Statistical Analysis Plan (SAP)

Randomized Evaluation of Default Access to Palliative Services (REDAPS) Trial

Sponsor	NIH/ National Institute on Aging (NIA)
NIH Grant Number	UH2-AG-050311
Principal Investigator	Scott D. Halpern, MD, PhD
Faculty Statistician	Dylan Small, PhD
Senior Statistician	Marzana Chowdhury, PhD
CT.gov number	NCT02505035
Penn IRB number	822134
Original SAP	June 1, 2016
Final SAP	February 8, 2022

Table of Contents

Final Statistical Analysis Plan

I. Design and randomization

The REDAPS study is a stepped-wedge, cluster-randomized, pragmatic clinical trial.

- Comparators: The study compares the status quo paradigm where physicians must identify patients who may benefit from inpatient palliative care (PC) consult services and actively order such services ("usual care") with an opt out paradigm in which patients meeting consensus criteria for eligibility for PC are identified by the electronic health record (EHR) and a PC consultation is ordered by default ("intervention"). Physicians may cancel the default order for PC after being alerted to it and patients or family members may decline such services.

- Stepped wedge intervals: All 11 participating hospitals first contribute a minimum of 4 months of data collection under the usual care paradigm of PC consultation. Then, using a stepped-wedge design, the hospitals are randomly assigned to begin the intervention in intervals spaced approximately 2.4 months apart, representing the average of 8 hospitals having 2.7-month intervals between them and the other 3 EHR–linked hospitals having 1.5-month intervals. By the end of the 32-month trial, all hospitals will have utilized the intervention of default PC consultation for at least 3 months.

- Randomization plan: Eleven hospitals were randomized to transition into the intervention arm sequentially. Among the 11 hospitals, 8 hospitals were randomized individually and the remaining 3 hospitals were treated as a single block because they shared a OneChart EHR platform that necessitated intervention implementation in close proximity (within six weeks). We assigned numbers 1 through 9 randomly to the 8 hospitals and the block and using the random number generator from Random.org. The 3 hospitals within the block were then individually randomized to decide on the sequence that they received the intervention.

See the accompanying Study Protocol for details of the study design, objectives, methods, and oversight. The Study Protocol Amendment Log details changes to the stepped wedge intervals reflected above.

II. Population

We sought to identify patients who (1) have sufficiently complex needs to likely benefit from specialized PC; (2) could be screened easily with a few criteria, thereby augmenting the pragmatic nature of the trial and real-world applicability of the results; and (3) differed from those patients most commonly included in prior or ongoing studies of PC, specifically cancer and heart failure patients, thereby optimally expanding the evidence base. These principles led us to include patients who met the following criteria:

1. Age ≥ 65 years old; **AND**
2. Hospital length of stay (LOS) >= 72 hours; **AND**
3. Any one of the following diagnoses:

Chronic Obstructive Pulmonary Disease (COPD)
- w/ chronic home oxygen use OR ≥ two hospitalizations within 12 months

End-Stage Renal Disease (ESRD)
- w/ chronic dialysis therapy

Dementia
- w/ admission from a long-term care facility (LTC), or surgical feeding tube in place at the time of admission, or ≥ two hospitalizations within 12 months

International Classification of Diseases, 9th and 10th Revisions, Clinical Modification codes can be found in Table S2 of the online supplement. Our accompanying Study Protocol details the rationale for changing the age criterion from ≥ 45 years to ≥ 65 years.

III. Primary outcome and analytic method

a. Primary outcome

The primary outcome is a combined metric of two traditional EHR-based outcomes, hospital mortality and hospital length of stay (LOS). The composite measure of in-hospital mortality and hospital LOS ranks deaths along the LOS distribution,[1] typically at or near the longest end of this distribution.

Defined as [Time 1: time of hospital discharge] – [Time 0: time of enrollment]

For the primary analysis, we will place death at the 99th percentile of the LOS distribution.

b. Primary analytic sample

Our inclusion criteria specify that a patient must be in hospital for at least 72 hours to be included in the study. However, at the time of eligibility determination and enrollment *(which occurs before 72 hours)* we are unable to know if a patient will still be in the hospital at 72 hours. We will undertake a modified intention-to-treat (mITT) approach, from which patients with a LOS <72 hours are excluded (i.e., *the 72-hour rule is a standard exclusion criterion but it is just one that is not known at the time of allocation*). The proposed mITT will more accurately reflect inpatient PC consult effectiveness and cost-effectiveness among the pre-specified eligible cohort.

c. Primary analysis

We will transform the primary outcome to its log value to account for skewness in LOS. Assuming the log of the LOS follows a normal distribution, we will use a mixed-effects model, for analyzing this stepped-wedge, cluster randomized trial as follows:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + X_{ij}\theta + \gamma_k + e_{ijk}$$

where Y_{ijk} denotes the response from individual (with suitable transformation) *k* at time *j* from cluster *i*; α_i is a random effect for cluster *i*; X_{ij} is the treatment indicator for cluster *i* at time *j*, θ is the treatment effect and the residual is *iid* with $e_{ijk} \sim N(0, \sigma_e^2)$. Since the intervention occurs over time, the proportion of clusters exposed to the intervention gradually increases. We will include step as a fixed effect (β_j) in the model to adjust for the potential confounding factor from calendar time. We will add a random effect for individuals (γ_k) to account for the repeated measures because some patients will be present in multiple periods because the level of analyses is the hospital encounter.

Further, though not listed above in the base model, primary analyses will also seek to adjust for imbalance in clusters' and/or patients' characteristics arising from the stepped wedge design. To do so, we will add terms $\gamma_1 \boldsymbol{Z}_{ijk}$ and $\gamma_2 \boldsymbol{W}_{ij}$ to the model, where $\boldsymbol{Z}$ and $\boldsymbol{W}$ represent vectors of patient and cluster characteristics, respectively, that could be related to outcomes. Note that the index *j* in these matrices allows us to include time-varying covariates. One particular $\boldsymbol{Z}_{ijk}$ we will attempt to include is the "time since first enrollment" for patients with subsequent eligible encounters, thereby accounting for possible differences in primary or secondary outcomes as disease progresses, independent of any effect of intervention vs. usual care or of calendar time. We will include these additional terms in primary models to augment precision in estimating the treatment effects.

If LOS or other appropriate transformation of LOS is not normally distributed, we will use a suitable modeling approach to account for the non-normality. Specific models that will be considered include count models. For example, negative binomial models, Quantile regression or others depending on the distribution of the data.

All analyses will be conducted using Stata (StataCorp, College Station, Texas) or R (Vienna, Austria).

Patient-level covariate factors for adjustment	
Patient level covariates	***Variable coding or definition***
• Age	Continuous
• Gender	Binary
• Ethnicity	Binary
• Race	Categorical
• Marital status	Categorical
• Time (in days) since first enrollment	Continuous
• Hospital admission source	Categorical
• Eligible diagnosis (COPD; ESRD; Dementia)	Binary (indicator) for each diagnosis
• Severity of illness (Elixhauser comorbidity index score)	Continuous
• In ICU at time of intervention	Binary

d. Secondary analysis

- Analysis of LOS in the ITT sample with adjusted mixed effects model with random effects for hospitals and fixed effects for time.

e. Sensitivity analysis

- An assessment of the impact of different rankings of death in the LOS distribution. We plan to place death at the 75^{th} through the 95^{th} percentiles in 10 percentile intervals (i.e., 75^{th}, 85^{th}, and 95^{th}) and will assess the change in the estimation.
- We will also conduct the following sensitivity analyses to determine whether extensions to this basic model change our results:

  (1) Unadjusted mixed effects model: To understand the potential contribution of patient-level baseline characteristics, we will conduct an unadjusted analysis of LOS in the mITT sample with random effects for hospital and fixed effects for time.

  (2) Time effect might differ across clusters: To account for the possibility that the time effect might not be the same for all clusters, we will add a random interaction term for the cluster-time combination τ_{ij}.

  (3) Treatment effect might be delayed: To account for the possibility that the effects of the intervention might not be observed until sometime after it is introduced, we will change $X_{ij}\theta$ to $X_{ijl}\theta_l$ to account for the number of steps (or months) since the intervention was introduced in cluster *i.*

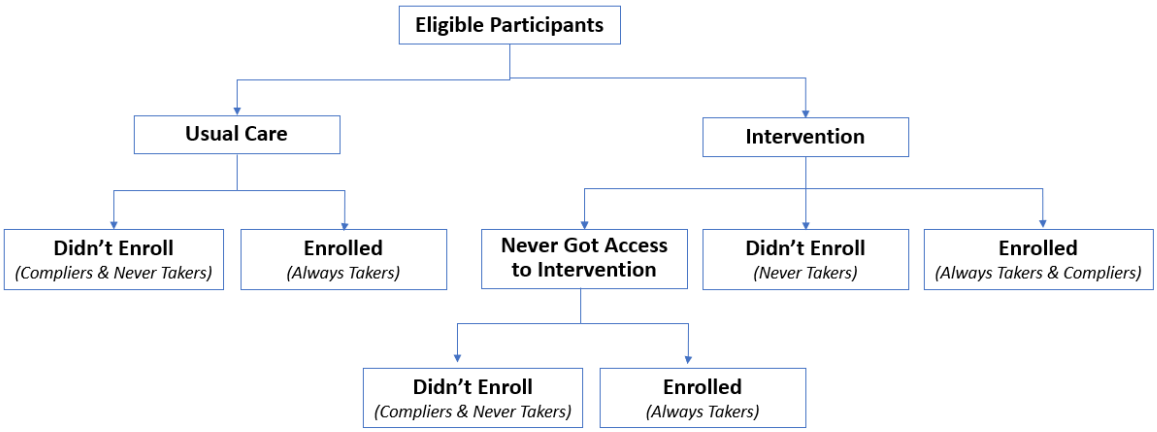
- **Time-to-event model**: To address common alternative approaches for analyzing LOS, we plan to analyze the LOS outcome (without ranking death) as time-to-event data using 2 approaches: (1) Clustered competing-risks model where the event of interest is discharged alive and we will treat 'death' as the competing event, and (2) Mixed-effects Cox proportional model where we will treat death as a censoring event.

f. Complier Average Treatment Effect (CATE) analysis: accounting for non-adherence[2]

As in most pragmatic trials, there are reasons that patients enrolled in this trial's intervention arm may not receive PC. Each of these forms of non-adherence creates differences between the intervention's *effectiveness* (which incorporates non-adherence and is assessed using intention-to-treat analyses) and its *efficacy* (which asks how well the intervention works for those who receive it). We will evaluate efficacy in secondary, explanatory analyses, using the mITT sample, that use modern methods of causal inference to address the fact that analyses restricted to those who receive the intervention do not maintain the virtues of randomization and may be influenced by selection effects. We will use methods developed by the Penn investigative team in which the randomization arm is modeled as an instrumental variable (IV) in analyses of RCTs with non-adherence, promoting unbiased estimates of the efficacy of the intervention.

The CATE approach entails a two-stage least-squares regression in which the randomization arm is modelled as an instrumental variable. This analysis will also be adjusted for cluster and time.

The adherence pathway for this study will be as follows:



Among those assigned to the intervention, there are three different reasons why they may not receive the intervention (from least to most common):

- Group 1: A procedural error in the EHR default alert rule resulted in the default order not firing as intended for some patients (i.e., *they should have received the treatment, but didn't)* – although this situation very rarely occurred, these patients are included in the mITT. These patients will be some combination of *Compliers* and *Never Takers.*

- Group 2: The default order fired (they received the treatment) but the primary team cancelled the default order within 24 hours (before the order was sent to the PC team) – this is a more common scenario and these patients will be considered as *Never Takers.*
- Group 3: The default order fired and was not cancelled (the order was sent to the PC team), but the patient did not receive PC – this is the most common scenario. We will consider these patients as *Never Takers.*

g. Analyses of effect modifiers

We will assess the extent to which the patient characteristics and PC consult characteristics modify the size of the effect of PC on the primary outcome. This will answer the question as to which PC characteristics are associated with the greatest benefits, and which types of patients are particularly likely to benefit from PC. By virtue of the trial's pragmatic design and large sample size, it provides high statistical power to determine how these and other service-level and patient-level factors alter the effectiveness of PC. We will explore effect modification by conducting analyses stratified by the different values of the PC or patient factors. If differences appear, we will formally evaluate for effect modification by testing the significance of the coefficients for statistical interaction terms between the potential effect modifier and the study period (intervention or usual care) on the primary outcome. Importantly, we evaluate PC-level factors (rolling up characteristics of the services they provide across patients) rather than patient-level receipt of specific services because analyses of the latter variables would be subject to confounding by indication (e.g., sicker patients may be seen by a physician).

Effect modifiers of intervention on hospital LOS	
Patient level covariates	***Variable coding or definition***
• Age	Continuous
• Gender (female vs male)	Binary
• Race (Black vs White)	Categorical
• Location at time of intervention (ICU vs ward)	Categorical
• Eligible dementia diagnosis vs not	Binary
• Eligible ESRD diagnosis vs not	Binary
• Eligible COPD diagnosis vs not	Binary
• Marital status (Married vs not)	Binary
• Hospital admission source (SNF vs home)	Categorical

IV. Secondary Outcomes

We will assess numerous clinical and palliative care process measures after enrollment including:

a. Secondary outcome measures

Secondary outcome measures	
Outcome measure	***Variable coding or definition***
• In-hospital death	Binary
• ICU mortality	Binary; yes if death occurs in ICU or within 24 hours of transfer from ICU to hospice
• Discharge to hospice	Binary
• DNR at discharge	Binary
• Receipt of CPR	Binary
• Transfer to ICU	Binary; excludes participants who spent the total duration of the study encounter in an ICU
• Invasive mechanical ventilation	Binary; initiation of MV after enrollment
• Count of 30-day hospital readmission	Count; if a new encounter is >12 hours after discharge from prior encounter
• 30-day hospital readmission rate	Rate; if a new encounter is >12 hours after discharge from prior encounter
• Time to PC consult	Continuous; time from hospital admission to the first signed PC consult note

b. Analysis of secondary outcomes

Secondary outcomes will be analyzed using the mITT and ITT samples. We plan to analyze the data with mixed-effects logistic regression for binary outcomes and mixed-effects count models depending on the dispersion parameter for count outcomes. Count models will be formulated with an offset term to account for differential length of follow-up.

V. Approach to missing data

We do not anticipate substantial missing data because all outcomes will be obtained through the hospitals' EHR systems. If any of the factors of adjustment (listed in the first table) has missing values >10% we will exclude those from the basic model and will look at the effect estimate changes.

VI. Sample size and statistical power calculations

a. Sample size

We queried the EHRs of two Ascension hospitals during calendar year 2013 to identify patients who would meet our eligibility criteria, and to determine how many of those patients had received PC. Around 961 patients met all eligibility criteria, of whom 79 (8.3%) received PC. By considering the proportion of total beds in the 11-hospital sample accounted for by these two hospitals and extrapolating to all 11 hospitals, we estimate that >15,000 patients will be eligible for trial enrollment.

Our final analyses included a total of 34,239 patients (ITT), reflecting that the number of patients in the 11-hospital sample exceeded initial estimates extrapolated from two hospitals.

b. Statistical power calculations

Even using the conservative assumption of an intra-cluster (within-hospital) correlation of patient outcomes of up to ρ=0.20 (design effect = 1+(n-1)*ρ = 2.6), this design provides greater than 80% power to detect a difference in the primary outcome between intervention and control of 0.5 days at the median with α = 0.05. It provides >99% power to detect differences of 1 day at the median. These analyses assume that the composite outcome of hospital length of stay and mortality follows a log normal distribution with a median of 5.5 days (mean of 8.3 days) in the control group (corresponding to a standard deviation of 7.9 days). Unlike prior RCTs of PC, we will be specifically powered to determine how different PC team structures and practices influence the benefits of PC, and the characteristics of patients who derive the greatest benefits. Our estimates using the formula of Hughes and Hussey and confirmed with Monte Carlo simulation suggest that we will have 80% power to detect small effects for both patient-level effect modifiers (which vary among patients within and among hospitals) and cluster-level effect modifiers (which vary among services).

VII. Data Safety & Monitoring Board (DSMB)

Data and Safety Monitoring Board (DSMB) has been convened by the NIA to oversee the conduct of the REDAPS trial and monitor patient safety. The DSMB is comprised of individuals with expertise in palliative care, bioethics, clinical trial conduct, vulnerable populations, decision making, and biostatistics. The DSMB Chair, Dr. Joan Teno, is a world-renowned expert on measuring and evaluating interventions to improve the quality of medical care for seriously ill and dying patients. Members of the DSMB will not be involved in the conduct of the trial. Once convened, the DSMB will perform several duties. First, they will review and approve the research protocol and plans for data and safety monitoring prior to the study. Second, they will evaluate the progress of the trial. This will include assessment of data quality, participant accrual and retention, participant risk versus benefit, and study outcomes. The DSMB will meet regularly and make recommendations to the trial sponsor (NIA) about study progress and safety and will make recommendations about trial continuation. We will charge DSMB members with using their judgment in simultaneously considering many data points in making decisions about trial design modifications and trial continuation or termination.

VIII. Safety data and interim analysis

a. Trial stopping rules / safety monitoring measures

We do not plan to stop the trial early for evidence of the effectiveness of the intervention because doing so would markedly reduce our power to detect which types of PC are most useful and which types of patients derive the most benefit from PC. However, we do propose to stop the trial for early evidence of harm based on the primary outcome (composite measure of in-hospital mortality and hospital LOS).

b. Timing of analysis and adjustment of significance level

The DSMB has recommended implementing the following statistical modifications to preserve the study-wide alpha level. We will calculate a Lan-DeMets spending function using O'Brien-Fleming boundaries to preserve the overall significance level of 0.05 for the interaction terms. The interim tests would use one-sided significance levels of 0.0007 and 0.0161. At the second interim analysis, for example, this is equivalent to calculating a one-sided 98.39% lower confidence bound for the treatment coefficient in

the primary outcome model. If the lower bound of the confidence interval for the change in the composite outcome exceeds two days at either interim analysis, we will propose early termination of the trial to the DSMB. Final analyses will be conducted at a significance level of 0.0451. This design allows for early termination if there is strong evidence of harm, without substantially impacting the significance level for the final analyses.

c. Person performing analysis

Statistical analysis will be performed by the trial's statistician who will be blinded to trial arm. The trial's data manager will be the only person who knows the identity of the arms during the trial's conduct and will provide the statistician with primary outcomes and safety data with dummy variables indicating each arm. The analyst will return the results to the data manager for inclusion in the interim analysis report to the DSMB. All members of the investigative team except the data manager will remain blinded to treatment arm and facility for the analysis.

IX. References

1. Lin, Winston, Scott D Halpern, Meeta Prasad Kerlin, and Dylan S Small. "A 'Placement of Death' Approach for Studies of Treatment Effects on ICU Length of Stay." *Statistical Methods in Medical Research* 26, no. 1 (February 2017): 292–311. doi:10.1177/0962280214545121.
2. Baiocchi, Michael, Jing Cheng, and Dylan S. Small. "Instrumental Variable Methods for Causal Inference." *Statistics in Medicine* 33, no. 13 (2014): 2297-340. doi:10.1002/sim.6128.

SAP Amendments

Concept:	Primary outcome definition
Date:	07.01.2017
Summary:	Hospital LOS is defined as [Time 1: Hospital discharge] – [Time 0: Time of enrollment, at 3:00 PM Hospital Day 1]
Justification:	This amendment was made *a priori*, before analysis and presentation of results in the first interim analyses to the DSMB in July 2017. This change is based on recommendations regarding the concept of "immutable" time laid out by PI Halpern in the 2017 *American Journal of Epidemiology* manuscript. Eligibility for the study is first assessed at 3:00 PM Hospital Day 1, at which time a future-dated PC consult order is generated by the EHR. Only the LOS after 3:00 PM Hospital Day 1 can be plausibly altered by the intervention. Although LOS was previously undefined in the Original SAP, it is traditionally defined as beginning at the time of hospital admission, thus warranting a description of this change herein.
Reference	*Harhay MO, Ratcliffe S, Halpern SD. Measurement Error Due to Patient Flow in Estimates of Intensive Care Unit Length of Stay. Am J Epidemiol; 2017 Dec 15;186(12):1389-1395. doi: 10.1093/aje/kwx222*

Concept:	Primary analysis
Date:	09.20.2018
Summary:	Primary analytic approach will include patient encounters, including rare circumstances in which the default order was not automatically generated for some apparently eligible cases in the intervention phase due to an EHR system glitch.
Justification:	Our primary analytic approach is maximally conservative by including all patient encounter data in the mITT and ITT. This is merely a clarification of our primary analytic approach, not a change; we had not specified our approach to handling this issue initially as we had not foreseen it arising.

Concept:	Secondary outcomes
Date:	05.10.2019
Summary:	Secondary outcome amended to 'New initiation of mechanical ventilation after enrollment'
Justification:	Our Original SAP stated that we wished to analyze Days of Mechanical Ventilation; however, this was not possible due to data limitations across the 8 Ascension ministries.

Concept:	Secondary outcomes
Date:	05.10.2019
Summary:	Secondary outcomes (including palliative care process measures) removed: (1) pain scores, (2) dyspnea, (3) documentation of goals of care, (4) documentation of family meetings, (5) documentation of durable power of attorney, (6) documentation of pain assessment, (7) PC team visits per patients, (8) use of bowel regimen for patients on opioids

Justification:	Data are collated by the central Ascension informatics team from separate EHRs of 8 Ascension ministries. These data are either not routinely collected in discrete fields, have high missing values, or are non-standardized in their collection. After extensive exploration of these data fields, these data were considered too unreliable to be reported.
Concept:	Effect modifiers
Date:	05.10.2019
Summary:	We had planned to measure whether duration of the existence of the PC program or the proportion of consults in a hospital staffed by a physician modified the effect of default orders, but eliminated these plans
Justification:	Data are collated by the central Ascension informatics team from separate EHRs of 8 Ascension ministries. These data are either not routinely collected in discrete fields, are non-standardized in their collection across hospitals, or are not represented at all hospitals (in the case of spiritual care). After extensive exploration of these data fields, these data were considered too unreliable to be reported. Duration of the existence of the PC program and proportion of consults staffed by a physician were not included in the analyses.
Concept:	Primary analysis, sensitivity analyses
Date:	06.18.2019
Summary:	Adjustment variables in models with be excluded if there are missing values >10%. Sensitivity analyses will exclude marital status and insurance status as patient-level covariate adjustment factors
Justification:	We do not anticipate substantial missing data because the outcomes data is obtained through the hospitals' EHR systems; however, we set a standard for what levels of missing data would warrant exclusion (and further analysis of the effect of excluding those variables on the estimates). Additionally, EHR data was re-extracted after completion of the study, pulling in the latest values for patient demographics; however, some patient covariates captured within the EHR can be non-stable over time, namely marital status and insurance.
Concept:	Primary outcome - sensitivity analysis
Date:	07.16.2019
Summary:	For the primary analysis, we will place death at the 99^{th} percentile of the LOS distribution in the linear mixed effects model and conduct sensitivity analyses placing death at the 75^{th} through the 95^{th} percentiles in 10 percentile intervals (i.e. 75^{th}, 85^{th}, and 95^{th}).
Justification:	Our measure of hospital LOS ranks deaths along the LOS distribution, typically at or near the longest end of this distribution and uses sensitivity analyses to assess the impact of different ranking. Although there are no changes to our intended approach, we did not fully specify the percentile intervals for the sensitivity analyses in the Original SAP and they are included in the Final SAP.
Concept:	Effect modifiers

Date:	02.23.2021
Summary:	We removed proposed patient-level effect modifiers: (1) insurance status, and (2) zip code
Justification:	When the age of eligibility was increased from ≥45 to ≥65, as described above, it resulted in nearly 100% of the sample having Medicare as the primary payer thus it was determined that it would be a noninformative analysis. Zip code was highly correlated with cluster and was dropped from the model.
Concept:	Primary analysis
Date:	02.24.2021
Summary:	Primary analytic model will add an adjustment for patients' "time since enrollment."
Justification:	We need to adjust for imbalance in patients' characteristics arising from the stepped wedge design. Including "time since enrollment" is to account for possible differences in primary or secondary outcomes as disease progresses, independent of any effect of intervention vs. usual care or of calendar time. We will include these additional terms in primary models to augment precision in estimating the treatment effects.
Concept:	Primary outcome, sensitivity analyses
Date:	04.30.2021
Summary:	Two additional pre-specified analyses of hospital LOS without ranking death: (1) Clustered competing risk model in which death was considered a competing event, and (2) Mixed-effect Cox proportional model in which death was considered a censoring event.
Justification:	We have added these analyses to check the robustness of our results using different modelling approaches as different models have different assumptions.
Concept:	Primary analysis, sensitivity analyses
Date:	6.03.2021
Summary:	We prespecified a sensitivity analysis that included an intervention-by-cluster interaction term to account for potential heterogeneity in treatment effect across clusters.
Justification:	Model convergence could not be achieved.
Concept:	Secondary outcomes
Date:	07.30.2021
Summary:	Definition of 'hospital discharge disposition' changed from multiple categories (home, home with home care, home hospice, inpatient hospice, nursing facility, long-term acute care facility, other) to a binary variable (hospice discharge vs not).
Justification:	The multi-category approach was not possible to report reliably due to significant variation in categories of discharge disposition reported by the 8 Ascension ministries. The only hypothesized causal pathway for inpatient palliative care to influence discharge disposition that is supported by existing evidence is on hospice use (regardless of location where receiving hospice care), and hospice disposition was reported reliably by all sites.

Concept:	Secondary outcomes
Date:	02.08.2022
Summary:	We will amend the definition of the previously specified secondary outcome 'date of the first palliative care consult' to 'time to completed consult,' defined as the time between hospital admission and the first signed palliative care consult note
Justification:	Our original SAP included the process outcome 'date of the first palliative care consult' however time-to-consult (measured from admission) is a more granular and informative way to capture this measure, and is aligned with prior reports in the literature.

Original Statistical Analysis Plan

X. Design and randomization

The REDAPS study is a stepped-wedge, cluster-randomized, pragmatic clinical trial.

- Comparators: The study compares the status quo paradigm where physicians must identify patients who may benefit from inpatient palliative care (PC) consult services and actively order such services ("usual care") with an opt out paradigm in which patients meeting consensus criteria for eligibility for PC are identified by the electronic health record (EHR) and a PC consultation is ordered by default ("intervention"). Physicians may cancel the default order for PC after being alerted to it and patients or family members may decline such services.

- Stepped wedge intervals: All 11 participating hospitals first contribute a minimum of 3.5 months of data collection under the usual care paradigm of PC consultation. Then, using a stepped-wedge design, the hospitals are randomly assigned to begin the intervention in intervals spaced approximately 2.4 months apart, representing the average of 8 hospitals having 2.7-month intervals between them and the other 3 EHR–linked hospitals having 1.5-month intervals. By the end of the 31-month trial, all hospitals will have utilized the intervention paradigm of PC consultation for at least 3.5 months.

- Randomization plan: Eleven hospitals were randomized into the intervention arm sequentially. Among the 11 hospitals, 8 hospitals were randomized individually and the remaining 3 hospitals were treated as a single block because they shared a OneChart EHR platform that necessitated intervention implementation in close proximity (within six weeks). We assigned numbers 1 through 9 randomly to the 8 hospitals and the block and using the random number generator from Random.org. The 3 hospitals within the block were then randomized to decide on the sequence that they received the intervention.

See the accompanying Study Protocol for details of the study design, objectives, methods, and oversight.

XI. Population

We sought to identify patients who (1) have sufficiently complex needs to likely benefit from specialized PC; (2) could be screened easily with a few criteria, thereby augmenting the pragmatic nature of the trial and real-world applicability of the results; and (3) differed from those patients most commonly included in prior or ongoing studies of PC, specifically cancer and heart failure patients, thereby optimally expanding the evidence base. These principles led us to include patients who met the following criteria:

4. Age ≥ 45 years old; **AND**
5. Hospital length of stay (LOS) >= 72 hours; **AND**
6. Any one of the following diagnoses:

Chronic Obstructive Pulmonary Disease (COPD)

- w/ chronic home oxygen use OR ≥ two hospitalizations within 12 months

End-Stage Renal Disease (ESRD)

- w/ chronic dialysis therapy

Dementia

- w/ admission from a long-term care facility (LTC), or surgical feeding tube in place at the time of admission, or ≥ two hospitalizations within 12 months

International Classification of Diseases, 9th and 10th Revisions, Clinical Modification codes can be found in Table S2 of the online supplement.

XII. Primary outcome and analytic method

a. Primary outcome

The primary outcome is a combined metric of two traditional EHR-based outcomes, hospital mortality and hospital LOS. The composite measure of in-hospital mortality and hospital LOS ranks deaths along the LOS distribution,[1] typically at or near the longest end of this distribution. For the primary analysis, we will place death at the 99th percentile of the LOS distribution.

b. Primary analytic sample

Our inclusion criteria specify that a patient must be in hospital for at least 72 hours to be included in the study. However, at the time of eligibility determination and enrollment *(which occurs before 72 hours)* we are unable to know if a patient will still be in the hospital at 72 hours. We will undertake a modified Intention-to-treat (mITT), from which patients with a LOS <72 hours are excluded (*the 72-hour rule is a standard exclusion criterion but it is just one that is not known at the time of allocation*). The proposed mITT will more accurately reflect inpatient PC consult effectiveness and cost-effectiveness among the pre-specified eligible cohort.

c. Primary analysis

We will transform the primary outcome to its log value to account for skewness in LOS. Assuming the log of the LOS follows a normal distribution, we will use a mixed-effects model, for analyzing this stepped-wedge, cluster randomized trial as follows:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + X_{ij}\theta + \gamma_k + e_{ijk}$$

where Y_{ijk} denotes the response from individual (with suitable transformation) *k* at time *j* from cluster *i;* α_i is a random effect for cluster *i*; X_{ij} is the treatment indicator for cluster *i* at time *j,* θ is the treatment effect and the residual is *iid* with $\mathrm{e_{ijk}} \sim N(0, \sigma_e^2)$. Since the intervention occurs over time, the proportion of clusters exposed to the intervention gradually increases. We will include step as a fixed effect (β_j) in the model to adjust for the potential confounding factor from calendar time. We will add a random effect for individuals (γ_k) to account for the repeated measures because some patients will be present in multiple periods because the level of analyses is the hospital encounter. Further, though not listed above in the base model, primary analyses will also seek to adjust for imbalance in clusters' and/or patients' characteristics arising from the stepped wedge design. To do so, we will add terms $\gamma_1 \boldsymbol{Z}_{ijk}$ and $\gamma_2 \boldsymbol{W}_{ij}$ to the model, where $\boldsymbol{Z}$ and $\boldsymbol{W}$ represent vectors of patient and cluster characteristics, respectively, that could be related to outcomes. Note that the index *j* in these matrices allows us to include time-varying covariates. We will include these additional terms in primary models to augment precision in estimating the treatment effects.

If LOS or other appropriate transformation of LOS is not normally distributed, we will use a suitable modeling approach to account for the non-normality. Specific models that will be considered include count models. For example, negative binomial models, Quantile regression or others depending on the distribution of the data.

All analyses will be conducted using Stata (StataCorp, College Station, Texas) or R (Vienna, Austria).

Patient-level covariates for Adjustment
• Source of hospital admission: home vs. nursing facility • Life-limiting illness: COPD vs. ESRD vs. Dementia • Marital status • Location in hospital: ICU vs. ward • Stage of disease: early vs. late

d. Secondary analysis

- Analysis of LOS in the ITT with adjusted mixed effects model with random effects for hospitals and fixed effects for time.

e. Sensitivity analysis

- **Unadjusted mixed effects model:** Analysis of the LOS in the mITT without patient-level characteristics, with random effects for hospital and fixed effects for time.

- **Composite primary outcome measure**: Sensitivity analyses of the impact of different rankings of death in the LOS distribution.

- **Time effect might differ across clusters:** To account for the possibility that the time effect might not be the same for all clusters, we will add a random interaction term for the cluster-time combination τ_{ij}.

- **Treatment effect might be delayed:** To account for the possibility that the effects of the intervention might not be observed until some time after it is introduced, we will change $X_{ij}\theta$ to $X_{ijl}\theta_l$ to account for the number of steps (or months) since the intervention was introduced in cluster *i.*

- **Treatment effect might differ across clusters:** To account for potential heterogeneity in treatment effect across clusters, we will add a random intervention-by-cluster interaction term.

f. Complier Average Treatment Effect (CATE) analysis: accounting for non-adherence[2]

As in most pragmatic trials, there are reasons that patients enrolled in this trial's intervention arm may not receive PC. Each of these forms of non-adherence creates differences between the intervention's *effectiveness* (which incorporates non-adherence and is assessed using intention-to-treat analyses) and its *efficacy* (which asks how well the intervention works for those who receive it). We will evaluate efficacy in secondary, explanatory analyses, using the mITT sample, that use modern methods of causal inference to address the fact that analyses restricted to those who receive the intervention do not

maintain the virtues of randomization and may be influenced by selection effects. We will use methods developed by the Penn investigative team in which the randomization arm is modeled as an instrumental variable (IV) in analyses of RCTs with non-adherence, promoting unbiased estimates of the efficacy of the intervention.

The CATE approach entails a two-stage least-squares regression in which the randomization arm is modelled as an instrumental variable. This analysis will also be adjusted for cluster and time.

g. Analyses of potential clinician and patient-level effect modifiers

We will assess the extent to which the patient characteristics and PC consult characteristics modify the size of the effect of PC on the primary outcome. This will answer the question as to which PC characteristics are associated with the greatest benefits, and which types of patients are particularly likely to benefit from PC. By virtue of the trial's pragmatic design and large sample size, it provides high statistical power to determine how these and other service-level and patient-level factors alter the effectiveness of PC. We will explore effect modification by conducting analyses stratified by the different values of the PC or patient factors. If differences appear, we will formally evaluate for effect modification by testing the significance of the coefficients for statistical interaction terms between the potential effect modifier and the study period (intervention or usual care) on the primary outcome. Importantly, we evaluate PC-level factors (rolling up characteristics of the services they provide across patients) rather than patient-level receipt of specific services because analyses of the latter variables would be subject to confounding by indication (e.g., sicker patients may be seen by a physician).

Potential modifiers of the effects of PC consultative services

PC consult factors	*Patient factors*
• Proportion of consults in which goals of care are documented	• Source of hospital admission: home vs. nursing facility
• Proportion of consults in which provision of spiritual support is documented	• Life-limiting illness: COPD vs. ESRD vs. Dementia
	• Marital status: married vs not
• Proportion of consults staffed by a physician	• Location in hospital: ICU vs. ward
• Duration of existence of the PC program	• Stage of disease: early vs. late

Abbreviations: COPD, chronic obstructive pulmonary disease; ESRD, end stage renal disease; ICU, intensive care unit

XIII. Secondary outcomes

a. Secondary outcomes

Secondary process and outcome measures

Process measure	*Variable coding or calculation*
Documentation of goals of care	Binary (coded by NLP algorithm)
Documentation of family meetings	Binary (coded by NLP algorithm)
Documentation of durable power of attorney, surrogate, or proxy	Binary (coded by NLP algorithm)

Documentation of pain assessment	Binary
Palliative care team visits per patient	Restricted to patients receiving consultation
Use of bowel regimen for patients on opioids	Binary; coded as present if a contraindication documented
Clinical outcome	***Variable coding or calculation***
Pain scores (excluding dementia patients)	Scores are standardized within hospitals
Dyspnea	Binary (coded by NLP algorithm)
Code status (most recent at time of death or discharge)	Categorical; full, do not resuscitate, do not intubation
Hospital mortality	Binary; yes if death occurred in the hospital (excluding ICU) or patient transferred to inpatient hospice and died within 24 hours
ICU mortality	Binary; yes if death occurred in the ICU or patient transferred from ICU to inpatient hospice and died within 24 hours
Transfer to ICU after randomization	Binary
CPR after randomization	Binary
Days of mechanical ventilation	Ordinal
Hospital discharge disposition	Categorical; home, home with home care, home hospice, inpatient hospice, nursing facility, long-term acute care facility, other
30-day hospital readmissions*	Binary
Economic outcome	***Variable coding or calculation***
Direct cost per hospitalization	Continuous; non-linear
Direct cost per day	Continuous; non-linear

Abbrv: NLP, natural language processing; ICU, intensive care unit; CPR, cardiopulmonary resuscitation
*Limited to readmissions at an Ascension hospital

b. Analysis of secondary outcomes

Secondary outcomes will be analyzed using the mITT and ITT samples. We plan to analyze the data with mixed-effects logistic regression for binary outcomes and mixed-effects count models depending on the dispersion parameter for count outcomes. Count models will be formulated with an offset term to account for differential length of follow-up.

XIV. Approach to missing data

We do not anticipate substantial missing data because outcomes will be obtained through the hospitals' EHR systems.

XV. Sample size and statistical power calculations

a. Sample size

We queried the EHRs of two Ascension hospitals during calendar year 2013 to identify patients who would meet our eligibility criteria, and to determine how many of those patients had received PC. Around 961 patients met all eligibility criteria, of whom 79 (8.3%) received PC. By considering the

proportion of total beds in the 11-hospital sample accounted for by these two hospitals and extrapolating to all 11 hospitals during a 31-month enrollment period, we estimate that >15,000 patients will be eligible for trial enrollment.

b. Statistical power calculations

Even using the conservative assumption of an intra-cluster (within-hospital) correlation of patient outcomes of up to $\rho=0.20$ (design effect = $1+(n-1)*\rho = 2.6$), this design provides greater than 80% power to detect a difference in the primary outcome between intervention and control of 0.5 days at the median with $\alpha = 0.05$. It provides >99% power to detect differences of 1 day at the median. These analyses assume that the composite outcome of hospital length of stay and mortality follows a log normal distribution with a median of 5.5 days (mean of 8.3 days) in the control group (corresponding to a standard deviation of 7.9 days). Unlike prior RCTs of PC, we will be specifically powered to determine how different PC team structures and practices influence the benefits of PC, and the characteristics of patients who derive the greatest benefits. Our estimates using the formula of Hughes and Hussey and confirmed with Monte Carlo simulation suggest that we will have 80% power to detect small effects for both patient-level effect modifiers (which vary among patients within and among hospitals) and cluster-level effect modifiers (which vary among services).

XVI. Data Safety & Monitoring Board (DSMB)

Data and Safety Monitoring Board (DSMB) has been convened by the NIA to oversee the conduct of the REDAPS trial and monitor patient safety. The DSMB is comprised of individuals with expertise in palliative care, bioethics, clinical trial conduct, vulnerable populations, decision making, and biostatistics. The DSMB Chair, Dr. Joan Teno, is a world-renowned expert on measuring and evaluating interventions to improve the quality of medical care for seriously ill and dying patients. Members of the DSMB will not be involved in the conduct of the trial. Once convened, the DSMB will perform several duties. First, they will review and approve the research protocol and plans for data and safety monitoring prior to the study. Second, they will evaluate the progress of the trial. This will include assessment of data quality, participant accrual and retention, participant risk versus benefit, and study outcomes. The DSMB will meet regularly and make recommendations to the trial sponsor (NIA) about study progress and safety and will make recommendations about trial continuation. We will charge DSMB members with using their judgment in simultaneously considering many data points in making decisions about trial design modifications and trial continuation or termination.

XVII. Safety data and interim analysis

a. Trial stopping rules / safety monitoring measures

We do not plan to stop the trial early for evidence of the effectiveness of the intervention because doing so would markedly reduce our power to detect which types of PC are most useful and which types of patients derive the most benefit from PC. However, we do propose to stop the trial for early evidence of harm based on the primary outcome (composite measure of in-hospital mortality and hospital LOS).

b. Timing of analysis and adjustment of significance level

If the DSMB agrees, we will implement the following statistical modifications to preserve the study-wide alpha level. We will calculate a Lan-DeMets spending function using O'Brien-Fleming boundaries to preserve the overall significance level of 0.05 for the interaction terms. The interim tests would use one-sided significance levels of 0.0007 and 0.0161. At the second interim analysis, for example, this is

equivalent to calculating a one-sided 98.39% lower confidence bound for the treatment coefficient in the primary outcome model. If the lower bound of the confidence interval for the change in the composite outcome exceeds two days at either interim analysis, we will propose early termination of the trial to the DSMB. Final analyses will be conducted at a significance level of 0.0451. This design allows for early termination if there is strong evidence of harm, without substantially impacting the significance level for the final analyses.

c. Person performing analysis

Statistical analysis will be performed by the trial's statistician who will be blinded to trial arm. The trial's data manager will be the only person who knows the identity of the arms during the trial's conduct and will provide the statistician with primary outcomes and safety data with dummy variables indicating each arm. The analyst will return the results to the data manager for inclusion in the interim analysis report to the DSMB. All members of the investigative team except the data manager will remain blinded to treatment arm and facility for the analysis.

XVIII. References

1. Lin, Winston, Scott D Halpern, Meeta Prasad Kerlin, and Dylan S Small. "A 'Placement of Death' Approach for Studies of Treatment Effects on ICU Length of Stay." *Statistical Methods in Medical Research* 26, no. 1 (February 2017): 292–311. doi:10.1177/0962280214545121.
2. Baiocchi, Michael, Jing Cheng, and Dylan S. Small. "Instrumental Variable Methods for Causal Inference." *Statistics in Medicine* 33, no. 13 (2014): 2297-340. doi:10.1002/sim.6128.