COMPASS

COMPREHENSIVE POST-ACUTE STROKE SERVICES

Phase 2
Statistical Analysis Plan

May 5, 2020

COMPASS
COMPREHENSIVE POST-ACUTE STROKE SERVICES

Phase 2 Statistical Analysis Plan

STUDY TITLE: Comprehensive Post-Acute Stroke Services (COMPASS) Study

COMMUNICATING PI: Pamela Duncan, PhD

JOINT PIs: Cheryl Bushnell, MD; Wayne Rosamond, PhD

FUNDING AGENCY: Patient-Centered Outcomes Research Institute

PROGRAM AWARD NUMBER: PCS-1403-14532

CLINICAL TRIALS NUMBER: NCT02588664

AUTHOR: Matthew A. Psioda, PhD

ANALYSIS PLAN VERSION: 1.0

DATE: May 5, 2020

SIGNATURE PAGE

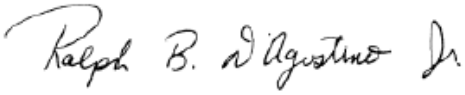
Name	Role	Signature	Date
Pamela Duncan	Principal Investigator		5/05/2020
Cheryl Bushnell	Co-Principal Investigator		5/05/2020
Wayne Rosamond	Co-Principal Investigator		5/05/2020
Matthew Psioda	Lead Statistician; Co-Investigator		5/05/2020
Ralph B. D'Agostino Jr.	Sn. Statistician; Co-Investigator		5/05/2020

Table of Contents

1 LIST OF ABBREVIATIONS

Term	Abbreviation
COMPASS	Comprehensive Post-Acute Stroke Services
NCSCC	North Carolina Stroke Care Collaborative
PSC	Primary stroke center
SIS-16	Stroke Impact Scale - 16
PHQ	Patient Health Questionnaire
MOCA	Montreal Cognitive Assessment
PROMIS	Patient reported outcome measurement information system
REAIM	Reach Effectiveness Adoption Implementation Maintenance
MICE	Multiple imputation by chained equations
PP	Per protocol
CACE	Complier average causal effect
SACE	Survivor average causal effect
BP	Blood pressure
ITT	Intent-to-treat
TIA	Transient ischemic attack
BHM	Bayesian Hierarchical Model

2 INTRODUCTION

The COMPASS Study is a pragmatic, cluster-randomized trial of 41 hospitals in North Carolina and is designed to evaluate the effectiveness of a model of post-acute stroke care (i.e. the COMPASS Care Model intervention) compared with usual care (control). The COMPASS Study includes two phases.[1] Forty-one hospitals were randomized as 40 units (two hospitals were randomized together as one unit) to either receive the COMPASS intervention at the beginning of the trial (Phase 1) or at the start of Phase 2. Hospitals transitioned to Phase 2 after approximately 1 year of enrollment or when key enrollment milestones were met. Hospitals randomized to receive the COMPASS intervention during Phase 2 represent the usual care comparator group during Phase 1. Hospitals randomized to receive the COMPASS intervention in Phase 1 transitioned to sustainability in Phase 2. This statistical analysis plan addresses the analysis of data collected during Phases 1 and 2 of the trial (**Figure 1**). The statistical analysis plan written for the primary & secondary Phase 1 endpoints can be found on clinicaltrials.gov (NCT02588664).

A separate analysis plan will be written for analysis of claims data for Phases 1 and 2.

3 STUDY DESIGN

3.1 Study Population

In 2013, data from hospitals in the North Carolina Stroke Care Collaborative (NCSCC) indicated that 46% of patients were discharged directly home from the hospital after a stroke or TIA (our proposed study population). In that population, the mean age was 65.0 years (SD 14.4), 25% were African American, and 48% were women. Stroke severity, measured by the NIH Stroke Severity score and ranging from 0 (no deficit) to 42 (maximum deficits), was on average 3.2 for those discharged home.

3.2 Randomization

Individual stroke patients cannot easily be randomized to receive the COMPASS intervention. Accordingly, the COMPASS Study utilizes stratified cluster randomization with each of the 41 individual hospitals being randomized to either receive the COMPASS intervention at the beginning of the study (Phase 1) or in Phase 2. Two of the participating 41 hospitals required paired randomization due to having shared staff, resulting in a total of 40 randomized units. The randomization of hospitals was stratified by annual stroke patient volume (2 levels: Large, Medium-to-Small) and whether the hospital is a primary stroke center (2 levels: primary or comprehensive stroke center, neither) resulting in a total of four strata.

3.3 Phase 2 – Usual Care to Intervention Crossover

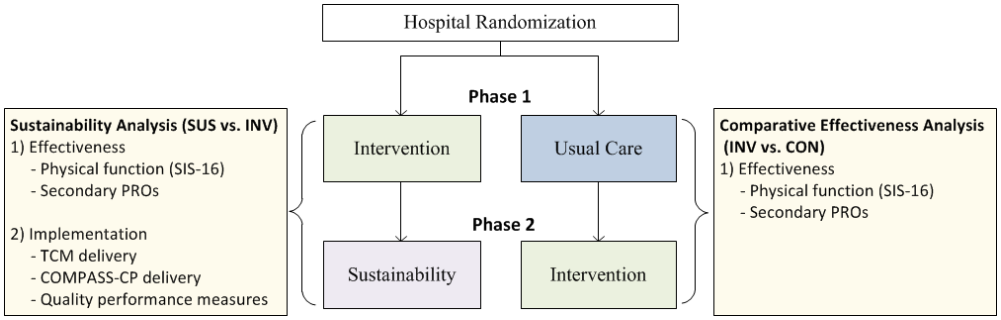
The design of the study allowed hospitals that were randomized to usual care during Phase 1 to transition to provide the intervention during Phase 2. At the end of Phase 1, 16 of the 20 Phase 1 usual care hospitals elected to continue into Phase 2 and transition to provide the intervention. They received comprehensive training on the COMPASS intervention including a 2-day centralized ‘boot camp’ and a hospital-specific site visit. They also received ongoing conference

calls, webinar trainings and case reimbursements. Hospitals crossed over in waves from November 2017 through April 2018, and patient enrollment continued through March 15, 2019.

3.4 Phase 2 – Intervention Sustainability

Hospitals randomized to intervention in Phase 1 received comprehensive support from the study during Phase 1 including one-on-one conference calls, webinar trainings, and case reimbursements. After at least 1 year of participation, intervention hospitals were asked to transition to Phase 2 sustainability. During sustainability, the hospital continued to deliver the COMPASS intervention as their standard of care without study support. Intervention hospitals crossed over from November 2017 through March 2018, and enrolled participants in Phase 2 through March 15, 2019.

Figure 1. COMPASS Study Design and Phase 2 Analysis Overview



4 PHASE 2 OBJECTIVES

4.1 Comparative Effectiveness Analysis

The primary comparative effectiveness objective for Phase 2 of the COMPASS Study is to evaluate whether Stroke Impact Scale 16 (SIS-16) scores are improved during Phase 2 (i.e., the intervention phase) for hospitals randomized to usual care in Phase 1. The primary objective is to simultaneously estimate *within* hospital differences in mean SIS-16 scores as well as the overall mean change (i.e., average change across hospitals).

Secondary comparative effectiveness objectives will evaluate whether several secondary endpoints are improved during Phase 2. These endpoints include home blood pressure monitoring, cognitive function, depression, fatigue, occurrence of falls, disability and dependence measured using the modified Rankin score, and satisfaction with care. For secondary comparative effectiveness analyses, both within hospital changes and overall changes (i.e., average change across hospitals) will be estimated.

All comparative effectiveness endpoints are formally defined in Section 5.1.

4.2 Sustainability Analysis

4.2.1 Effectiveness Sustainability

The effectiveness sustainability objective for Phase 2 of the COMPASS Study is to characterize whether patient outcomes are sustained, worsen, or improve during the sustainability period of the COMPASS study for hospitals randomized to the intervention during Phase 1. For effectiveness sustainability analyses, we will evaluate the primary endpoint for Phase 1 of the COMPASS Study (physical function as measured by the SIS-16) and the secondary outcomes as listed in 4.1.

The two effectiveness sustainability endpoints are shared with the comparative effectiveness analysis and are thus formally defined in Section 5.1.

4.2.2 Implementation Sustainability

The implementation sustainability objectives for Phase 2 of the COMPASS Study are to characterize whether rates of transitional care management (TCM) delivery and eCare plan delivery are sustained, worsen, or improve during the sustainability period of the COMPASS study for hospitals randomized to the intervention during Phase 1. We will also evaluate sustainability of implementation using two post-acute performance measures developed as part of the study (i.e., 2-d follow-up telephone call, 14-d clinic visit).

The implementation sustainability endpoints are formally defined in Section 5.2.2.

5 PHASE 2 ENDPOINTS

5.1 Comparative Effectiveness Endpoints

5.1.1 Primary Endpoint – Stroke Impact Scale 16

The primary endpoint for the COMPASS Study is physical function as measured by the Stroke Impact Scale 16 (SIS-16)[2] at 90 days post-discharge. The SIS-16 instrument is a 16-item survey that assesses the difficulty level of performing basic physical activities (e.g., dressing oneself) over the most recent two-week period. For each item, responses are provided on the following 5-point Likert scale:

Not difficult at all	A little difficult	Somewhat difficult	Very difficult	Could not do at all
5	4	3	2	1

A raw score is obtained for each participant by summing individual item scores. For participants who complete all 16 items, the maximum possible raw score is 80 and the minimum possible raw score is 16. A survey is considered scoreable if a participant answers at least 12 of the 16 items. Standardized scores are computed for each valid raw score using the following formula:

$$\text{analysis score} = \frac{\text{raw score} - n}{5n - n} \times 100,$$

where n is the number of items answered by the participant. Thus, the standardized scores for all participants will have a possible range from 0 to 100 with larger scores corresponding to outcomes that are more favorable.

Ninety-day patient outcomes were assessed through telephone interviews by trained and blinded interviewers using Computer Assisted Telephone Interviewing software. For participants who could not be reached by phone (e.g., invalid phone number, 10 call attempts with no answer), an abbreviated survey was subsequently sent by mail to the participant address on record. This mailed survey included only the SIS-16, self-rated health, and last measured blood pressure. In the rare event that a participant completed a telephone and mailed survey, the SIS-16 from the telephone survey was selected as the participant's SIS-16 outcome unless fewer than 12 items were completed. In such cases, the SIS-16 outcome from the mailed survey was selected. Although the 90-day survey would ideally be completed approximately 90 days after discharge, all scoreable surveys will be included in the primary analysis regardless of the time of completion.

5.1.2 Secondary Endpoints Based on 90-day Surveys

5.1.2.1 Secondary Prevention – Home Blood Pressure Monitoring

Participants are asked whether they monitor their blood pressure at home (yes or no) and, if they answer in the affirmative, how frequently (daily, weekly, and monthly). Home blood pressure monitoring will be analyzed as a dichotomous endpoint (monitoring with any frequency versus no monitoring).

5.1.2.2 Cognition – Montreal Cognitive Assessment (MoCA) 5-Minute Protocol

The MoCA 5-minute protocol is a brief cognitive protocol for screening for vascular cognitive impairment.[3,4] The tool includes 4 items from the full MoCA and examines attention, verbal learning and memory, executive functions/language, and orientation. Each of the four items are scored separately and the scores are summed to obtain a total score that falls between 0 and 30, with higher scores representing better cognition, for analysis as a continuous variable.

5.1.2.3 Depression

The PHQ-2 is a 2-item questionnaire that inquires about the frequency of depressed mood and anhedonia over the past 2 weeks.[5] The first question asks how often the participant had little interest or pleasure in doing things and the second asks how often the participant felt down, depressed, or hopeless. Each of the two questions are answered using a Likert scale with the following scoring rubric:

- 0 = Not at all
- 1 = Several days
- 2 = More than half the days
- 3 = Nearly every day

The total score is the sum of the scores for the two questions and ranges from 0-6. For analysis, following standard screening criteria,[5] the total score will be dichotomized (total score >= 3 versus total score <3).

5.1.2.4 Fatigue

Degree of fatigue will be assessed using the PROMIS Fatigue Instrument.[6] This 4-question self-report instrument asks participants about their level of fatigue over the past 7-day period. Each of the 4 questions are answered using a Likert scale with the following scoring rubric:

1	2	3	4	5
Not at all	A little bit	Somewhat	Quite a bit	Very much

The total raw score is obtained by summing individual question scores and has a range of 4-20. For analysis, raw scores are translated into T-scores using the table show on the right. The T-score rescales the raw score into a standardized score with a mean of 50 and a SD of 10.

Fatigue 4a
Short Form Conversion Table

Raw Score	T-score	SE*
4	33.7	4.9
5	39.7	3.1
6	43.1	2.7
7	46.0	2.6
8	48.6	2.5
9	51.0	2.5
10	53.1	2.4
11	55.1	2.4
12	57.0	2.3
13	58.8	2.3
14	60.7	2.3
15	62.7	2.4
16	64.6	2.4
17	66.7	2.4
18	69.0	2.5
19	71.6	2.7
20	75.8	3.9

*SE = Standard Error

5.1.2.5 Falls

Participants are asked whether they have fallen (yes versus no) since hospital discharge, whether or not the fall resulted in a doctor/emergency room visit, whether they have fallen multiple times since discharge, and how many times they have fallen since discharge. Analysis of falls will be based on incidence of any fall since hospital discharge (no falls versus at least one fall).

5.1.2.6 Disability and Dependence – Modified Rankin Scale

The Modified Rankin Scale measures the degree of disability or dependence in daily activities for people who have suffered a stroke.[7] The scale ranges from 0-6 according to the following criteria:

0 - No symptoms

1 - No significant disability. Able to carry out all usual activities, despite some symptoms.

2 - Slight disability. Able to look after own affairs without assistance, but unable to carry out all previous activities.

3 - Moderate disability. Requires some help, but able to walk unassisted.

4 - Moderately severe disability. Unable to attend to own bodily needs without assistance, and unable to walk unassisted.

5 - Severe disability. Requires constant nursing care and attention, bedridden, incontinent

6 - Dead

Since 90-day survey respondents are alive at the time of survey completion, survey responses will fall between 0 and 5. A value of 6 will be assigned for all participants who are confirmed to have died prior to completion of the 90-days outcomes protocol based on the North Carolina state mortality database. For analysis, values will be categorized as 0, 1, 2-3, and 4-6.

5.1.2.7 Satisfaction with Care

Participant satisfaction with care will be assessed using the Consumer Assessment of Health Plans and Services Clinician and Group Survey version 3.0. This instrument includes 6 questions that ask about how often the patient's doctors explained concepts in a way that was easy to understand, listened carefully, knew important medical history, showed respect for what the patient had to say, spent sufficient time with the patient, and talked about all of the patient's prescription medications. The individual questions are answered using a Likert scale with the following scoring rubric:

- 1 = Never
- 2 = Sometimes
- 3 = Usually
- 4 = Always

A total raw score is obtained by summing the individual question scores and has a range of 4-24 if all questions are answered. Standardized scores will be computed for each valid raw score using the following formula:

$$\text{analysis score} = \frac{\text{actual raw score} - \text{n}}{4\text{n} - \text{n}} \times 100,$$

where n is the number of items answered by the participant. Thus, the standardized scores for all participants will have a possible range from 0 to 100 with greater scores indicating higher satisfaction with care.

5.2 Sustainability Endpoints

5.2.1 Effectiveness Sustainability Endpoints

The effectiveness sustainability endpoints include physical function as measured by the Stroke Impact Scale 16 (SIS-16) and all secondary endpoints. These endpoints are defined in Sections 5.1.1 - 5.1.2.7 of this analysis plan.

5.2.2 Implementation Sustainability Endpoints

5.2.2.1 TCM Delivery

Successful TCM delivery is defined at the patient level as delivering the COMPASS care model intervention according to TCM billing requirements. Successful TCM delivery requires that 1) the patient receive a telephone call within two business days of discharge (or documentation that two call attempts are made); 2) the patient attend a clinic visit within 14 calendar days post-discharge; and 3) the patient receive an eCare Plan at that clinic visit. This definition is equivalent to the Reach definition from the Phase 1 implementation analysis performed in accordance with the REAIM framework.[8]

5.2.2.2 eCare Plan Delivery

Successful eCare Plan delivery is defined as a patient receiving an eCare Plan at the COMPASS clinic visit within 30 calendar days post discharge. These criteria were used to define the per-protocol population in the Phase 1 analysis (available at https://clinicaltrials.gov/ct2/show/NCT02588664).

5.2.2.3 Performance Measures

5.2.2.3.1 2-day Phone Call

Successful implementation of the 2-day phone call is defined as a patient receiving a phone call within 2 business days of hospital discharge or documentation of 2 attempts on the part of hospital staff.

5.2.2.3.2 14-day Clinic Visit

Successful implementation of the 14-day clinic visit is defined as a patient being scheduled for or offered a clinic visit within 14 calendar days regardless of whether they ultimately attended the visit (i.e., if the patients declined to attend when offered or scheduled or was a no-show).

6 ANALYSIS POPULATIONS

For comparative effectiveness and sustainability outcomes, patients from all hospitals that provided the intervention (per-protocol) to at least one patient during Phase 2 of the study will be included.

6.1.1 ITT Analyses

For all ITT analyses, participants will be analyzed as having received the intervention (or usual care) regardless of whether or not that was actually the case. Intent-to-treat analyses will compare outcomes for intervention patients (i.e., Phase 2 patients) to those from usual care patients (i.e., Phase 1 patients).

7 COMPARATIVE EFFECTIVENESS ANALYSIS

7.1 Primary Endpoint

As standardized SIS-16 scores are semi-continuous in nature, the primary analysis will be based on a linear model. It is of primary interest to evaluate the ITT effect for each hospital but also globally (i.e., the average ITT effect across hospitals). We will do so without making the assumption that the ITT effect is common to all hospital units. From Phase 1 of the COMPASS Study, we have observed that compliance with the intervention is highly variable across hospital units and thus the assumption of a common ITT effect across hospitals is not plausible.

Note that here we use the term ITT effect as a synonym for "phase effect" which would be appropriate since the analysis compares patient outcomes from the two consecutive phases of the COMPASS Study. When there is no temporal confounder (e.g., temporal changes in patient characteristics not accounted for in the analysis), the phase effect can be interpreted as an ITT effect.

The linear model for the primary endpoint will incorporate hospital-specific ITT effects with overall model estimation performed using a Bayesian Hierarchical Model (BHM)[9,10] that shrinks estimates of hospital-specific ITT effects towards the overall average effect based on the degree

of homogeneity of the observed effects across the hospital units. The proposed model is given as follows:

$$Y_{phi} = \mu_{0h} + \mu_{1h} \cdot w_{phi} + \boldsymbol{z}_h^T\gamma + \boldsymbol{x}_{phi}^T\beta + \epsilon_{phi}$$
$$\epsilon_{phi} \sim \text{Normal}(0; \sigma^2)$$

where *p* indexes study phase, *h* indexes hospital unit, and *i* indexes patient within hospital unit and study phase.

In the model w_{phi} is an indicator that a patient enrolled in Phase 2 of the study. Thus, the effects of interest are the μ_{1h} parameters, which are the hospital *h* ITT effects (if we assume no temporal or other confounding).

The hospital-specific covariates $\boldsymbol{z}_h$ include randomization stratum (stroke volume; primary stroke center status). Specifically, $\boldsymbol{z}_h$ will include three indicator variables: an indicator that hospital *h* is a high volume PSC, an indicator that hospital *h* is a low volume non-PSC, and an indicator that hospital *h* is a high volume non-PSC. Thus, the reference hospital group corresponds to low volume PSCs. With a pre-post design, it is possible that patient characteristics will be imbalanced between phases and they will be included as needed. The patient-specific covariates $\boldsymbol{x}_{phi}$ may include but are not limited to diagnosis (stroke versus TIA), NIH stroke scale score (categories: 0, 1-4, 5-15, and 16-42), race (white versus non-white), age, whether the patient has documented insurance coverage in their medical record (yes versus no), and whether the patient had a history of stroke or TIA in their medical record (yes versus no). The formal covariate adjustment set will include variables identified to be strongly prognostic for the outcome as well as weakly prognostic if the variable is meaningfully imbalanced in the enrolled patient populations for the two study phases.

7.1.1 Specification of the Bayesian Hierarchical Model

To borrow information across hospitals in the estimation of hospital-specific ITT effects, both the hospital-specific intercepts and ITT effects will be modeled hierarchically. This will allow the analysis to capture heterogeneity in patient outcomes as well as in the effectiveness of the intervention, the latter being of primary interest in this analysis.

The following prior BHM framework will be used:

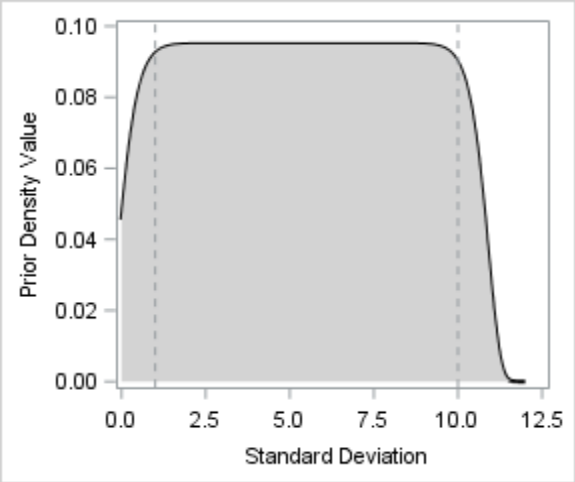
$$(\mu_{0h}|\mu_0; \sigma_0) \sim \text{Normal}(\mu_0; \sigma_0) \quad (\mu_{1h}|\mu_1; \sigma_1) \sim \text{Normal}(\mu_1; \sigma_1)$$
$$\pi(\mu_0) \propto \text{Uniform}(-200,200) \quad \pi(\mu_1) \propto \text{Uniform}(-200,200)$$
$$\pi(\sigma_0) \propto \text{Uniform}(0,50) \quad \pi(\sigma_1) \propto \text{TGN}(5.40,5.50,16.00)$$

In the model formulation, the parameter μ_1 represents the *average* ITT effect across the hospitals. All priors and hyperpriors are elicited to be non-informative other than to rule out implausible values for the parameter in question. The Truncated Generalized Normal (TGN)[11] prior has the form:

$$\pi(\sigma_1) \propto \text{TGN}(5.40,5.50,16.00) \propto \exp\left(-\left|\frac{\sigma_1 - 5.4}{5.5}\right|^{16}\right) \times 1\{\sigma_1 > 0\},$$

and is depicted in Figure 1.

Figure 1: Prior for σ_1



The prior for σ_1 captures the degree to which hospital-specific ITT effects vary about the average effect. The prior $\pi(\sigma_1)$ is approximately uniform on the interval [1, 10]. Note that $\sigma_1 = 10$ would imply that the hospital-specific ITT effects are likely to deviate from the average by as much as 10 SIS-16 percentage points or more. Based on analysis of Phase 1 data, such heterogeneity is not plausible. The prior also suggests that values of $\sigma_1 \approx 0$ are relatively less likely than values in the interval [1, 10]. This is suggestive that there is some degree of heterogeneity in hospital-specific ITT effects, which is consistent with the heterogeneity in compliance levels (a key influence of the ITT effect) observed for Phase 1 intervention hospitals.

7.1.2 Specification of Priors for Other Parameters

For all regression parameters other than μ_{0h} and μ_{1h} (i.e., priors for components of γ and β), independent priors from the same generalized normal family as those used for μ_{0h} and μ_{1h} will be used. Specifically, we will use the prior $\pi(x) \propto \mathrm{GN}(0.0,31.6,16)$ where x represents arbitrary regression parameter. The prior $\pi(x)$ is depicted in Figure 2.

Figure 2: Prior for Components of γ and β

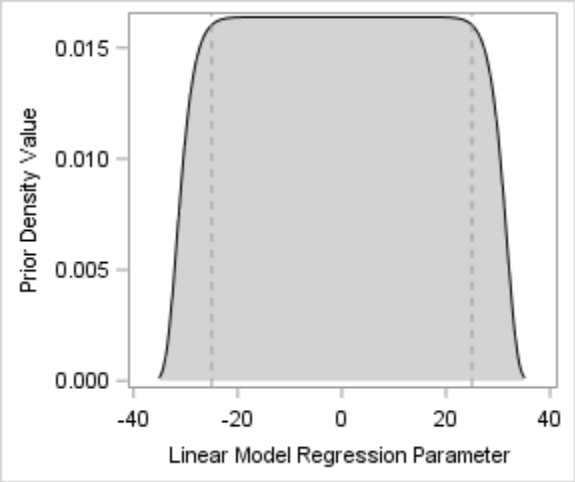
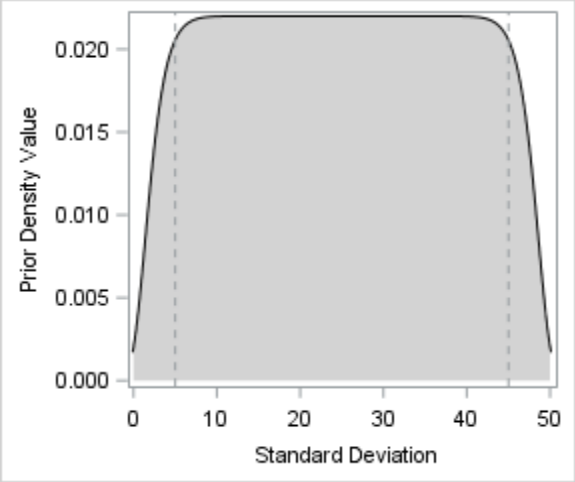


Figure 3: Prior for Standard Deviation Parameter



Note that the prior is essentially uniform over the interval [-25, 25] which encompasses the range of plausible values for all regression parameters based on analyses of Phase 1 data.

A non-informative prior will be used for σ. Specifically, we will consider the prior $\pi(\sigma) \propto \mathrm{TGN}(25,23.1,16)$. This prior has the property that it is approximately uniform over the interval [5, 45] which encompasses the plausible values of σ and effectively rules out impossible values greater than 50. The prior is presented graphically in Figure 3.

7.1.3 Missing Outcome Data

Based on data from Phase 1 of the COMPASS Study, we expect that outcome ascertainment will occur for approximately 55%-65% of patients. Due to the significant number of missing outcomes, a complete case analysis could be biased. Our fully Bayesian analysis will account for missing outcome data by sampling missing outcomes in the Markov Chain Monte Carlo algorithm for model fitting. Missing outcomes will be imputed from a normal full conditional distribution for the missing value given the parameters. Values observed at the floor or ceiling of the instrument will contribute cumulative distribution function (CDF) or survival distribution function (1-CDF) terms, respectively, to the likelihood – essentially using a Tobit regression approach.[12] This approach should adequately address missingness in the outcomes under the assumption that the outcomes are missing at random (MAR) and that the fitted outcome model is a reasonable approximation of the true model.

Patients who die prior to the date of the 90-day survey will be excluded from the analysis. Thus, the primary analysis will effectively be performed *conditional on 90-day survival*. During Phase 1 of the COMPASS study the 90-day mortality incidence was very low and consistent in both the intervention and usual care arms (2.0%, 1.8%) and thus more sophisticated methods such as those that attempt to estimate the survivor average causal ITT effect (i.e., ITT effect among those patients who would survive regardless of receiving usual care or the COMPASS intervention) are not warranted or able to be used.

7.1.4 Missing Covariate Data

Several covariates planned for inclusion in the primary endpoint analysis are not always observed. These covariates include race (white versus non-white), NIH Stroke Scale score (categorical: 0, 1-4, 5-15, and 16-42), and whether the patient has insurance documented in their medical record (yes versus no).

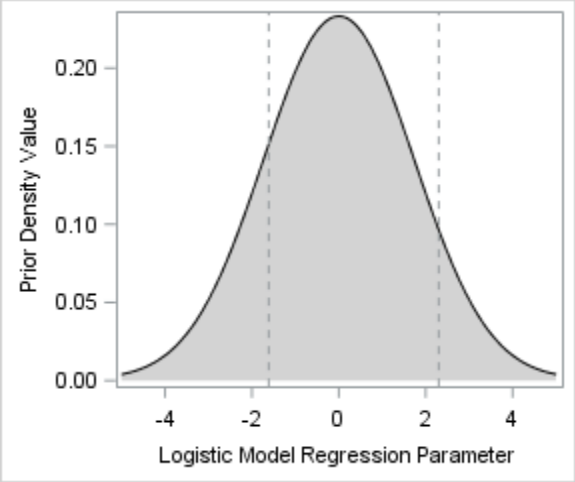
To account for missing values in these covariates, they will also be modeled in our fully Bayesian analysis using standard techniques. Specifically, we will model the sometimes-missing covariates using a marginal/conditional factorization of the covariate joint distribution.[13] This approach induces a joint distribution between the set of sometimes-missing covariates and the outcome from which missing values can be imputed during model fitting. The approach has been shown to perform similarly to multiple imputation by chained equations (MICE),[14] has a sound theoretical justification, and is easily implementable in our fully Bayesian analysis.[15]

To supplement the outcome model given in Section 7.1, we will fit a joint regression model for race, insurance status, and NIH Stroke Scale score category using a factorization of a posited joint distribution. Let RACE_{phi}, INS_{phi}, and NIH_{phi} represent the race, insurance status, and NIH Stroke Scale score, respectively, for phase p participant i from hospital h. We consider the following factorization model:

$$\begin{aligned}
\mathrm{logit}\left(P\left(\mathrm{NIH}_{phi} \leq j\right)\right) &= \alpha_j + \boldsymbol{x}_{phi}^T \cdot \boldsymbol{\beta}_1 + \mathrm{RACE}_{phi} \cdot \beta_{1\mathrm{R}} + \mathrm{INS}_{phi} \cdot \beta_{1\mathrm{I}}, \\
\mathrm{logit}\left(P\left(\mathrm{INS}_{phi} = 1\right)\right) &= \boldsymbol{x}_{phi}^T \cdot \boldsymbol{\beta}_2 + \mathrm{RACE}_{phi} \cdot \beta_{2\mathrm{R}}, \\
\mathrm{logit}\left(P\left(\mathrm{RACE}_{phi} = 1\right)\right) &= \boldsymbol{x}_{phi}^T \cdot \boldsymbol{\beta}_3,
\end{aligned}$$

where $\boldsymbol{x}_{phi}$ is a set of patient-specific variables predictive of the variable being modeled. For each regression parameter, we will use the prior $\pi(x) \propto \mathrm{GN}(0.0, 2.43, 2.0)$ where x represents an arbitrary regression parameter. The prior $\pi(x)$ is depicted in Figure 4. An absolute value of approximately 1.61 (left dashed reference line) for a logistic model regression coefficient corresponds to an odds ratio of approximately 5, illustrating that the prior suggests odds ratios at the extremes of [-5,5] are only modestly less likely than null odds ratios. An absolute value of 2.30 (right dashed reference line) illustrates that odds ratios approximately equal to 10 are also well supported by the prior and are approximately half as likely as null values.

Figure 4: Logistic Regression Parameter Prior



For NIH Stroke Scale score, the three α_j parameters will be used to model intercepts of the cumulative logistic model (aka proportional odds model). The following non-informative prior will be used

$$\pi(\alpha) \propto 1[\alpha_1 < \alpha_2 < \alpha_3] \sum_{j=1}^{3} 1\big[\big|\alpha_j\big| < 10\big].$$

Essentially the prior specifies a uniform prior distribution for each α_j over the interval [-10,10] aside from imposing the ordering constraint on the intercepts required of the cumulative logistic model.

The approach outlined above should adequately address missingness in the covariates under the assumption that they are MAR and that the model for each missing covariate is a reasonable approximation of the true model.

7.2 Analysis of Secondary Comparative Effectiveness Endpoints

All secondary comparative effectiveness endpoints listed in Section 5.1 will be analyzed using the same strategies as described in Section 7 for the primary endpoint.

7.2.1 Continuous Secondary Endpoints

For continuous outcomes (e.g., cognition, fatigue, and satisfaction with care), linear models with hierarchical priors will be employed. The priors used for analysis will be elicited using the same strategy as described for the primary outcome. In all cases, priors will be elicited to have minimal influence on the analysis apart from ruling out implausible values.

7.2.2 Binary Secondary Endpoints

For binary outcomes (e.g., home blood pressure monitoring, depression, and falls), logistic models with hierarchical priors will be employed. The priors used for analysis will match with those described above for the regression coefficients and treatment effect parameters.

7.2.3 Ordinal Secondary Endpoint

For the disability and dependence endpoint based on the Modified Rankin Scale, a cumulative logistic regression model with hierarchical priors will be employed. The priors used for analysis will match with those described in Section 7.1.4. Due to the multiple intercepts in the ordinal logistic model, these parameters will not be modeled hierarchically for computational reasons. However, treatment effects will be modeled hierarchically as described above.

7.3 Sensitivity Analyses for the Bayesian Hierarchical Model

The analyses described above in Section 7 make use of the Bayesian hierarchical model to borrow information across hospitals in the estimation of hospital-specific intervention effects. For hospitals that enroll relatively few patients, the information borrowing procedure may strongly shrink the hospital-specific intervention effect estimate towards the average observed across hospitals. As a sensitivity analysis to the planned analysis using the hierarchical model, we will perform a secondary analysis that limits the degree of information borrowing by modifying the hierarchical variance prior (e.g., $\pi(\sigma_1)$ in Section 7.1.1) so that minimal information is borrowed to estimate hospital-specific effects.

8 SUSTAINABILITY ANALYSIS

8.1 Effectiveness Sustainability

For effectiveness sustainability analyses, the only difference between the comparative effectiveness analysis and the effectiveness sustainability analysis is the set of hospitals included in the analysis. The comparative effectiveness analysis is restricted to Phase 1 usual care hospitals and the effectiveness sustainability analysis is restricted to Phase 1 intervention hospitals. In both cases, we also require that the hospital successfully delivered the intervention per protocol to at least 1 participant in Phase 2. The statistical methodology is the same. Accordingly, all ITT analysis methods described in Section 7 are applicable for the sustainability effectiveness analyses and will be implemented as described in Section 7.

8.2 Implementation Sustainability

The implementation sustainability endpoints are binary. The statistical methodology used to analyze the data for the implementation sustainability endpoints will match that described in Section 7 for comparative effectiveness analysis of binary endpoints with one exception. For these analyses, we will stratify the analysis according to stroke diagnosis type (stroke, TIA) to determine if sustainability differed by diagnosis.

8.3 Sensitivity Analyses for the Bayesian Hierarchical Model

Sensitivity analyses will be performed as described in Section 7.3.

BIBLIOGRAPHY

1. Duncan PW, Bushnell CD, Rosamond WD, et al. The Comprehensive Post-Acute Stroke Services (COMPASS) study: design and methods for a cluster-randomized pragmatic trial. *BMC neurology.* 2017;17(1):133.
2. Duncan PW, Lai SM, Bode RK, Perera S, DeRosa J. Stroke Impact Scale-16: A brief assessment of physical function. *Neurology.* 2003;60(2):291-296.
3. Wong A, Nyenhuis D, Black SE, et al. Montreal Cognitive Assessment 5-minute protocol is a brief, valid, reliable, and feasible cognitive screen for telephone administration. *Stroke.* 2015;46(4):1059-1064.
4. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society.* 2005;53(4):695-699.
5. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Medical care.* 2003;41(11):1284-1292.
6. Lai JS, Cella D, Choi S, et al. How item banks and their application can influence measurement practice in rehabilitation medicine: a PROMIS fatigue item bank example. *Archives of physical medicine and rehabilitation.* 2011;92(10 Suppl):S20-27.
7. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke.* 1988;19(5):604-607.
8. Gesell SB, Bushnell CD, Jones SB, et al. Implementation of a billable transitional care model for stroke patients: the COMPASS study. *BMC health services research.* 2019;19(1):978.
9. Hoff PD. *A First Course in Bayesian Statistical Methods.* Springer Publishing Company, Incorporated; 2009.
10. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis, Third Edition.* Taylor & Francis; 2013.
11. Nadarajah S. A generalized normal distribution. *Journal of Applied Statistics.* 2005;32(7):685-694.
12. Takeshi A. *Tobit Models.* Vol Oxford: Basil Blackwell1985.
13. Ibrahim JG, Chen M-H, Lipsitz SR. Bayesian Methods for Generalized Linear Models with Covariates Missing at Random. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique.* 2002;30(1):55-78.
14. Ibrahim JG, Chen M-H, Lipsitz SR, Herring AH. Missing-Data Methods for Generalized Linear Models: A Comparative Review. *Journal of the American Statistical Association.* 2005;100(469):332-346.
15. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research.* 2007;16:23.