
Analytical Plan for Time-adjusted effect of socioeconomic status in mortality rates after brain injury: cohort study

DOCUMENT: SAP-2023-016-BH-v02

From: Felipe Figueiredo To: Brennan Hickson

2023-07-19

TABLE OF CONTENTS

1	ABBREVIATIONS.....	2
2	CONTEXT.....	2
2.1	Objectives.....	2
2.2	Hypotheses.....	2
3	DATA.....	3
3.1	Raw data.....	3
3.2	Analytical dataset.....	3
4	STUDY PARAMETERS.....	4
4.1	Study design.....	4
4.2	Inclusion and exclusion criteria.....	4
4.3	Exposures.....	4
4.4	Outcomes.....	4
4.5	Covariates.....	5
5	STATISTICAL METHODS.....	5
5.1	Statistical analyses.....	5
5.1.1	Descriptive analyses.....	5
5.1.2	Inferential analyses.....	5
5.1.3	Statistical modeling.....	6
5.1.4	Missing data.....	6
5.2	Significance and Confidence Intervals.....	6
5.3	Study size and Power.....	6
5.4	Statistical packages.....	6
6	OBSERVATIONS AND LIMITATIONS.....	7
7	REFERENCES.....	7
8	APPENDIX.....	8
8.1	Availability.....	8
8.2	Associated analyses.....	8

Analytical Plan for Time-adjusted effect of socioeconomic status in mortality rates after brain injury: cohort study

Document version

Version	Alterations
01	Initial version
02	Imputation + time-varying SES

1 ABBREVIATIONS

- FIM: Functional Independence Measure
- CI: confidence interval
- DCI: Distress community index
- HR: hazards ratio
- LOCF: Last observation carried forward
- SD: standard deviation
- SES: socioeconomic status
- TBI: Traumatic brain injury

2 CONTEXT

2.1 Objectives

To determine the effect of socioeconomic status of the neighborhood on mortality of patients with brain injury, accounting for time-dependent covariates.

2.2 Hypotheses

The neighborhood to which an individual is discharged from acute care affects the mortality rates of individuals that suffered a brain injury.

3 DATA

3.1 Raw data

The raw data was received in three distinct databases, one containing clinical and epidemiological data, a second one with follow up on the status of individuals and a third one containing SES information of each Zip code. The clinical and followup tables were merged by study ID, and this was joined with the SES table by the Zip code of the neighborhood of discharge, resulting in the original raw data base.

The original data base had 711 variables collected on 76665 observations from 19303 individuals.

Following the findings of **SAR-2023-017-BH-v01** the multiple observations per individual should be favored where the time-varying SES exposure allows the full model specification in this analytical plan to be used. The raw data for this analysis was prepared and procedures were described in that document, as well as its accompanying analysis plan (**SAP-2023-017-BH-v02**).

3.2 Analytical dataset

The date of death of participants was available, and its presence was used as an indicator of the living status of the participant at each follow up. All variables in the raw dataset had varying missing data codes in the data dictionary made available by the researcher, which were used to attribute missingness status to each datum. Most of the categorical variables were measured with many levels that were condensed into fewer levels for analysis. The quartiles for each of the both FIM scores were calculated for each individual and saved as two separate variables as covariates for the analysis.

After the cleaning process 23 variables were included in the analysis. The total number of observations excluded due to incompleteness and exclusion criteria will be reported in the analysis.

All variables in the analytical set were labeled according to the raw data provided and values were labeled according to the data dictionary for the preparation of production-quality results tables and figures.

4 STUDY PARAMETERS

4.1 Study design

This is a retrospective analysis of a prospective cohort study.

4.2 Inclusion and exclusion criteria

Inclusion criteria

1. Participants with at most 10 years of follow up;
2. Participants included in the cohort between 2010-01-01 and 2018-12-31;
3. Only the last valid observation of each individual will be included in the analysis
4. Participants with non-negative time under observation*.
 - this criterion will be treated specially in this analysis (see section 5.1.4).

Exclusion criteria

1. Observations after 2019-12-31 will be excluded in order to mitigate risk of confounding by COVID-19 related deaths.
2. Observations prior to this date will still be considered for participants where such data is available.

4.3 Exposures

SES of the neighborhood to which the participant was discharged. The SES measure was stratified into its quintiles, and labelled according to the data dictionary to facilitate interpretation of the results.

4.4 Outcomes

Specification of outcome measures (Zarin, 2011):

1. (Domain) Mortality
2. (Specific measurement) Death
3. (Specific metric) Time-to-event
4. (Method of aggregation) Hazard ratio

Primary outcome

Death after a brain injury.

4.5 Covariates

Covariates will be included in the models in sets (section 5.1.3).

Demographic covariates:

- Sex
- Race
- Age at injury
- Education
- Employment status

Clinical covariates:

- Cause of injury
- Substance Problem Use
- Previous seizure disorder diagnosis
- Spinal cord injury
- Primary rehabilitation payer
- Days From Injury to Rehab Discharge
- FIM Motor at Discharge quartiles
- FIM Cognitive at Discharge quartiles

Geographical covariates:

- Rural area
- Residence after rehab discharge

5 STATISTICAL METHODS

5.1 Statistical analyses

5.1.1 Descriptive analyses

The epidemiological profile of the study participants will be described. Demographic and clinical variables will be described as mean (SD) or as counts and proportions (%), as appropriate. The distributions of participants' characteristics will be summarized in tables and visualized in exploratory plots.

5.1.2 Inferential analyses

All inferential analyses will be performed in the statistical models (described in the next section).

5.1.3 Statistical modeling

The hazard of mortality will be assessed with multivariate Cox regression models. In order to assess if there is an effect of the SES of the neighborhood to which the participant was discharged on mortality will be assessed with four nested models.

1. A crude estimate of the HR between each SES quintiles and mortality will be calculated as the basis of interpretation of the effect.
2. Model 2 will include demographic covariates on top of the crude estimate.
3. Model 3 will add clinical covariates on top of model 2.
4. Model 4 will add geographical covariates to model 3, i.e., it will be calculated adjusting for all covariates described in section 4.5.

The estimates on the association between SES quintiles and mortality will be compared across models in a sensitivity analysis to gauge the cumulative effect of each set of covariates.

The proportional hazards assumption will be tested by assessing the the distribution of Schoenfeld residuals against time, the Schoenfeld test and the distribution of Martingale residuals of time-dependent covariates against the null model. In order to keep the maximum number of time-dependent covariates in the final model, we plan to perform the residual analysis combining the following approaches: stratification by categorical variables, adding explicit interactions with time, time splitting and non-linear fitting of covariates using polynomials or splines. Time-dependent covariates that fail these approaches will be removed before the final model is evaluated and described.

5.1.4 Missing data

A simple imputation using an LOCF approach will be used on zip codes prior to merging the SES. All data cleaning procedures, as well as the details of the missing data imputation were performed and are described in **SAP-2023-017-BH-v02**. Missing data counts and proportions will be reported in tables.

5.2 Significance and Confidence Intervals

All analyses will be performed using the significance level of 5%. All significance hypothesis tests and confidence intervals computed will be two-tailed.

5.3 Study size and Power

N/A

5.4 Statistical packages

This analysis will be performed using statistical software R version 4.3.0.

6 OBSERVATIONS AND LIMITATIONS

Recommended reporting guideline

The adoption of the EQUATOR network (<http://www.equator-network.org/>) reporting guidelines have seen increasing adoption by scientific journals. All observational studies are recommended to be reported following the STROBE guideline (von Elm et al, 2014).

7 REFERENCES

- **SAR-2023-016-BH-v01** – Time-adjusted effect of socioeconomic status in mortality rates after brain injury: cohort study
- **SAR-2023-017-BH-v01** – Sensitivity of mortality rates to the imputation of missing socioeconomic data: cohort study
- **SAP-2023-017-BH-v02** – Analytical Plan for Sensitivity of mortality rates to the imputation of missing socioeconomic data: cohort study
- Zarin DA, et al. The ClinicalTrials.gov results database – update and key issues. N Engl J Med 2011;364:852-60 (<https://doi.org/10.1056/NEJMsa1012065>).
- Gamble C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337–2343 (<https://doi.org/10.1001/jama.2017.18556>).
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg. 2014 Dec;12(12):1495-9 (<https://doi.org/10.1016/j.ijsu.2014.07.013>).

8 APPENDIX

This document was elaborated following recommendations on the structure for Statistical Analysis Plans (Gamble, 2017) for better transparency and clarity.

8.1 Availability

All documents from this consultation were included in the consultant's Portfolio.

The portfolio is available at:

<https://philsf-biostat.github.io/SAR-2023-004-BH/>

8.2 Associated analyses

This analysis is part of a larger project and is supported by other analyses, linked below.

Effect of socioeconomic status in mortality rates after brain injury: cohort study

<https://philsf-biostat.github.io/SAR-2023-004-BH/>

Sensitivity of mortality rates to the imputation of missing socioeconomic data: cohort study

<https://philsf-biostat.github.io/SAR-2023-017-BH/>