

Association of Stroke and Traumatic Brain Injury with Healthcare Access and Utilization in the U.S. Adult Population: A Propensity Score–Matched Analysis of NHANES 2011–2014

Mike Baiocchi, Jonathan Pipping, Andrea Schneider, Ashil Srivastava and Dylan Small

Contents

1	Title and Abstract	3
1.1	Title	3
1.2	Abstract	3
2	Introduction	3
3	Research Question	4
3.1	Primary Research Question	4
3.2	Hypotheses	4
3.3	Objectives	5
3.4	Generalization Target	5
4	Inclusion and Exclusion Criteria	5
4.1	Data Source	5
4.2	Inclusion Criteria	6
4.3	Exclusion Criteria	6
4.4	Data Counts and Flow	6
5	Study Outcomes	7
5.1	Exposures	7
5.2	Primary Outcomes	7
5.3	Secondary Outcomes	7
5.4	Marginal Distributions and Descriptive Summaries	8
6	Study Design	8
6.1	Overview and Primary Design Features	8
6.2	Covariates: Observed Confounders and Effect Modifiers	8
6.3	Potential Unobserved Confounders	8
6.4	Handling Missing and Corrupted Data	9
6.5	Propensity Score Estimation	9
6.6	Matching Algorithm	9
6.7	Covariate Balance and Diagnostics	10
6.8	Outcome Analysis and Bias Control	10
6.9	Implementation	10
7	Study Analysis Plan	11
7.1	Primary Analysis	11
7.2	Connection to Hypotheses	11
7.3	Multiple Testing	11

7.4	Inferential Methods	11
7.5	Challenges and Solutions	11
7.6	Empirical Summary	11
7.7	Result Interpretation	13
7.8	Generalizability	13
8	Sensitivity Analysis	13
8.1	Formal Sensitivity Methods	13
8.2	Interpretation of Formal Sensitivity Analyses	13
8.3	Stability Analyses	13
9	Simulation of Protocol	13
9.1	Simulation of Data	13
9.2	Data Pipelining Code	13
9.3	Baseline Summary Code	13
9.4	Analyses Code	13
9.5	Results Summary Code	13
10	Amendments	13
11	Funding	13
12	Future Steps	13

1 Title and Abstract

1.1 Title

Association of Stroke and Traumatic Brain Injury with Healthcare Access and Utilization in the U.S. Adult Population: A Propensity Score–Matched Analysis of NHANES 2011–2014

1.2 Abstract

Background: Traumatic brain injury (TBI) and stroke are common neurologic conditions associated with substantial morbidity and mortality in the United States. Both conditions can have long-term cognitive, functional, and psychosocial consequences that may increase the need for ongoing healthcare. At the same time, barriers to healthcare access and utilization may worsen outcomes among individuals with a history of TBI or stroke.

Objectives: This study will estimate the association between self-reported lifetime history of (1) stroke and (2) traumatic brain injury (TBI) with loss of consciousness and multiple indicators of healthcare access and utilization in a nationally representative sample of U.S. adults.

Data and Setting: We will use data from the 2011–2012 and 2013–2014 cycles of the National Health and Nutrition Examination Survey (NHANES), a complex, multistage probability sample of the noninstitutionalized U.S. civilian population.

Design: This is an observational, cross-sectional study using existing survey data. For each exposure (stroke and TBI), we will construct separate analytic cohorts and estimate propensity scores for exposure as a function of demographic, socioeconomic, health behavior, cardiovascular risk, and survey design variables (strata, primary sampling unit, and interview weights), along with missingness indicators. We will perform optimal 1:k propensity score matching with a caliper of 0.2 pooled standard deviations on the logit of the propensity score and assess covariate balance using standardized mean differences and graphical diagnostics (love plots).

Primary Outcomes: The primary outcomes are (1) having a usual place to go for healthcare and (2) having any health insurance coverage. Secondary outcomes include type of insurance (public vs. private), type of usual source of care, time since last healthcare visit, mental health service use, hospitalizations, and prescription coverage.

Analytic Plan: Within the matched samples, we will estimate the population average treatment effect on the treated (PATT), i.e., the effect among adults with a history of stroke or TBI, using survey-weighted models that incorporate both NHANES sampling weights and the matched design, following the framework of DuGoff et al. (2014) for combining propensity score methods with complex survey data.

Results and Conclusions: This document describes the protocol and planned analyses. Numerical results and final conclusions will be added once analyses are completed.

2 Introduction

Traumatic brain injury (TBI) and stroke are major causes of death and disability in the United States. TBI, typically resulting from a bump, blow, or jolt to the head—for example from falls or motor vehicle crashes—ranges in severity from mild concussions to severe injuries causing permanent neurologic impairment. Stroke, caused by interruption of blood flow to the brain (ischemic) or rupture of a cerebral blood vessel (hemorrhagic), similarly leads to significant cognitive and functional impairment and often requires intensive acute and chronic care.

In the United States, hundreds of thousands of adults each year experience stroke, and many die from TBI-related injuries, with many more living with long-term aftereffects. Adequate access to healthcare, including continuous insurance coverage, a usual source of care, and timely primary and specialty services, is likely critical to optimizing outcomes among people with a history of stroke or TBI. Conversely, gaps in insurance

coverage, lack of a usual source of care, and delayed or forgone care may exacerbate disability and increase the risk of preventable complications.

Despite the plausibility of a strong link between severe neurologic injury and subsequent healthcare access and utilization, evidence on this association at the population level using nationally representative data with comprehensive covariate adjustment remains limited. While some studies have examined healthcare utilization following stroke or TBI, existing work often relies on single health systems, selected subpopulations (e.g., hospitalized patients), or administrative data that may not simultaneously capture neurologic history, detailed healthcare access measures, and rich demographic and socioeconomic covariates needed for robust causal inference.

The National Health and Nutrition Examination Survey (NHANES) provides an opportunity to address this gap. NHANES is a nationally representative, multistage probability survey of the noninstitutionalized U.S. civilian population that collects standardized information on demographics, health behaviors, chronic conditions, health insurance, and healthcare utilization. In the 2011–2012 and 2013–2014 survey cycles, NHANES asked respondents about lifetime history of stroke and, in adults aged 40 years and older, history of head injury resulting in loss of consciousness.

Using these data, this study will:

1. Quantify how healthcare access and utilization differ between adults with and without a history of stroke and between adults with and without a history of TBI.
2. Evaluate whether these associations vary across demographic and socioeconomic subgroups (e.g., age, sex, race/ethnicity, education, income).
3. Implement propensity score methods combined with NHANES survey weights, following DuGoff et al. (2014), to obtain effect estimates that are both adjusted for observed confounding and generalizable to the national target population.

By doing so, the study aims to inform whether individuals with prior stroke or TBI experience systematic differences in access to and use of healthcare services relative to the broader U.S. adult population.

3 Research Question

3.1 Primary Research Question

Among noninstitutionalized U.S. adults represented by NHANES 2011–2014, what is the association between self-reported lifetime history of:

1. Stroke, and
2. Traumatic brain injury (TBI) with loss of consciousness

and multiple indicators of healthcare access and utilization, including having a usual source of care and health insurance coverage?

Formally, for each exposure (stroke, TBI) we seek to estimate the *population average treatment effect on the treated* (PATT)—the difference in healthcare access and utilization outcomes between:

- adults who report a history of stroke (or TBI), and
- the same population, had they not experienced stroke (or TBI),

generalized to the NHANES target population via survey weighting.

3.2 Hypotheses

Primary hypotheses

- Adults with a history of TBI are less likely to have adequate healthcare access and utilization (e.g., less likely to have a usual place for care, more likely to lack coverage or have longer time since last visit).
- Adults with a history of stroke are more likely to have healthcare access and utilize healthcare services (e.g., more likely to have a usual source of care and insurance) compared with adults without stroke, reflecting increased contact with the healthcare system after a major vascular event.

Secondary hypothesis

- The magnitude and direction of these associations differ across subgroups defined by age, sex, race/ethnicity, education, and income-to-poverty ratio (i.e., effect modification by demographic and socioeconomic characteristics).

3.3 Objectives

1. Describe healthcare access and utilization indicators among U.S. adults overall and stratified by history of stroke and TBI.
2. Estimate the association (PATT) between stroke history and each healthcare access/utilization outcome, adjusting for observed confounders using propensity score matching combined with survey design-based analysis.
3. Estimate the analogous association for TBI.
4. Explore effect modification by key demographic and socioeconomic variables.

3.4 Generalization Target

The target population for inference is the U.S. noninstitutionalized civilian population represented by NHANES 2011–2014. Findings are intended to apply to:

- Adults aged ≥ 20 years for the stroke analysis (the age range for which stroke history was collected), and
- Adults aged ≥ 40 years for the TBI analysis (the age range for which the TBI module was administered).

Findings will not generalize to:

- Institutionalized populations (e.g., nursing homes, long-term care facilities),
- Active-duty military personnel, or
- People experiencing homelessness or others outside the NHANES sampling frame.

4 Inclusion and Exclusion Criteria

4.1 Data Source

We will construct analytic cohorts from the combined 2011–2012 and 2013–2014 NHANES cycles. These cycles include:

- Demographics (DEMO),
- Health insurance and access (HIQ, HUQ),
- Medical conditions including stroke (MCQ),
- Traumatic brain injury (CSQ; administered to adults ≥ 40 years),
- Alcohol use (ALQ), smoking (SMQ), hypertension (BPQ), and diabetes (DIQ).

These components will be merged by respondent identifier (SEQN) and survey cycle.

4.2 Inclusion Criteria

We will define separate analytic samples for the stroke and TBI analyses.

Stroke analysis

- Adults aged ≥ 20 years at the time of the NHANES examination.
- Nonmissing response to the stroke history question (MCQ160F: “Has a doctor or other health professional ever told you that you had a stroke?” coded as yes/no).
- Nonmissing data on the primary outcomes (usual source of care and health insurance coverage).
- Availability of NHANES design variables (strata, primary sampling unit) and interview sampling weights (WTINT2YR).

TBI analysis

- Adults aged ≥ 40 years (NHANES eligibility for the TBI module).
- Nonmissing response to the TBI question (CSQ240: “Have you ever had loss of consciousness because of a head injury?” yes/no).
- Nonmissing data on the primary outcomes.
- Availability of design variables and interview sampling weights.

4.3 Exclusion Criteria

We will exclude:

- Respondents with missing or indeterminate exposure status (stroke history or TBI),
- Respondents missing all primary outcomes,
- Respondents without valid sampling weights or design variables required for survey analysis,
- Participants outside the age ranges for each analytic cohort (e.g., < 20 for stroke analysis, < 40 for TBI analysis).

We will not exclude participants solely based on missing covariate values used for propensity score estimation; instead, we will address covariate missingness as described in Section 6 (missingness indicators and single imputation).

4.4 Data Counts and Flow

In the final manuscript, we will report:

- The total number of NHANES participants in 2011–2014,
- Counts excluded at each step (e.g., ineligible age, missing exposure, missing outcomes, missing design variables),
- Number of exposed and unexposed participants entering the propensity score models, and
- Number retained in the matched samples for each exposure.

These will be summarized in a CONSORT-style flow diagram, as recommended in observational protocol guidance.

5 Study Outcomes

5.1 Exposures

We consider two separate binary exposures, each analyzed in its own matched cohort.

History of stroke

- Derived from NHANES item MCQ160F (“Ever told you had a stroke?”).
- Coded as: 1 = Yes (stroke_exposed), 0 = No.

History of traumatic brain injury (TBI)

- Derived from NHANES item CSQ240 (“Ever had a loss of consciousness because of a head injury?”).
- Coded as: 1 = Yes (tbi_exposed), 0 = No.

5.2 Primary Outcomes

We will define two primary healthcare access outcomes (binary).

Usual place to go for healthcare Derived from NHANES item HUQ030 (“Is there a place you usually go when you are sick or need advice about your health?”). Coded as: 1 = Has a usual place (response = 1 or 3), 0 = No usual place (response = 2).

Any health insurance coverage Derived from NHANES item HIQ011 (“Are you covered by health insurance or some other kind of health care plan?”). Coded as: 1 = Any coverage (public or private), 0 = Uninsured at the time of interview (response = 2).

These primary outcomes directly reflect core dimensions of healthcare access: having a regular source of care and being insured.

5.3 Secondary Outcomes

Secondary outcomes will characterize additional aspects of access and utilization:

- Type of insurance plan (e.g., government vs. private coverage among those insured),
- Type of usual source of care (e.g., doctor’s office/clinic, hospital outpatient clinic, emergency department, other),
- Time since last healthcare visit (categorical measure based on HUQ visit timing, e.g., ≤ 6 months, 6–12 months, > 12 months),
- Mental health service use (visit to a mental health professional in the past year; yes/no),
- Hospital utilization (number of overnight hospital stays in the past year, modeled as count or categorized),
- Insurance continuity and prescription coverage (any gaps in health insurance in the past year; whether the plan covers prescription medications).

All secondary outcomes will be coded in a way that aligns with NHANES public documentation and allows for interpretable contrasts between exposed and unexposed groups.

5.4 Marginal Distributions and Descriptive Summaries

Before matching, we will:

- Present weighted marginal distributions (e.g., proportions, means) of all primary and secondary outcomes in the overall sample and stratified by exposure status (stroke/TBI yes/no),
- Produce tables and figures (e.g., bar charts, histograms) to describe the distribution of outcomes across survey cycles (2011–2012 vs. 2013–2014), consistent with the shell tables.

These summaries will provide context for the matched analyses and help assess whether the matched sample remains representative of the target population.

6 Study Design

6.1 Overview and Primary Design Features

This is an observational, cross-sectional study using existing NHANES data to estimate the association between lifetime neurologic injury (stroke, TBI) and healthcare access/utilization. The principal design elements are:

1. Use of a nationally representative complex survey (NHANES) to enable population-level inference.
2. Propensity score matching to reduce confounding by observed covariates.
3. Combination of propensity score methods with survey weighting and design-based analysis, following DuGoff et al. (2014), to obtain effect estimates that are both confounding-adjusted and generalizable to the NHANES target population (PATT).

We will conduct two parallel analyses: one with stroke as the exposure, one with TBI as the exposure.

6.2 Covariates: Observed Confounders and Effect Modifiers

Based on prior literature and data availability, we identify the following as key observed confounders and potential effect modifiers:

- **Demographic characteristics:** age (RIDAGEYR), sex (RIAGENDR), race/ethnicity (RIDRETH3).
- **Socioeconomic status:** education level (DMDEDUC2), income-to-poverty ratio (INDFMPIR).
- **Health behaviors and risk factors:** alcohol use and potential alcohol abuse (derived from ALQ151, ALQ141Q), smoking status (derived from SMQ020 and SMQ040; categories: every day, some days, not at all), hypertension (BPQ020), diabetes (DIQ010, including borderline as “yes”).
- **History of stroke:** for the TBI analysis, we will include stroke.history (MCQ160F) as a covariate, given its strong relation to both TBI and healthcare utilization.
- **Survey design characteristics:** primary sampling unit (SDMVPSU), stratum (SDMVSTRA), and NHANES interview sampling weight (WTINT2YR) will be included as predictors in the propensity score model, consistent with DuGoff et al. (2014)’s recommendations for Stage 1.

We will treat age, sex, race/ethnicity, education, and income as candidate effect modifiers and will explore interactions in secondary analyses.

6.3 Potential Unobserved Confounders

We acknowledge that some important determinants of both exposure and outcomes may be unmeasured or imperfectly measured in NHANES, including:

- Severity, timing, and mechanism of stroke or TBI,
- Cognitive impairment severity, functional status, and disability,

- Social support and detailed health system characteristics.

These factors may induce residual confounding. We plan to address robustness to unmeasured confounding via sensitivity analyses (see planned Section 8), for example gamma-based sensitivity analyses for matched observational studies.

6.4 Handling Missing and Corrupted Data

We will use a structured approach to missing data:

1. **Missingness indicators.** For covariates with missing data (e.g., RIDAGEYR, RIAGENDR, RIDRETH3, DMDEDUC2, INDFMPIR, alcohol_abuse, smoking_status, hypertension, diabetes, stroke_history), we will create binary indicators (e.g., RIDAGEYR_missing) capturing whether the value was originally missing.
2. **Single imputation for covariates.** Continuous variables (e.g., age, income-to-poverty ratio) will be imputed using the sample mean. Categorical variables (e.g., sex, race/ethnicity, education, smoking_status) will be recoded with an explicit “Missing” category. Binary health conditions (alcohol_abuse, hypertension, diabetes, stroke_history) will be imputed to their modal value.
3. **Exposures and primary outcomes.** Exposure variables (stroke_exposed, tbi_exposed) and primary outcomes will not be imputed; records with missing exposure or primary outcome will be excluded from the analytic sample as described in Section 4.

Missingness indicators will then be included as covariates in the propensity score models, allowing the propensity score to account for systematic patterns of missingness in covariates.

6.5 Propensity Score Estimation

For each exposure (stroke and TBI), we will estimate propensity scores using logistic regression. The propensity score models will include all measured covariates (demographic characteristics, socioeconomic status, health behaviors and risk factors, and missingness indicators), along with the NHANES survey weights (WTINT2YR) and design variables (SDMVPSU, SDMVSTRA). For the TBI analysis, we will also include stroke history as a covariate. This approach is consistent with DuGoff et al. (2014)’s Stage-1 recommendation for incorporating survey design information in propensity scores. We will predict propensity scores and their logits for all individuals with nonmissing exposure status.

6.6 Matching Algorithm

We will implement optimal propensity score matching using the `MatchIt` package in R. For each exposure, we will:

1. Restrict to participants with nonmissing exposure status (stroke data, TBI data).
2. Fit a logistic regression model to estimate propensity scores for all participants with nonmissing exposure status.
3. Compute the pooled standard deviation of the logit propensity score among treated and control participants.
4. Define a caliper as 0.2 times the pooled standard deviation of the logit propensity score.
5. Construct a distance matrix based on absolute differences in logit propensity score between treated and control units.
6. Apply the caliper by adding a large penalty (1000) to distances exceeding the caliper.
7. Perform optimal 1:k matching ($k = 1, \dots, 6$) without replacement using `MatchIt`’s `method = "optimal"` and the pre-computed distance matrix.

For each matching ratio (1:1 through 1:6), we will:

- Extract the matched dataset,
- Record the number of matched treated and control individuals and any treated units discarded due to the caliper, and
- Save matched datasets for later analysis.

This design primarily targets the average treatment effect on the treated (ATT) in the matched sample; in combination with survey weights in the outcome analysis, we will interpret estimates as approximating the population average treatment effect on the treated (PATT) in the NHANES target population.

6.7 Covariate Balance and Diagnostics

We will assess covariate balance using the `cobalt` package in R. For each matched dataset, we will:

- Compute standardized mean differences (SMD) before and after matching using `bal.tab`,
- Generate love plots summarizing covariate balance pre- and post-matching,
- Summarize residual imbalance by counting covariates with $|\text{SMD}| > 0.05$, 0.10, and 0.20, and report the maximum $|\text{SMD}|$ and the percentage of covariates with $|\text{SMD}| > 0.10$.

We will select a final matching ratio (e.g., 1:1, 1:2, etc.) for each exposure based on:

- Achieving $\text{SMD} < 0.10$ for all or nearly all covariates,
- Minimizing the number of treated units discarded, and
- Preserving adequate sample size for stable estimation.

Diagnostic plots and summary statistics will be included in the final manuscript to demonstrate successful balance in the matched samples.

6.8 Outcome Analysis and Bias Control

Following DuGoff et al. (2014), the outcome analysis will:

- Use survey-weighted regression models (e.g., logistic regression for binary outcomes, Poisson/negative binomial for counts) applied to the matched data,
- Incorporate NHANES sampling weights and design variables (strata and PSU) to ensure proper variance estimation and population-level inference, and
- Reflect the matching design (by using the matched sample and appropriate weights).

For each exposure and outcome, the primary estimand will be the PATT—the difference in mean outcomes between exposed and unexposed, among those exposed in the target population.

This design addresses bias from observed covariates by:

- Including a rich set of confounders and missingness indicators in the propensity score model,
- Ensuring good covariate balance after matching, and
- Using survey design-based inference to retain generalizability.

Unobserved confounding will be addressed in planned sensitivity analyses (Section 8).

6.9 Implementation

All data management and analyses will be implemented in R, using the following core packages: `haven` and `tidyverse` for data ingestion and merging; `readr` and `tableone` for descriptive summaries; `MatchIt` for propensity score matching; and `cobalt` for covariate balance diagnostics.

7 Study Analysis Plan

7.1 Primary Analysis

7.2 Connection to Hypotheses

7.3 Multiple Testing

7.4 Inferential Methods

7.5 Challenges and Solutions

7.6 Empirical Summary

The following tables present shell structures for the key summary statistics, tables, and visualizations that will be used to report findings from the analyses.

Table 1: Demographic and clinical characteristics of adults with and without traumatic brain injury (NHANES 2011–2014 TBI subsample)

Characteristic	Total	TBI	No TBI
Age (years)			
Sex			
Male			
Female			
Race / Ethnicity			
Income-to-poverty ratio			
Health insurance coverage			
Education level			
Alcohol consumption			
Hypertension			
Diabetes			
Smoking			
Ever told you had a stroke?			
If yes: age at stroke (years)			
If yes: time since stroke (years)			

Table 2: Demographic and clinical characteristics of adults with and without stroke (NHANES 2011–2014 stroke subsample)

Characteristic	Total	Stroke	No stroke
Age (years)			
Sex			
Male			
Female			
Race / Ethnicity			
Income-to-poverty ratio			
Health insurance coverage			
Education level			
Alcohol consumption			
Hypertension			
Diabetes			
Smoking			

Table 3: Covariate balance before propensity score matching for TBI and stroke

Variable (% or mean)	TBI vs No TBI			Stroke vs No stroke		
	No TBI	TBI	SMD	No stroke	Stroke	SMD
Age (years)						
Age range (e.g. 40–49, 50–59, ...)						
Sex: Male						
Sex: Female						
Race / Ethnicity						
Income-to-poverty ratio						
Health insurance coverage						
Education level						
Alcohol consumption						
Hypertension						
Diabetes						
Smoking						

Table 4: Covariate balance after propensity score matching for TBI and stroke

Variable (% or mean)	TBI vs No TBI			Stroke vs No stroke		
	No TBI	TBI	SMD	No stroke	Stroke	SMD
Age (years)						
Age range (e.g. 40–49, 50–59, ...)						
Sex: Male						
Sex: Female						
Race / Ethnicity						
Income-to-poverty ratio						
Health insurance coverage						
Education level						
Alcohol consumption						
Hypertension						
Diabetes						
Smoking						

Table 5: Healthcare access and utilization indicators by TBI status

Measure of access / utilization	Total	TBI	No TBI
Has routine place to go for healthcare			
Type of usual place (e.g. clinic, doctor's office, ER, other)			
Any health insurance coverage			
Type of insurance: government			
Type of insurance: private			
Type of place most often used for healthcare			
Time since last healthcare visit			
Seen a mental health professional in past year			
Number of overnight hospital stays (past year)			
Any time without health insurance in past year			
Plan covers prescription medicines			
Plan covers mental-health related services			

Table 6: Healthcare access and utilization indicators by stroke status

Measure of access / utilization	Total	Stroke	No stroke
Has routine place to go for healthcare			
Type of usual place (e.g. clinic, doctor’s office, ER, other)			
Any health insurance coverage			
Type of insurance: government			
Type of insurance: private			
Type of place most often used for healthcare			
Time since last healthcare visit			
Seen a mental health professional in past year			
Number of overnight hospital stays (past year)			
Any time without health insurance in past year			
Plan covers prescription medicines			
Plan covers mental-health related services			

7.7 Result Interpretation

7.8 Generalizability

8 Sensitivity Analysis

8.1 Formal Sensitivity Methods

8.2 Interpretation of Formal Sensitivity Analyses

8.3 Stability Analyses

9 Simulation of Protocol

9.1 Simulation of Data

9.2 Data Pipelining Code

9.3 Baseline Summary Code

9.4 Analyses Code

9.5 Results Summary Code

10 Amendments

11 Funding

12 Future Steps

References

DuGoff, E. H., Schuler, M., and Stuart, E. A. (2014). Generalizing observational study results: applying propensity score methods to complex surveys. *Health Services Research*, 49(1):284–303.