorder, unspecified	F11.159
Opioid abuse with opioid-induced disorder	F11.18
Opioid abuse with opioid-induced sexual dysfunction	F11.181
Opioid abuse with opioid-induced sleep disorder	F11.182
Opioid abuse with other opioid-induced disorder	F11.188
Opioid abuse with unspecified opioid-induced disorder	F11.19
Opioid dependence, uncomplicated	F11.20
Opioid dependence, in remission	F11.21
Opioid dependence with intoxication, uncomplicated	F11.220
Opioid dependence with intoxication delirium	F11.221
Opioid dependence with intoxication with perceptual disturbance	F11.222
Opioid dependence with intoxication, unspecified	F11.229
Opioid dependence with withdrawal	F11.23
Opioid dependence with opioid-induced mood disorder	F11.24
Opioid dependence with opioid-induced psychotic disorder with delusions	F11.250
Opioid dependence with opioid-induced psychotic disorder with hallucination	F11.251
Opioid dependence with opioid-induced psychotic disorder, unspecified	F11.259
Opioid dependence with other opioid-induced disorder	F11.28
Opioid dependence with opioid-induced sexual dysfunction	F11.281
Opioid dependence with opioid-induced sleep disorder	F11.282
Opioid dependence with other opioid-induced disorder	F11.288
Opioid dependence with unspecified opioid-induced disorder	F11.29
Opioid type dependence, unspecified	304.00
Opioid type dependence, continuous	304.01
Opioid type dependence, episodic	304.02
Opioid type dependence, in remission	304.03
Combinations of opioid type drug with any other drug dependence, unspecified	304.70
Combinations of opioid type drug with any other drug dependence, continuous	304.71
Combinations of opioid type drug with any other drug dependence, episodic	304.72
Combinations of opioid type drug with any other drug dependence, in remission	304.73
Opioid abuse, unspecified	305.50

Opioid abuse, continuous	305.51
Opioid abuse, episodic	305.52
Opioid abuse, in remission	305.53

- Define the list of opioid overdose ICD-9 and ICD-10 diagnosis codes

Poisoning by, adverse effect of and underdosing of opium	T40.0
Poisoning by, adverse effect of and underdosing of opium	T40.0X
Poisoning by opium, accidental (unintentional)	T40.0X1
Poisoning by opium, accidental (unintentional), initial encounter	T40.0X1A
Poisoning by opium, accidental (unintentional), subsequent encounter	T40.0X1D
Poisoning by opium, accidental (unintentional), sequela	T40.0X1S
Poisoning by opium, intentional self-harm	T40.0X2
Poisoning by opium, intentional self-harm, initial encounter	T40.0X2A
Poisoning by opium, intentional self-harm, subsequent encounter	T40.0X2D
Poisoning by opium, intentional self-harm, sequela	T40.0X2S
Poisoning by opium, assault	T40.0X3
Poisoning by opium, assault, initial encounter	T40.0X3A
Poisoning by opium, assault, subsequent encounter	T40.0X3D
Poisoning by opium, assault, sequela	T40.0X3S
Poisoning by opium, undetermined	T40.0X4
Poisoning by opium, undetermined, initial encounter	T40.0X4A
Poisoning by opium, undetermined, subsequent encounter	T40.0X4D
Poisoning by opium, undetermined, sequela	T40.0X4S
Poisoning by and adverse effect of heroin	T40.1
Poisoning by and adverse effect of heroin	T40.1X
Poisoning by heroin, accidental (unintentional	T40.1X1
Poisoning by heroin, accidental (unintentional), initial encounter	T40.1X1A
Poisoning by heroin, accidental (unintentional), subsequent encounter	T40.1X1D
Poisoning by heroin, accidental (unintentional), sequela	T40.1X1S
Poisoning by heroin, intentional self-harm,	T40.1X2
Poisoning by heroin, intentional self-harm, initial encounter	T40.1X2A
Poisoning by heroin, intentional self-harm, subsequent encounter	T40.1X2D
Poisoning by heroin, intentional self-harm, sequela	T40.1X2S
Poisoning by heroin, assault	T40.1X3
Poisoning by heroin, assault, initial encounter	T40.1X3A
Poisoning by heroin, assault, subsequent encounter	T40.1X3D
Poisoning by heroin, assault, sequela	T40.1X3S
Poisoning by heroin, undetermined	T40.1X4
Poisoning by heroin, undetermined, initial encounter	T40.1X4A
Poisoning by heroin, undetermined, subsequent encounter	T40.1X4D
Poisoning by heroin, undetermined, sequela	T40.1X4S
Poisoning by, adverse effect of and underdosing of other opioids	T40.2
Poisoning by, adverse effect of and underdosing of other opioids	T40.2X

Poisoning by other opioids, accidental (unintentional)	T40.2X1
Poisoning by other opioids, accidental (unintentional), initial encounter	T40.2X1A
Poisoning by other opioids, accidental (unintentional), subsequent encounter	T40.2X1D
Poisoning by other opioids, accidental (unintentional), sequela	T40.2X1S
Poisoning by other opioids, intentional self-harm,	T40.2X2
Poisoning by other opioids, intentional self-harm, initial encounter	T40.2X2A
Poisoning by other opioids, intentional self-harm, subsequent encounter	T40.2X2D
Poisoning by other opioids, intentional self-harm, sequela	T40.2X2S
Poisoning by other opioids, assault	T40.2X3
Poisoning by other opioids, assault, initial encounter	T40.2X3A
Poisoning by other opioids, assault, subsequent encounter	T40.2X3D
Poisoning by other opioids, assault, sequela	T40.2X3S
Poisoning by other opioids, undetermined	T40.2X4
Poisoning by other opioids, undetermined, initial encounter	T40.2X4A
Poisoning by other opioids, undetermined, subsequent encounter	T40.2X4D
Poisoning by other opioids, undetermined, sequela	T40.2X4S
Poisoning by methadone, accidental (unintentional)	T40.3X1
Poisoning by methadone, accidental (unintentional), initial encounter	T40.3X1A
Poisoning by methadone, accidental (unintentional), subsequent encounter	T40.3X1D
Poisoning by methadone, accidental (unintentional), sequela	T40.3X1S
Poisoning by methadone, intentional self-harm, initial encounter	T40.3X2
Poisoning by methadone, intentional self-harm, initial encounter	T40.3X2A
Poisoning by methadone, intentional self-harm, subsequent encounter	T40.3X2D
Poisoning by methadone, intentional self-harm, sequela	T40.3X2S
Poisoning by methadone, assault	T40.3X3
Poisoning by methadone, assault, initial encounter	T40.3X3A
Poisoning by methadone, assault, subsequent encounter	T40.3X3D
Poisoning by methadone, assault, sequela	T40.3X3S
Poisoning by methadone, undetermined	T40.3X4
Poisoning by methadone, undetermined, initial encounter	T40.3X4A
Poisoning by methadone, undetermined, subsequent encounter	T40.3X4D
Poisoning by methadone, undetermined, sequela	T40.3X4S
Poisoning by other synthetic narcotics, accidental (unintentional)	T40.4X1
Poisoning by other synthetic narcotics, accidental (unintentional), initial encounter	T40.4X1A
Poisoning by other synthetic narcotics, accidental (unintentional), subsequent encounter	T40.4X1D
Poisoning by other synthetic narcotics, accidental (unintentional), sequela	T40.4X1S
Poisoning by other synthetic narcotics, intentional self-harm	T40.4X2
Poisoning by other synthetic narcotics, intentional self-harm, initial encounter	T40.4X2A
Poisoning by other synthetic narcotics, intentional self-harm, subsequent encounter	T40.4X2D

Poisoning by other synthetic narcotics, intentional self-harm, sequela	T40.4X2S
Poisoning by other synthetic narcotics, assault	T40.4X3
Poisoning by other synthetic narcotics, assault, initial encounter	T40.4X3A
Poisoning by other synthetic narcotics, assault, subsequent encounter	T40.4X3D
Poisoning by other synthetic narcotics, assault, sequela	T40.4X3S
Poisoning by other synthetic narcotics, undetermined	T40.4X4
Poisoning by other synthetic narcotics, undetermined, initial encounter	T40.4X4A
Poisoning by other synthetic narcotics, undetermined, subsequent encounter	T40.4X4D
Poisoning by other synthetic narcotics, undetermined, sequela	T40.4X4S
Poisoning by opioids	965.0
Poisoning by opium (alkaloids), unspecified	965.00
Poisoning by heroin	965.01
Poisoning by methadone	965.02
Poisoning by other opiates and related narcotics	965.09
Accidental poisoning by heroin	E850.0
Accidental poisoning by methadone	E850.1
Accidental poisoning by other opiates and related narcotics	E850.2

- Define the list of AUD codes

Alcohol abuse, uncomplicated	F10.10
Alcohol abuse, remission	F10.11
Alcohol abuse with intoxication, uncomplicated	F10.120
Alcohol abuse with intoxication delirium	F10.121
Alcohol abuse with intoxication, unspecified	F10.129
Alcohol abuse with alcohol-induced mood disorder	F10.14
Alcohol abuse with alcohol-induced psychotic disorder with delusions	F10.150
Alcohol abuse with alcohol-induced psychotic disorder with hallucinations	F10.151
Alcohol abuse with alcohol-induced psychotic disorder, unspecified	F10.159
Alcohol abuse with alcohol-induced anxiety disorder	F10.180
Alcohol abuse with alcohol-induced sexual dysfunction	F10.181
Alcohol abuse with alcohol-induced sleep disorder	F10.182
Alcohol abuse with other alcohol-induced disorder	F10.188
Alcohol abuse with unspecified alcohol-induced disorder	F10.19
Alcohol dependence, uncomplicated	F10.20
Alcohol dependence, in remission	F10.21
Alcohol dependence with intoxication, uncomplicated	F10.220
Alcohol dependence with intoxication delirium	F10.221
Alcohol dependence with intoxication, unspecified	F10.229
Alcohol dependence with withdrawal, uncomplicated	F10.230
Alcohol dependence with withdrawal delirium	F10.231
Alcohol dependence with withdrawal with perceptual disturbance	F10.232
Alcohol dependence with withdrawal, unspecified	F10.239
Alcohol dependence with alcohol-induced mood disorder	F10.24
Alcohol dependence with alcohol-induced psychotic disorder with delusions	F10.250

Alcohol dependence with alcohol-induced psychotic disorder with hallucinations	F10.251
Alcohol dependence with alcohol-induced psychotic disorder, unspecified	F10.259
Alcohol dependence with alcohol-induced persisting amnestic disorder	F10.26
Alcohol dependence with alcohol-induced persisting dementia	F10.27
Alcohol dependence with alcohol-induced anxiety disorder	F10.280
Alcohol dependence with alcohol-induced sexual dysfunction	F10.281
Alcohol dependence with alcohol-induced sleep disorder	F10.282
Alcohol dependence with other alcohol-induced disorder	F10.288
Alcohol dependence with unspecified alcohol-induced disorder	F10.29
Alcohol withdrawal delirium	291.0
Alcohol withdrawal	291.81
Acute alcoholic intoxication in alcoholism, unspecified	303.00
Acute alcoholic intoxication in alcoholism, continuous	303.01
Acute alcoholic intoxication in alcoholism, episodic	303.02
Acute alcoholic intoxication in alcoholism, in remission	303.03
Other and unspecified alcohol dependence	303.9
Other and unspecified alcohol dependence, unspecified	303.90
Other and unspecified alcohol dependence, continuous	303.91
Other and unspecified alcohol dependence, episodic	303.92
Other and unspecified alcohol dependence, in remission	303.93
Alcohol abuse	305.0
Alcohol abuse, unspecified	305.00
Alcohol abuse, continuous	305.01
Alcohol abuse, episodic	305.02
Alcohol abuse, in remission	305.03

- Drop naltrexone injections in the pre randomization period if there are not 2+ visits with an OUD/OD diagnosis (can be 1 OUD and 1 OD code) in the pre randomization period.
- Drop naltrexone injections in the post randomization period if there are not 2+ visits with an OUD/OD diagnoses (can be 1 OUD and 1 OD code) in the pre or post randomization period.
- Adjudicate subjects with both 2+ OUD/OD and 1+ AUD diagnosis codes to decide whether to include or exclude as OUD treatment.

16) Using the start and end dates of **each bup (including probuphine and sublocade) and naltrexone order or procedure** (and adding dispensed buprenorphine and/or methadone in a sensitivity analysis), create continuous episodes of use. This will result in a start and end date for each continuous OUD treatment. Continuous is defined as <=7 days gap between end date and subsequent start date or <= 14 days gap if bup end date and XR injectable naltrexone start date (allow for washout period prior to starting XR injectable naltrexone). There may be multiple continuous episodes for any given subject because there will be gaps (see below) or breaks in continuous treatment.

17) <u>Determine treatment day</u>

a. Order all of a patient's orders and procedures by start and end dates. This should include rx and procedures 180 days before day 1 of the time period of interest because these may run out into the time period of interest.

b. For example: Within a period of interest starting on 1/1/20, a bup order occurring on 12/25/20 for 30 days (end date 1/24/20) would contribute 20 treatment days to January.
c. During time periods of interest, consider it a day "covered" by buprenorphine or injectable naltrexone during the period of interest if there is bup or naltrexone that covers that day based on rx start and end date. Do not double count if >1 bup or naltrexone on the same day.

Estimate begin and stop dates for each continuous episode (continuous defined as a gap ≤ 7 days between end of one bup rx and start of the next bup rx; <=14 days between bup end and naltrexone start, and <=7 days between naltrexone end and bup or naltrexone start) for all the bup and naltrexone in the time period of interest. In other words, the <=14 rule may only be applied when the last day before the gap is covered by bup only and the first day after the bup gap is covered by naltrexone only.

a. Treatment days = sum of all days covered by bup or naltrexone during the period of interest. No overlap or double counting of treatment days is allowed. In other words, a patient can have a max of 365 days covered by treatment in a year.
    i. Left and right truncate when time period begins and ends
        1. For example: Bup order that covers 12/15/2019 to 1/14/220 would stop contributing treatment days on 12/31/2019 if the period of interest ends on 12/31/2019.

Sensitivity analyses (SA): The above steps will be replicated for

1) A SA that only includes episodes with at least 2 buprenorphine orders omitting last order (and any refills on the order) in an oral buprenorphine treatment episode
2) A SA that varies the allowable 7-day gap of continuous use to a lower threshold (based on distribution of the gaps; example: 2 days).
3) A SA that uses a combination of orders and dispensings at health systems with some or all (one health system) dispensings data.

For each dispensing, if its rxdate is in "the range" of an order date or a refill date, then it is considered to be linked to that order or refill and dropped in calculations; otherwise, this dispensing is considered as not linked to any order and added to the calculation of days covered. The range is defined as dispensed within 30 days after the order/refill's start date (an order's start date is the order's date; a refill's start date is derived from the original order's date and days supply).

This calculation will be done after all data cleaning and imputation.

1) A SA that adds in procedure codes for methadone maintenance therapy at sites where this data is available
2) A SA that combines SA # 3 and #4 at sites with both dispensing and methadone maintenance therapy data.

A3. Combine OUD Treatments of Buprenorphine and Naltrexone

1) Order all OUD treatments by their respective start and runout dates. See Examples below.
2) During time periods of interest, consider it a day "covered" by OUD treatment during the period of interest if there is buprenorphine or injectable naltrexone that covers that day based on start and runout dates. Do not double count covered days.
3) Estimate start and end dates for each continuous episode (continuous defined as a gap ≤ 7 days between runout date of one OUD treatment and start date of the next

treatment OR ≤14 days between end date of one buprenorphine prescription and start date of naltrexone) based on days covered for all the buprenorphine and naltrexone in the time period of interest.

4) Any gaps not meeting these criteria above result in the end of a continuous episode and these gaps between episodes are NOT part of any episode and do not contribute to treatment days. See examples below.
5) Pre and post randomization episodes are not allowed to overlap. They will be left and right censored accordingly. For example, a treatment episode of 90 days with 15 days in the post randomization period (includes randomization date) will be split into two episodes with one episode contributing 75 days of treatment pre-randomization and a second episode contributing 15 days of treatment post randomization.
6) OUD treatment days = sum of all episodes [(episode end date – episode start date) + 1)] in the pre and post randomization periods.

Examples 1-8 below are each a single patient's prescriptions. These examples are shortened to a 20-day period of interest as an example only. Our perspective here is the prescribers' "intent to treat" (i.e., we do not know how or if patients actually took the oral medications but the assumptions we are making are typical of studies that use electronic health data for studying medication treatment and adherence). (Bup=buprenorphine; NTX=naltrexone)

Example 1: Bup	**TIME PERIOD OF INTEREST**																				
	-1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Bup1																					
Bup2																					
Bup3																					
Bup4																					
Bup5																					

Bup 1 covers day 1 (from a rx occurring during -90 days of period of interest as depicted by coverage on -1 day) and ends day 8; Bup 2 starts day 8 and ends day 11; Bup 3 starts day 12 and ends day 15; Bup 4 starts day 15 and ends day 16; Bup 5 starts day 19 and covers day 19 and 20 – may end on day 20 or after period of interest (truncated).

One continuous episode: Day 1 to Day 20. The gap (day 17 and day 18) is 2 days (≤7 days) so smooth it to be continuous.

20 of the 20 day time period of interest is covered by treatment.

Example 2: Bup	**TIME PERIOD OF INTEREST**																			
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Bup1																				
Bup2																				
Bup3																				
Bup4																				
Bup5																				

Bup 1 covers day 1 and ends day 6; Bup 2 starts day 1 and ends day 11; Bup 3 starts day 6 and ends day 8; Bup 4 starts day 15 and ends day 16; Bup 5 starts day 16 and ends day 19.

One continuous episode: Day 1 to Day 20. The gap is 4 days (<=7 days) so smooth it to be continuous.

Example 3: Bup & NTX	**TIME PERIOD OF INTEREST**																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Bup1																				
NTX1																				
Bup 1 covers day 1 and ends day 4; NTX1 starts day 10 and covers through day 20 (right truncated since end date would be past time period of interest of 20 days). One continuous episode: Day 1 to Day 20. The gaps in treatment (day 5-9) are ≤14 days. 20 of the 20 day time period of interest are covered by treatment.																				

Example 4: Bup & NTX	**TIME PERIOD OF INTEREST**																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
NTX1																				
Bup1																				
NTX1 1 covers day 1 and ends day 9 (left truncated because it must have begun prior to day 1 given inj ntx lasts 28 days); Bup1 starts day 18 and covers though day 20 (likely right truncated) in the period of interest. Two continuous episodes: 1st episode is Day 1 to Day 9. Then there is a gap >7 days so a new episode begins. 2nd episode starts day 18 and ends day 20. 12 of the 20 day time period of interest are covered by treatment.																				

Example 5: Bup	**TIME PERIOD OF INTEREST**																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Bup1																				
Bup2																				
BP1 1 covers day 1 and ends day 9 (possibly left truncated if it began prior to day 1; Bup2 starts day 7 and covers though day 17. One continuous episode: Day 1 to Day 17. No double counting of overlap days (7-9). 17 of the 20 day time period of interest are covered by treatment.																				

Example 6: Bup & NTX	**TIME PERIOD OF INTEREST**																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Bup1																				
NTX1																				
NTX2																				
BP1 1 covers day 1 and ends day 4 (possibly left truncated if it began prior to day 1; NTX1 also covers day 1 and ends day 4 (left truncated); NTX2 starts day 10 and ends day 20 (right truncated) One continuous episode: Day 1 to Day 20. The gap is 5 days (<=7) so smooth it to be continuous.																				

Example 7: Bup & NTX	**TIME PERIOD OF INTEREST**																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Bup1																				
NTX1																				
NTX2																				
BP1 1 covers day 1 and ends day 4 (possibly left truncated if it began prior to day 1; NTX1 covers day 1 and ends day 3 (left truncated); NTX2 starts day 10 and ends day 20 (right truncated). One episode: Day 1 to Day 20. No double counting of overlap in days between bup1 and ntx1. There is <=14 day gap between when bup1 ends (last drug before NTX2) and when NTX2 starts so it is continuous. 20 of the 20 day time period of interest are covered by treatment.																				

Example 8: Bup & NTX	**TIME PERIOD OF INTEREST**																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Bup1																				
NTX1																				
NTX2																				
BP1 1 covers day 1 and ends day 4 (possibly left truncated if it began prior to day 1; NTX1 covers day 1 and ends day 5 (left truncated); NTX2 starts day 10 and ends day 20 (right truncated). One continuous episode: Day 1 to Day 20. The gap is 4 days (<=7) so smooth it to be continuous.																				

Appendix B. Initial Values for Power Simulations

The following is a justification of the initial values used in Step 2 of the data generation model for the Primary Objective 1 Outcome power simulations.

Consider a lag-1 autoregressive time series given by

$$y_{t+1} = \xi + \psi\, y_t + e$$

where ξ and ψ are constants, *e* is normally distributed with mean zero and variance σ_e^2 and *t* indexes time interval Letting η denote the mean of any particular y_t, and ς the variance, then we have

$$\begin{gathered} E(y_{t+1}) = \xi + \psi\, E(y_t) \\ \Rightarrow \eta = \xi + \psi\, \eta \\ \Rightarrow \eta = \frac{\xi}{1-\psi} \end{gathered}$$

and

$$\begin{gathered} Var(y_{t+1}) = \psi^2\, Var(y_t) + Var(e) \\ \Rightarrow \varsigma = \psi^2\, \varsigma + \sigma_e^2 \\ \Rightarrow \varsigma = \frac{\sigma_e^2}{1-\psi^2}. \end{gathered}$$

Appendix C. Small-sample Degree of Freedom (DF) Correction Method for Objective 2 Primary Analysis

Recent literature has highlighted the need to incorporate small-sample correction methods when analyzing data from cluster-randomized trials (CRTs) when there are few (e.g., <30) clusters (Kahan et al. 2016; Leyrat et al. 2017). In particular, it has been documented that when there are few clusters the usual approaches to analyzing correlated data in CRTs, including generalized estimating equations (GEE) and generalized linear mixed models (GLMM; the proposed analysis approach), can yield inflated type 1 error rates (Kahan et al. 2016). For example, applying a traditional test based on these models that has a nominal type 1 error rate of 5%, may lead to actual type 1 error rates that are much larger (e.g., 10-20%) such that we would reject the null hypothesis that the intervention is effective, even if it is truly not effective, above the pre-specified acceptable level.

Although corrections for tests of treatment effects under GLMM with few clusters have been studied, there remain gaps in knowledge of which test to use in the context of PROUD Objective 2 analyses. In particular, although it is common to adjust for covariates (such as the baseline value of the outcome as in PROUD analyses), covariate-adjusted analyses have not received attention in existing literature. Moreover, despite extensive research on corrections when the outcome is continuous or binary, to our knowledge the performance of the methods for count outcomes (such as the number of days of acute care utilization) has not been studied.

To address these gaps in the statistical literature (and to gain insight on how to account for the small number of clusters for PROUD Objective 2 analyses), we investigated the performance of various testing procedures for GLMM via simulation. The tests we considered were all eight combinations of two forms of test and four methods to compute the denominator degree of freedom (DDF). The forms of test we considered were the Wald t-test and the likelihood ratio F-test (LRT). The methods to compute DDF we considered were residual, containment, between-within 1 (BW1), and between-with 2 (BW2) (also referred to as “inner-outer”). BW1 and BW2 are two generalizations of the BW method to covariate-adjusted models, and BW has been shown to perform well via various scenarios in simulation (Li & Redden 2015).

We conducted simulations for count outcomes (i.e., Poisson GLMM) with various numbers of clusters, mean and coefficient of variation (CV) of clusters sizes, and intraclass correlation coefficients (ICC) characterizing the correlation of patients within a cluster. We also vary the number and the level (individual- vs. cluster-level) of (i) prognostic variables in the data-generating model, and (ii) extra (i.e., non-prognostic) covariates in the correctly specified fitted model. We consider including extra covariates to account for the potential that non-prognostic covariates may be included in secondary analyses. Additional details on the simulation study setup are described in a manuscript in preparation and are available upon request.

Selected Results

Here we present selected simulation study results that correspond to the settings of the primary Objective 2 analysis that adjusts for a single covariate (baseline value of the outcome), as well as for the secondary analysis that adjusts for additional person-level covariates. For these settings, the GLMM fitted to the data was the correctly specified model (i.e., the model did not include redundant covariates).

Figure C1. Type 1 Error Rates from the Setting with Data-generating Model that Includes a Single Person-level Covariate (setting for primary Objective 2 analysis that adjusts solely for the baseline value of the outcome)

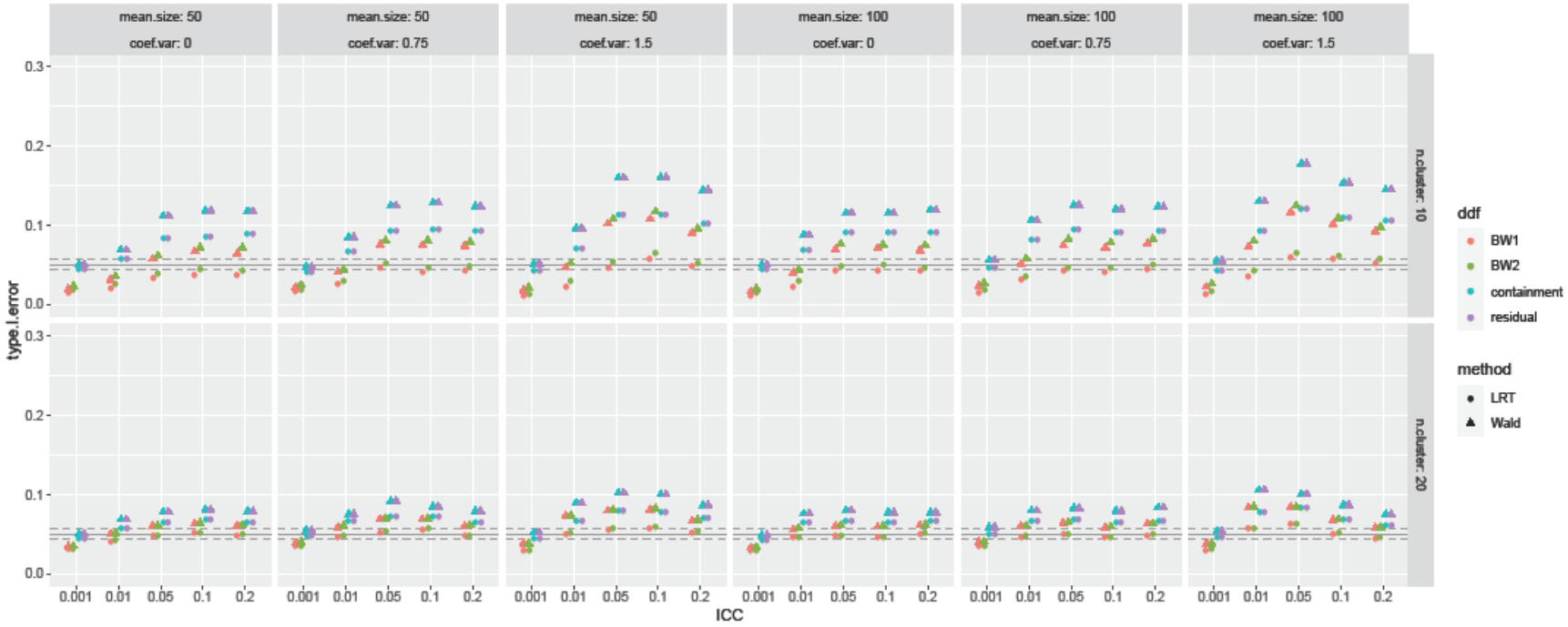
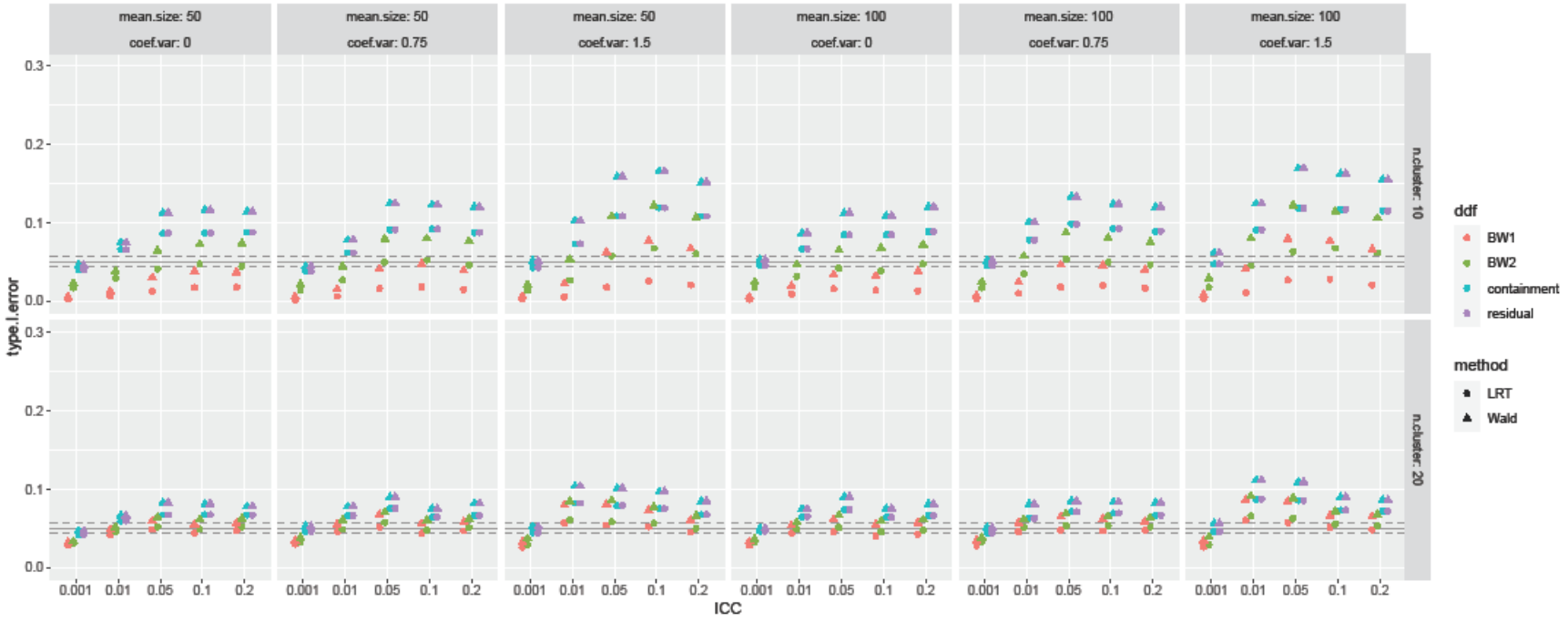


Figure C2. Type 1 Error Rates from the Setting with Data-generating Model that Includes Multiple Person-level Covariate (setting for sensitivity Objective 2 analysis that adjusts for multiple covariates in addition to the baseline value of the outcome)



Implications for PROUD Objective 2 Analyses

None of the tests considered performed uniformly well across all data generating scenarios considered, and the optimally performing test (i.e., with type 1 error rate closest to the nominal 0.05 level), varied highly depending on the scenario.

With the Objective 2 analysis clustered at the clinic level (see Section 5.3), the scenario that most closely aligns with our setting, based on PROUD Phase 1 data, is given in the top-right plot of the two Figures above. (For Phase 1, the ICC under a linear mixed model was approximately 0.02, average sample size within a cluster was approximately 100, and CV of cluster sizes was 1.2.)

For the results that align with the covariate adjustment setting under the primary analysis approach (Figure C1), the LRT with either BW1 or BW2 have reasonable type 1 error rates. For the results that align with the covariate adjustment setting under the secondary analysis approach (Figure C2), LRT with BW2 performs reasonably well (whereas with BW1 performance was overly conservative). We therefore selected LRT with BW2 (also referred to as “inner-outer”) for the testing approach that incorporates small sample methods.