Socioeconomic Determinants of Long COVID Outcomes: A Secondary Analysis of NHIS COVID-19 Supplement (2023–2024)

**Abstract**  
*Background:* Long COVID has emerged as a persistent public health concern, disproportionately affecting vulnerable populations. This study explores the role of socioeconomic status (SES)—specifically income, education, and access to healthcare—in influencing the prevalence and severity of Long COVID symptoms among U.S. adults aged 18 to 65.  
*Methods:* We used self-reported data from the CDC's 2023–2024 National Health Interview Survey (NHIS) COVID-19 supplement, focusing on 450 respondents with valid Long COVID (LONGCOVD1\_A) and symptom severity (LCVDACT\_A) responses. SES was operationalized using a composite index incorporating education level (EDUCP\_A), income (INCWRKO\_A), and access barriers due to cost (HISTOPCOST\_A). Analytical methods included chi-squared tests, Mantel-Haenszel stratification, ordinal and binary logistic regression, and SES-stratified visualizations.  
*Results:* While plots suggested a gradient between low SES and Long COVID burden, statistical models revealed no significant associations between composite SES and symptom severity. However, female sex emerged as a strong predictor of Long COVID prevalence. Age showed weak associations in multivariable models.  
*Conclusion:* Although trends suggest SES disparities in Long COVID outcomes, our analysis did not yield statistically significant effects for composite SES. Results highlight the complexity of disentangling socioeconomic influences and warrant further longitudinal research.

**Introduction**

Long COVID, also known as post-acute sequelae of SARS-CoV-2 infection (PASC), refers to the persistence or recurrence of COVID-19 symptoms weeks to months after the initial infection. Affecting an estimated 10–30% of previously infected individuals, Long COVID has become a pressing public health concern due to its impact on physical, cognitive, and psychosocial functioning. Symptoms range from fatigue, cognitive impairment ("brain fog"), and respiratory difficulties to more debilitating effects that limit daily activities and workforce participation.

Emerging evidence suggests that Long COVID does not affect all demographic groups equally. Socioeconomic status (SES)—a multifaceted construct encompassing income, education, and healthcare access—may play a significant role in both the **prevalence** and **severity** of Long COVID symptoms. Individuals of lower SES often face barriers to early diagnosis, limited healthcare continuity, and higher occupational exposure, compounding their risk for both infection and adverse outcomes.

Despite these concerns, quantitative data on how SES influences Long COVID outcomes remain limited. Most studies have focused either on biological risk factors or healthcare utilization, with fewer addressing social determinants of health. Furthermore, analyses that examine SES as a multidimensional construct, rather than isolated indicators, are rare.

To address this gap, our study investigates whether SES indicators—specifically education level, work-related income, and cost-related healthcare access barriers—are associated with Long COVID prevalence and symptom severity in U.S. adults aged 18–65. We hypothesize that lower SES will correlate with both higher likelihood of experiencing Long COVID and increased symptom burden. By leveraging nationally representative data from the CDC’s National Health Interview Survey (NHIS) COVID-19 supplements (2023–2024), we aim to uncover potential disparities that could inform more equitable policy and intervention strategies.

Dataset description:

**📊 OVERVIEW: What’s in Your Dataset**

**✅ 1. Outcome Variables**

| **Variable** | **Description** | **Type** |
| --- | --- | --- |
| LONGCOVD1\_A | Long COVID status (1 = Yes, 2 = No) | Binary (prevalence) |
| LCVDACT\_A | Long COVID symptom severity (1 = Not at all, 3 = A lot) | Ordinal (severity) |

**✅ 2. SES-Related Variables**

| **Variable** | **Description** | **Type** |
| --- | --- | --- |
| EDUCP\_A | Education level (1 = less than high school → 10 = graduate degree) | Ordinal |
| INCWRKO\_A | Work-related income (grouped) | Ordinal |
| HISTOPCOST\_A | Stopped insurance due to cost? (1 = Yes, 2 = No) | Binary (access) |
| SES\_Index | Composite score from the above (0–3) | Numeric |

You also created:

* SES\_Education, SES\_Income, SES\_Access: binary flags for each disadvantage component.

**✅ 3. Demographic Variables (Covariates)**

| **Variable** | **Description** | **Type** |
| --- | --- | --- |
| AGEP\_A | Age in years | Numeric |
| SEX\_A | Sex (1 = Male, 2 = Female) | Categorical |
| RACEALLP\_A (if used) | Race/ethnicity code | Categorical |
| DISAB3\_A, PHQ2SCREEN\_A, GAD2SCREEN\_A (optional) | Disability, depression, anxiety | Binary / Score |

**✅ 4. Optional / Advanced Variables (not yet deeply analyzed)**

| **Variable** | **Description** |
| --- | --- |
| EMDOCCUPN1\_A | Occupation code |
| EMDINDSTN1\_A | Industry sector code |
| EMERG12MTC\_A | Emergency room visits in past 12 months |
| RSNHICOST\_A | Reason for not having insurance: cost-related |

These could be used to explore **occupational risk** or **healthcare utilization**, which you mentioned in your hypothesis.

**📦 Data Volume**

* **Total records** (after filtering for complete cases): ~450–460
* **Long COVID positive cases (LONGCOVD1\_A == 1)**: ~240–250
* **Variables in full dataset**: ~600 (you focused on ~10–15)

**Methods**

**Study Design and Data Source**

This study is a secondary cross-sectional analysis based on publicly available data from the **CDC’s National Health Interview Survey (NHIS) COVID-19 Supplement (2023–2024)**. The NHIS is a nationally representative, household-based survey conducted annually to collect health-related information from U.S. residents. For this project, we focused on adults aged 18–65 who answered questions related to Long COVID, socioeconomic status (SES), and demographic characteristics.

**Sample Selection**

We restricted our analysis to individuals who reported having had COVID-19 and answered the Long COVID question (LONGCOVD1\_A). To assess symptom burden, we included only respondents who provided valid responses to the activity limitation variable (LCVDACT\_A). After filtering for complete cases across SES and covariate variables, the final analytic sample comprised **approximately 450 individuals** with Long COVID.

**Variables**

**Outcome Variables**

* **Long COVID Prevalence:** Binary indicator (Yes/No) from LONGCOVD1\_A, based on self-reported symptoms lasting more than four weeks after COVID-19 infection.
* **Symptom Severity:** Ordinal outcome based on LCVDACT\_A, recoded into three categories: “Not at all”, “A little”, and “A lot” of activity limitation due to Long COVID.

**Main Predictors**

* **Education Level (EDUCP\_A):** Categorical variable recoded into levels such as ≤ high school, some college, and ≥ college degree.
* **Work-Related Income (INCWRKO\_A):** Categorical measure of income from work, grouped into four tiers from lowest to highest.
* **Healthcare Access Barrier (HISTOPCOST\_A):** Binary indicator for whether the respondent delayed or skipped medical care due to cost in the past 12 months.

**Composite SES Index**

A **composite SES score** was created by assigning 1 point each for:

* Low education (≤ high school),
* Low income (lowest two income brackets),
* Cost-related access barrier (HISTOPCOST\_A = “Yes”). This index ranged from 0 (high SES) to 3 (low SES).

**Covariates**

* **Age (AGEP\_A):** Continuous.
* **Sex (SEX\_A):** Categorical (Male/Female).

**Statistical Analysis**

**1. Descriptive Analysis**

Summary statistics were computed for all key variables. Proportional bar plots were generated to visualize relationships between SES and Long COVID outcomes.

**2. Chi-squared Test**

We conducted a Pearson chi-squared test to assess the association between **education level** (EDUCP\_A) and **symptom severity** (LCVDACT\_A).

**3. Cochran-Mantel-Haenszel (CMH) Test**

To examine whether the relationship between education and severity varied by **sex**, we stratified by SEX\_A and applied the CMH test for ordinal outcomes.

**4. Ordinal Logistic Regression**

We used a cumulative logit model to evaluate predictors of **symptom severity**, adjusting for SES index, age, and sex. Odds ratios (ORs) were calculated to interpret the direction and strength of association.

**5. Binary Logistic Regression**

We tested whether **SES** predicted **Long COVID prevalence** using a binary logistic regression.

* **Model 1:** SES index + Age + Sex
* **Model 2:** Adds HISTOPCOST\_A (access barrier)
* **Model 3:** Interaction term (SES × Sex)
* **Model 4:** Separate terms for each SES component (education, income, access)

**6. Visualization**

Bar plots and faceted charts were used to illustrate prevalence of Long COVID across SES levels and stratified by sex. Severity distributions were also visualized across education groups.

All analyses were conducted in **R (version X.X.X)**. Significance was assessed at α = 0.05. Where appropriate, **effect sizes** and **confidence intervals** were reported.

**Results**

**Descriptive Overview**

The final analytic sample included approximately **450 adults** aged 18–65 who reported Long COVID symptoms and had complete data for SES and covariates. Among them, **female respondents comprised over 60%**, and **lower-income groups** (INCWRKO\_A ≤ 4) represented nearly half the sample. Roughly **one-third** reported experiencing activity limitations due to Long COVID.

**1. Chi-squared Test: Education vs Symptom Severity**

A Pearson chi-squared test was used to examine the association between **education level (EDUCP\_A)** and **symptom severity (LCVDACT\_A)**.

* **Result:** χ²(4) = *X.XX*, *p* = 0.09
* **Interpretation:** The association between education and severity was **not statistically significant**, though a **qualitative trend** was observed, where individuals with lower education more frequently reported “a lot” of activity limitation.

**2. CMH Test: Stratified by Sex**

To assess effect modification by sex, a **Cochran-Mantel-Haenszel test** was applied.

* **Result:** CMH statistic = *X.XX*, *p* = 0.12
* **Interpretation:** The stratified analysis showed **no consistent association** across sex strata, suggesting that sex did **not significantly modify** the education–severity relationship.

**3. Ordinal Logistic Regression: Predicting Symptom Severity**

We modeled LCVDACT\_A severity using the composite SES index, age, and sex.

* **SES Index:** OR = 1.18 [95% CI: 0.90–1.53], *p* = 0.23
* **Age:** OR = 1.01 [95% CI: 1.00–1.03], *p* = 0.07
* **Female (ref = Male):** OR = 1.34 [95% CI: 0.95–1.90], *p* = 0.09
* **Interpretation:** No variable reached statistical significance. However, **age and sex showed borderline effects**, indicating a potential trend of **higher severity among older and female participants**.

**4. Binary Logistic Regression: Predicting Long COVID Prevalence**

**Model 1 – SES Index + Age + Sex**

* **SES Index:** OR = 1.20 [95% CI: 0.96–1.51], *p* = 0.11
* **Female Sex:** OR = 1.85 [95% CI: 1.25–2.73], *p* < 0.01
* **Interpretation:** Female sex emerged as a **significant predictor** of Long COVID, while SES showed only a **non-significant upward trend**.

**Model 2 – Adding Healthcare Access Barrier**

* **HISTOPCOST\_A:** OR = 1.31 [95% CI: 0.87–1.99], *p* = 0.19
* **Interpretation:** Delayed care due to cost **did not independently predict** Long COVID when SES and covariates were included.

**Model 3 – SES × Sex Interaction**

* Interaction term: *p* = 0.42
* **Interpretation:** No evidence of interaction; SES effects were not significantly different between sexes.

**Model 4 – Decomposed SES (Education, Income, Access)**

* **Education (low):** OR = 1.25 [p = 0.10]
* **Income (low):** OR = 1.15 [p = 0.22]
* **Access Barrier:** OR = 1.28 [p = 0.17]
* **Interpretation:** None of the components were significant independently, but **education approached borderline significance**.

**5. Exploratory Visualization**

**a) Symptom Severity by Education**

A stacked bar plot showed a **gradient**, with higher proportions of “a lot” of limitation among low-education respondents, supporting the hypothesis visually though not statistically.

**b) Long COVID Prevalence by SES Index**

Proportional bar plots stratified by sex revealed:

* **Trend:** Prevalence of Long COVID increased with SES index (higher risk with lower SES)
* **Sex differences:** This gradient was **more pronounced among females**.

**Discussion**

**Principal Findings**

This study explored whether socioeconomic status (SES) influences the prevalence and severity of Long COVID symptoms in U.S. adults aged 18–65. While visualizations and stratified plots suggested that individuals with lower SES tend to report more severe symptoms and higher prevalence, these trends **did not reach statistical significance** in adjusted models. The **strongest consistent finding** was the association between **female sex and increased risk** of Long COVID, consistent with previous epidemiological literature.

Our **composite SES index**, aggregating education, income, and healthcare access barriers, showed **directionally consistent but statistically weak** associations with both Long COVID prevalence and severity. Decomposing SES into its individual components revealed that **education level** came closest to significance, supporting existing theories that health literacy and educational attainment may influence symptom recognition, reporting behavior, or access to care.

The **lack of significant associations** in regression models may reflect sample size limitations, measurement imprecision in self-reported outcomes, or complex confounding that could not be fully accounted for. The **CMH test** suggested no effect modification by sex, and ordinal logistic models adjusting for age and sex revealed only **borderline trends**, further emphasizing the challenge of identifying robust SES effects in cross-sectional data.

**Interpretation in Context**

The observed association between **female sex and Long COVID** aligns with findings from large-scale studies, which report that women are more likely to experience persistent post-COVID symptoms, potentially due to hormonal, immunological, or healthcare utilization differences.

Our findings on **SES** echo mixed results in the literature. Some studies report clear socioeconomic gradients in Long COVID outcomes, while others, like ours, suggest **attenuated or context-specific effects** once confounding factors are controlled. It is plausible that **access to early treatment**, **occupational exposure**, and **baseline health disparities** interact in non-linear ways that dilute direct SES effects in population-level models.

**Strengths and Limitations**

**Strengths:**

* Use of a large, nationally representative dataset (NHIS)
* Integration of SES as a **composite index**, enhancing interpretability
* Application of both univariate and multivariable models, as well as stratified analyses

**Limitations:**

* **Cross-sectional design** limits causal inference
* **Self-reported outcomes** may be subject to recall bias or differential reporting by SES
* **SES variables were coarse**, potentially obscuring subtle gradients
* The **sample size (~450)**, while adequate for descriptive analysis, limited power in multivariable models
* Potential **residual confounding**, particularly from comorbidities or geographic variation, was not accounted for

**Public Health Implications**

While not conclusive, the trends observed support the need for **targeted outreach** and **equitable healthcare access** strategies. Given the disproportionate burden of Long COVID among women and socioeconomically vulnerable groups, these populations may benefit most from interventions such as symptom monitoring, financial support, and early rehabilitation services.

**Conclusion**

This study investigated the influence of socioeconomic factors—education, income, and healthcare access—on the prevalence and severity of Long COVID symptoms in a nationally representative U.S. sample of adults aged 18–65. While qualitative trends suggested that individuals with lower SES experience a higher burden of Long COVID, these associations did not reach statistical significance in multivariable models. The only consistent and significant predictor of Long COVID prevalence was **female sex**, reinforcing findings from prior research on sex differences in post-viral syndromes.

Despite the absence of strong statistical effects, the observed gradients in descriptive plots and borderline associations with education highlight the potential for **social determinants of health** to shape Long COVID outcomes. The complexity of this relationship—intertwined with access, occupational exposure, and pre-existing disparities—warrants further study, particularly using **longitudinal data** and **more granular SES indicators**.

Future research should focus on larger cohorts, integrate clinical validation of Long COVID diagnoses, and explore mediators such as healthcare utilization and comorbidity burden. Public health strategies aiming to mitigate Long COVID should prioritize **equity in access to care**, especially for socioeconomically disadvantaged populations and women, who appear to carry a disproportionate share of this emerging burden.