PROTOCOL AND STATISTICAL ANALYSIS PLAN

Study Title: **Achieving Patient-Centered Care and Optimized Health In Care Transitions by Evaluating the Value of Evidence (ACHIEVE)**

Institution/Site:	University of Kentucky
Document (Approval/Update) Date:	December 20, 2017
NCT Number:	NCT02354482
IRB Number	14-0789-F3R
Coversheet created:	09/30/2019

II. Prospective Analysis, Patient survey

The prospective analysis will use detailed measures collected directly from patients about the types of TC components they receive and experience during care episodes. This analysis will define and classify TC clusters (comparators) on a much more granular level. However, to maximize the study's power to detect the individual and combined effects of TC components and to realize efficiencies in primary data collection, a Fractional Factorial (FF) design will be used to select fractions (subgroups) of the TC clusters and care settings in which to collect detailed, prospective patient data.

II.A. Eligibility and Target Population for Prospective Study, Patient survey

Site Selection Recognizing hospital and community contexts do not remain static, we will develop a purposively selected hospital list for recruitment based on how TC improvement efforts were selected, what components were deployed and whether they were modified based on the local delivery system, the nature of the population served, and other contextual factors, to ensure representation of: 1) urban and rural areas; 2) safety-net; 3) critical access; 4) integrated delivery system; 5) involvement in care delivery demonstrations (e.g., ACO, BPCI). Since lists of discharged patients need to be provided by hospitals, the CBOs and downstream providers will be recruited using snowball methodology. The purposive hospital sampling will consider the HEN, QIO ICPC, and CCTP participation. The CBO selection will be based on CCTP participation and input from n4a. There are many hospital demographics that we can use for sampling. After reviewing AHA hospital survey file layout and CMS IPPS impact file, the recruitment and engagement and data collection and management workgroups went through multiple cycles of refining sampling variables. Appendix B shows the primary and secondary sampling variables.

Patient Population Project ACHIEVE will focus on Medicare fee-for-services beneficiaries and study diverse high risk patient populations, including those with: 1) multiple chronic conditions; 2) mental health issues; 3) rural area domicile; 4) limited English proficiency or low health literacy; 5) low socioeconomic status; 6) Medicare and Medicaid dual eligible; 7) disabled and younger than 65.

II.B. Selection of Factors and Levels for Prospective Data Collection

The sampling frame for the prospective FF design includes the full range of TC clusters along with the full range of hospital and community settings and patient subgroups in which these clusters are implemented. Results from the retrospective analysis will be used together with results from the hospital site visits and responses of Organizational TC Implementation Assessment (OTCIA) form to identify which combinations of TC clusters, care settings, and patient subgroups are most important to study in detail. A FF design matrix will be

constructed that lists all TC components and all important organizational, community, and patient characteristics as possible factors and interactions. The organizational, community and patient characteristics to include in the matrix will be based on the retrospective analysis results showing which cluster/setting/patient interactions achieved significant improvements in outcomes. Results from the OTCIA will be used to determine which cells can be dropped from the design matrix altogether (given zero weight) because the specific TC cluster/setting/patient combination does not exist in practice. Other cells will be assigned increased weight because they include clusters that are highly valued by patients and caregivers, or because the retrospective analysis shows that the cell is associated with improved outcomes. From this design matrix, a subset of comparators will be selected that maximize the study's ability to detect unconfounded main effects and interaction effects involving TC components, care settings, and patient subgroups.

Based on our team's existing knowledge of and experience with the organized TC programs and their TC components, we anticipate that it will be necessary to include at least 6 TC clusters in the prospective FF design with approximately 5 hospitals per cluster in order to adequately characterize the existing heterogeneity in TC program implementation, care settings, and patient subgroups (Figure 1). The exact number of clusters and hospitals per cluster will be determined by analysis of the FF design matrix. Surveys will be fielded with a total of 46 hospitals and an average of 300 patients per hospital.

In the event that an insufficient number of TC/setting/patient combinations are available to utilize the fractional factorial design for a subpopulation, ***data visualization*** will be used to identify potential comparators for ad-hoc analyses and descriptions. This more ad-hoc approach has potential analytical issues if one hospital contributes a significant portion of the patients for a subpopulation, and separating the TCC/cluster from the hospital-effect may be difficult. Therefore, in the selection of hospitals, availability of subpopulations and presence of particular TCs will be carefully considered; obtaining a sample of hospitals based on TCs *and* needed combinations of subpopulations per the fractional factorial design will avoid the potential for analyzing and describing patients within a single hospital.

II.C. Study Outcome Measures

Patient utilization of health services include 30-day ED visits, 30-day rehospitalizations

II.D. Development and Validation of the Patient Survey Instruments

Westat will develop a draft survey instruments consisting of items designed to elicit patient experiences with the delivery and receipt of TC components, as well as patient-reported outcomes relevant to transitions in care.

Draft items will be derived from TC measurement constructs identified from an updated literature review, existing validated instruments such as the Coleman CAHPS Care Transition questions, items listed in the NIH PROMIS repository, and constructs elicited from focus groups and hospital site visits. Draft instruments will be refined based on at least two cycles of iterative review by members of the ACHIEVE study team, including members with relevant clinical expertise, TC program development and implementation expertise, and patient and caregiver experience.

Once draft items have been developed, cognitive interviews will be conducted to examine item clarity, comprehension, difficulty, and face validity. Cognitive interviews will be conducted with up to 14 participants in two iterative rounds . The patient survey will also be translated into Spanish and up to 6 cognitive interviews will be conducted to pretest the translated items. The surveys will be revised as needed based on feedback from the cognitive interviews, in preparation for pilot testing.

Significant changes were required in the administration of the patient and caregiver survey component of Project ACHIEVE based on uncovering potential conflict with CMS HCAHPS. Numerous follow-up conversations with CMS, HSAG, PCORI and the ACHIEVE research team yielded agreement on Project ACHIEVE being allowed to field its patient survey 51 days after patient hospital discharge; after most patients contacted for HCAHPS have completed their survey and without causing significant disruption to hospitals' participation in HCAHPS. Of note, we originally proposed to call patients between 30 and 45 days post discharge.

Using the data from the survey pilot, item analysis and internal consistency reliability tests will be conducted to determine if there are items that should be dropped and/or to identify items that are not operating as expected. Item analysis will consist of examining item frequencies and missing percentages to ensure that there is item variability and low levels of missing data when the item should have been answered. High levels of missing data may signal either an issue with the way the item is worded or an inappropriate topic for the population. On the patient and caregiver surveys, telephone interviewer comments will also be considered to determine if certain items were problematic or unclear for respondents. Internal consistency reliability tests will be conducted on the a priori theoretical constructs of interest within the survey using a threshold of a Cronbach's alpha of at least 0.70 to be considered acceptable. Item grouping and inclusion will be refined based on the small-scale pilot validation results in order to improve reliability and validity of item scales.

After completion of survey validation work from the pilot tests, the survey instruments will be revised prior to the main survey data collection.

The ACHIEVE survey pilot study was conducted from December 6 2016 to April 30, 2017, with **132** completed patient surveys, **38%** overall response rate (42% mail with phone follow up; 30% phone only).

The pilot survey response rates were lower than estimated in original proposal. Literature demonstrates that a multi-wave, mail survey with telephone follow-up yields a better response rate than a phone-only protocol. Further, the addition of an incentive, whether promised or pre-paid, will also improve response rates, compared to a design without any incentive. Of note, we are unable to promise an incentive to hospitalized patients when they are approached about Project ACHIEVE per guidance from CMS. Proposed changes to the data collection protocol for patients:

1. Initial survey packet mailed to all patients, including $5 cash incentive payment in the envelope

2. Thank You reminder post card (mailed to all)
3. 2nd survey packet mailed to those who do not respond to the first survey
4. Telephone follow-up with those who do not respond to either mail survey

Before the main survey administration, using pilot survey estimates, we expected to approach and obtain HIPAA authorization from 27,000 patients (45 participating hospitals approaching an estimated average of 20 patients per hospital per week during the 30-week recruitment period), and have 12,150 completed patient survey (a 45% response rate). Since launching the main survey administration in June 2017, several factors have conspired to reduce the expected patient and caregiver recruitment from participating hospitals. Therefore, we propose to reduce the target number of completed patient surveys from 12,000 to 9,000 and extended the recruitment end date from Jan 21, 2018 to April 15, 2018 for adequate sample size to detect the comparative effects of transitional care strategies on important patient outcomes of interest.

II.E. Response Rates, Missing Values and Data Imputation on the Survey Data Collection Datasets

Response rates for the main data collection will be calculated using AAPOR response rate formula 4 (RR4).[6] This formula includes both complete and partial completed surveys/interviews for inclusion in the numerator and allocates cases of unknown eligibility in the denominator. This is particularly important given that the phone number data may be incomplete, incorrect, or phone numbers may be disconnected and we therefore may not be able to know the eligibility of the case. We will examine survey refusals and nonresponse to determine if there is the potential for nonresponse bias based on certain patient subgroups and/or hospital subgroups and make corrections if possible.

For item nonresponse, we will examine if certain items have higher missing data than others. Given that the patient surveys will be designed to assess experiences with care, if the respondent does not indicate that they experienced a specific care component, we will not impute data on their experience with it.

II.F. Analysis of Patient Survey Datasets

Psychometric analysis. Westat will compile patient survey data from the main study, conduct data cleaning, and generate descriptive statistics from the survey results. Psychometric analyses will be conducted to examine items for variability of response, missing percentages, calculate reliability statistics, perform confirmatory factor analysis to examine the factor structure of the measures, and examine inter-correlations of items and composites. Final composite scores will be created and the survey data will then be used to examine the study's aims and hypotheses regarding TC delivery and outcomes.

Descriptive statistics. Continuous variables will be summarized with descriptive statistics (n, mean, standard deviation, median, quintiles, and min and max); categorical variables will be described with counts and percentages. All data will be unweighted. For each survey type

(i.e., patient, caregiver, provider), numerical and/or graphical summaries will be provided overall and by subpopulations.

II.G. Analysis of TC Component Exposure. Analyses will be conducted using both hospital- and patient/caregiver-level data. Comparisons on outcomes will be made first using TCCs identified by the hospital; additional analyses will use patient/caregiver identified components. Moreover, additional analyses will also use the number and occurrence of specific categories received within a component as a measure of the TCC received. Finally, an additional issue for consideration is that indicators for TCC presence or absence may not be sufficient for identifying TC component "exposure." It is expected that TCCs in different communities/hospitals will be implemented to varying degrees. Non-uniform implementation may provide a natural experiment where communities/hospitals serve as their own control and the longitudinal effects of the TC component can be investigated. Using process and implementation data collected at Phase 2, segmented regression analysis of interrupted time series data will allow for comparisons and the estimation of effects for different intensities of TC component exposure.

Although every effort will be made to include the possible combinations of TC components, it is unlikely that all combinations will be present in each subpopulation. Additionally, although many TC component factors will be measured, it is not expected that all of them will impact the outcomes (sparsity of effects). As an alternative to the factorial design utilized in an experiment, fractional factorials may also be utilized, where a large number of factors can be "screened" to detect important factors even in the presence of a small number of hospitals.

As an additional visualization tool, tree-based models (CART and Random Forests) may also be used to identify potential higher-order interactions. Although the primary analysis will be conducted within a subpopulation, exploratory analyses will be used to investigate the interaction of subpopulation characteristics, patient/caregiver characteristics, and TC components on outcomes. The identification of these potential combinations will help inform additional comparative analyses in the presence of a large number of potentially correlated variables.

- **<u>Pooled Analysis</u>**
  Based on the statistical power available with a survey response rate of 35 percent, we conclude that the resulting sample size of approximately 9000 patients from 46 hospitals represents the minimum sample required to ensure that the ACHIEVE prospective analysis can detect the comparative effects of transitional care strategies on important patient outcomes of interest. Our power analysis is based on the study's ability to detect meaningful differences in outcomes across groups of hospitals that implement different combinations of transitional care (TC) components. Results from the ACHIEVE survey pilot test indicate that the patient survey response rate could be as low as 35 percent. To assess the impact of this response rate on statistical power, we conducted a Monte Carlo simulation using data from the ACHIEVE retrospective analysis, which includes Medicare claims data from nearly 400 hospitals and all Medicare patients hospitalized at these facilities

over five years. This simulation allowed us to assess the likely statistical power of the prospective analysis even though we do not yet know with certainty the distribution of TC components and patients across participating hospitals. This simulation assumes that the distribution of TC components and patients among hospitals recruited for the ACHIEVE prospective study will be consistent with the distributions observed in the larger population of hospitals included in the ACHIEVE retrospective study.

We conducted the simulation as follows: (1) we drew a random sample of 46 hospitals from the nearly 400 hospitals included in the ACHIEVE retrospective data; (2) we drew a random sample of 560 patients from each of the 46 selected hospitals, corresponding with the planned data collection period of 28 weeks and the planned recruitment rate of 20 patients per week at each facility; (3) we drew a random sample of 35 percent of the recruited patients at each hospital, corresponding with the assumed minimum response rate; (4) we used the 5 clusters of TC components identified from the ACHIEVE retrospective analysis to classify hospitals into comparison groups based on which TC components they implemented for their patients; and (5) we used a multivariable logistic regression model to estimate minimum detectible differences (using a 5% significance level) in 30-day all-cause readmission rates across hospitals grouped by TC clusters, while controlling for patient characteristics that may impact readmission (e.g., age, race, sex, comorbidities, and dual eligible status, as well as others), and accounting for patient clustering within hospitals. We used 1000 replications of this simulation to estimate the variances in minimum detectible differences.

Simulation results indicate that at an average patient response rate of 35 percent, the study will have sufficient power to detect differences in readmission rates of 2.1 to 2.9 percentage-points across the TC clusters at p=0.05. This level of statistical power will allow the study to detect effect sizes that are somewhat smaller than the effects estimated from prior studies of care transition interventions, such as the 3.6 percentage-point reduction in readmissions observed in Eric Coleman's care transition intervention trial. Because the ACHIEVE analysis compares outcomes across multiple active treatment groups defined by TC clusters—rather than comparisons between a single treatment group and control group—and because readmission rates have been declining nationally since 2012, it is important for the ACHIEVE study to be powered to detect somewhat smaller effect sizes than those found in previous two-group studies. For this reason, we consider a response rate of 35 percent to be the minimum acceptable response rate for the ACHIEVE prospective study. This would be approximately 9000 patients.

- **Subgroup Analysis**

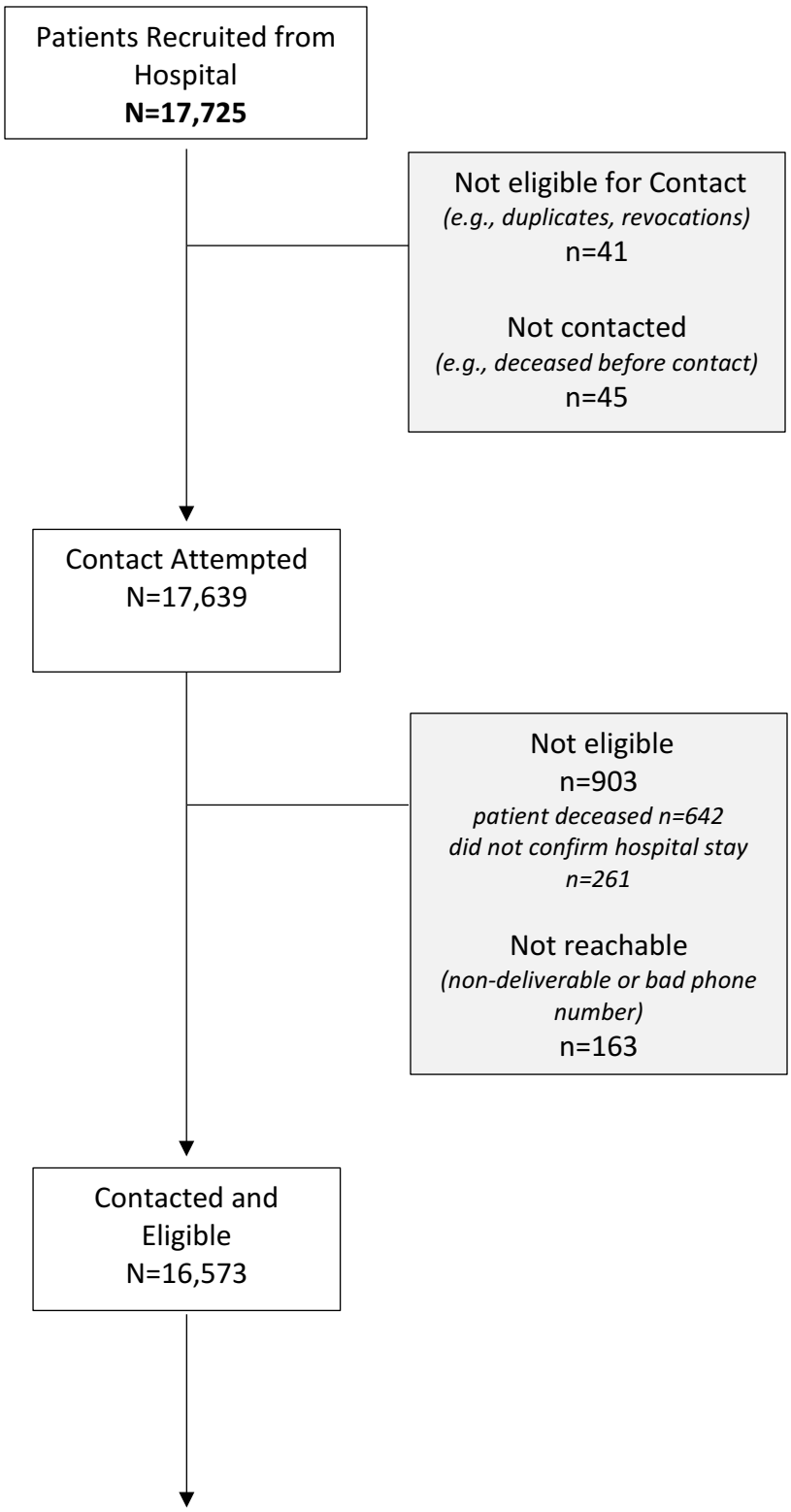
  Several patient subgroups and community settings are of special interest in the ACHIEVE analysis, including patients with multiple comorbidities, dual eligibles, low-literacy populations, and rural communities. It is not possible to estimate minimum sample sizes for subgroup analysis with any reasonable level of precision given the lack of a priori information about how patient covariates and TC clusters

are distributed within and across these subgroups. Our simulation analyses suggest that the minimum effect sizes we will be able to detect for subgroups could be more than two times larger than the 2.1 to 2.9 percentage points estimated for the pooled analysis as a whole, when using traditional multivariable analysis methods. For this reason, we anticipate using a modified analytical approach that relies on Bayesian estimation methods that borrow statistical power from the full ACHIEVE sample in order to support statistical inferences about subgroup differences. This method has been used successfully in several recent federal studies of similar magnitude and scope, including the 2016 Mathematica Policy Research evaluation of the CMS Comprehensive Primary Care Initiative. The primary limitation of Bayesian estimation with small subgroup sample sizes is that results can be sensitive to incorrect assumptions about the prior probability distributions of the effect sizes. A major strength of the ACHIEVE study is that we can use estimated effects from the ACHIEVE retrospective analysis to specify informative prior probability distributions for the effect sizes to be estimated in the prospective analysis, rather than basing these distributions on uninformed assumptions.

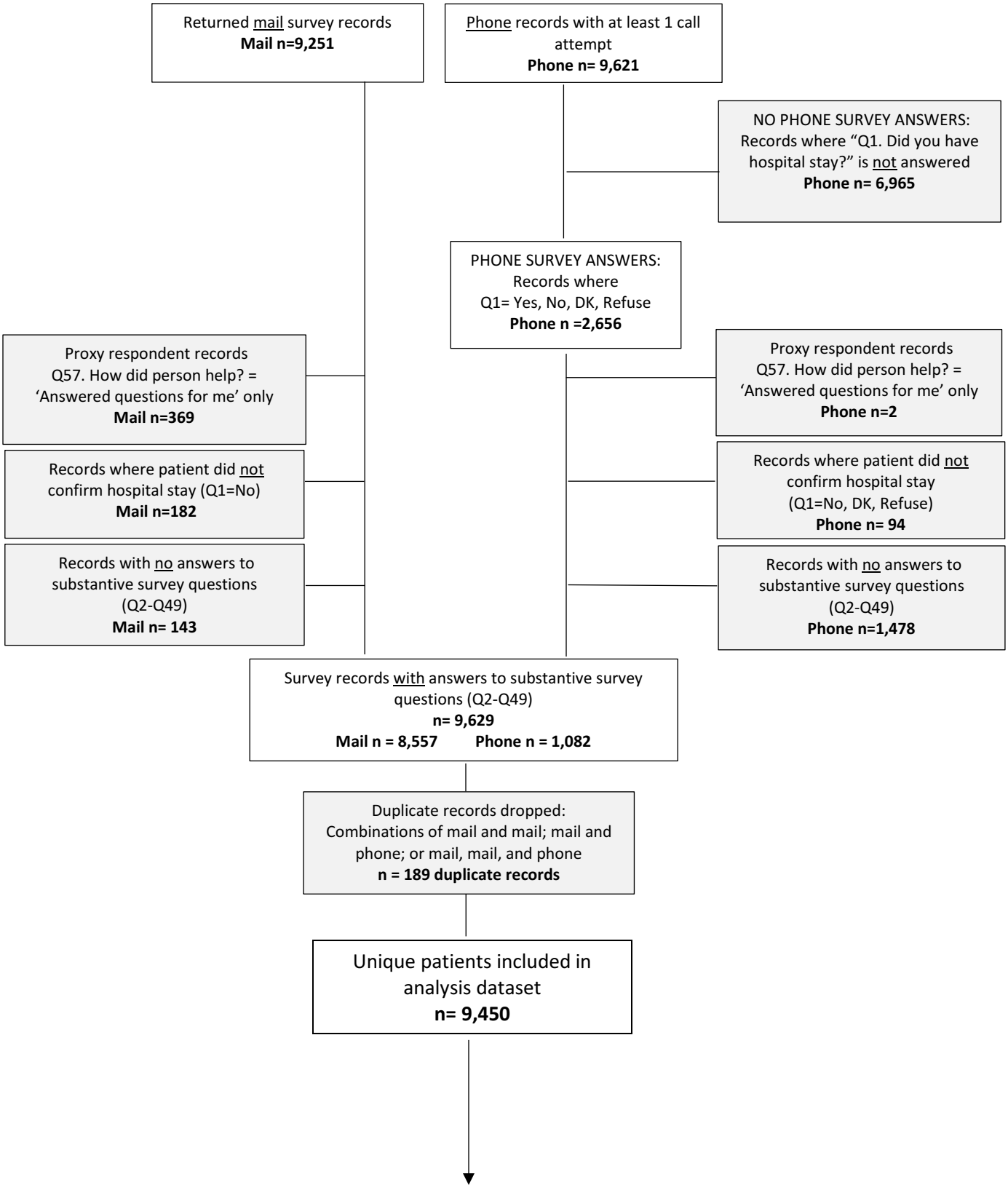
- **Limitations of the Simulation Analysis**
  Our simulated power analyses use retrospective Medicare claims data from 2013-14 and are based upon relatively limited a priori knowledge about the TC components implemented by individual hospitals and self-reported by hospital staff. In particular, the measures of TC components are hospital-level measures and do not account for within-hospital differences in the implementation of TC components for individual patients. The ACHIEVE prospective study will collect much more detailed, patient-level information about the TC components received by individual patients, as reported by patients and caregivers. These enhanced, patient-level measures of TC exposure will further improve the study's statistical power to detect differences in outcomes attributable to TC components. As such, this power analysis provides conservative, lower-bound estimates of minimum detectible differences.

Inferential Analysis. Utilizing the information from fractional factorials and tree-based approaches, hierarchical models will be used to compare TC component exposures while accounting for hospital-level characteristics, patientcharacteristics, and community demographics. Both GEEs and GLMMs (with random intercepts and coefficients) will be used to examine the relationship of TC component exposure using hierarchical characteristics of the community-hospital-patient triad.

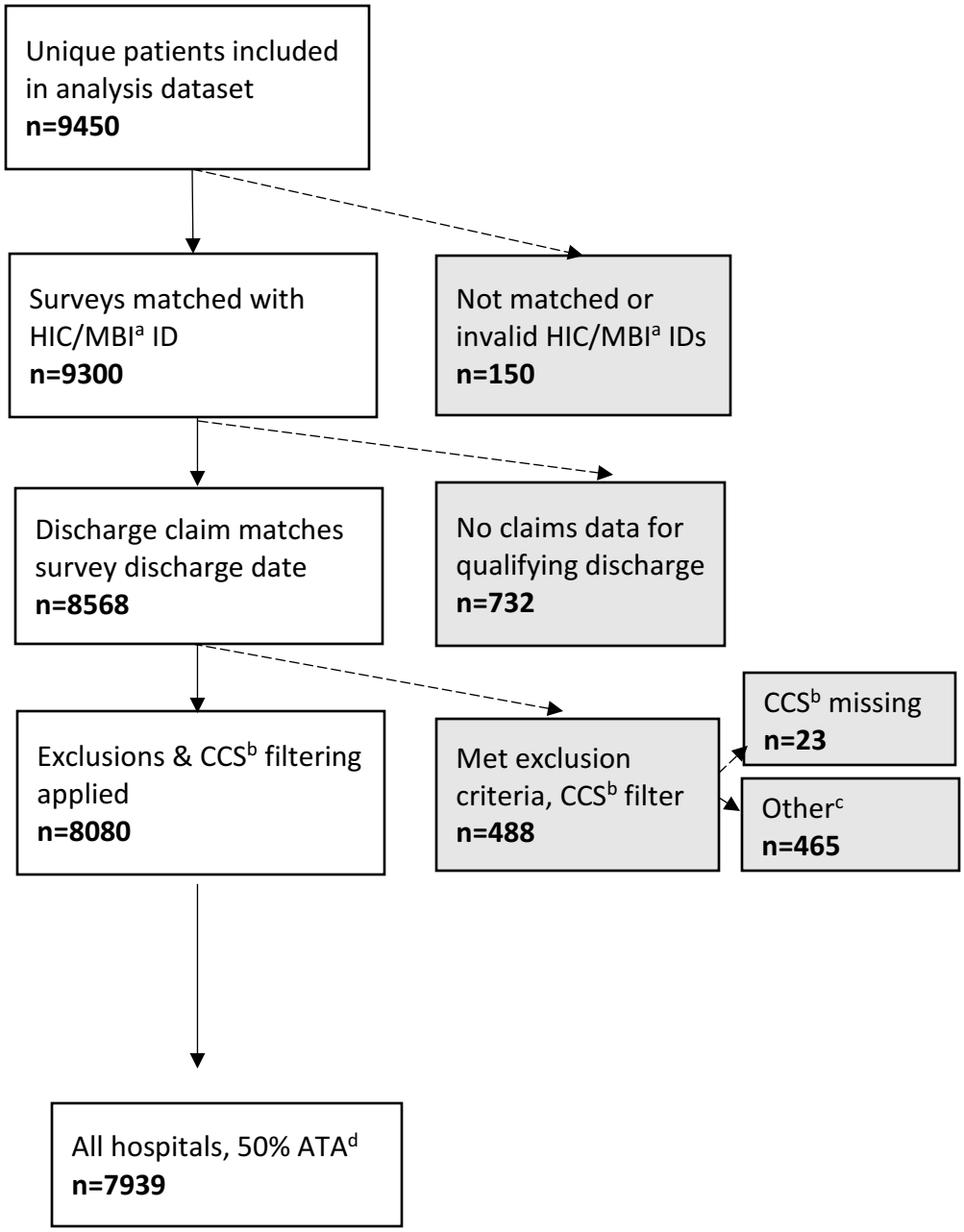
Flow Chart of Patient Records in Project ACHIEVE Analytic File



Flow Chart of Patient Records in Project ACHIEVE Analytic File



Flow Chart of Patient Records in Project ACHIEVE Analytic File



[a] HIC/MBI—Health Insurance Claims/Medicare Beneficiary Identification Number. This was required to be able to link patient responses to their health care utilization data.

[b] Clinical Classification System (CCS). CMS methodology requires mapping certain diagnosis codes to CMS diagnosis coding system. If the patient's diagnosis code was unable to be mapped, the record was excluded.

[c] Other reasons for exclusion include ineligible diagnosis codes (e.g., rehab = 274, cancer = 101, psychiatric = 13, discharged against medical advice, AMA = 19, transfer = 71).

[d] ATA = Applicable to All questions. To ensure higher quality data, we determined that 50% of all questions that were applicable to all participants must be complete for a record to be included in the final dataset.