

# Preparing the NHANES Data for the Replication of: Association between sleep duration on workdays and blood pressure in non-overweight/obese population in NHANES: a public database research\*

Short title goes here!

Amos Okutse, Brown University      Second Author, Another Uni  
Graduate Student, Graduate School

09 December, 2023

## Abstract

**Introduction:** Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sit amet libero justo. Pellentesque eget nibh ex. Aliquam tincidunt egestas lectus id ullamcorper.

**Methods:** Proin tellus orci, posuere sed cursus at, bibendum ac odio. Nam consequat non ante eget aliquam. Nulla facilisis tincidunt elit. Nunc hendrerit pellentesque quam, eu imperdiet ipsum porttitor ut. Interdum et malesuada fames ac ante ipsum primis in faucibus. Suspendisse potenti. Duis vitae nibh mauris. Duis nec sem sit amet ante dictum mattis. Suspendisse diam velit, maximus eget commodo at, faucibus et nisi. Ut a pellentesque eros, sit amet suscipit eros. Nunc tincidunt quis risus suscipit vestibulum. Quisque eu fringilla massa.

## Introduction

- importance of reproducibility
- As a case on point ...?
- Underscores the importance of reproducibility within the frame of statistical analysis. We use the study by to ... The current study builds on the paper by .. and starts in the manner of a replication study and extends the analyses therein. We focus on statistical analyses of these data and compare models based on varied specification of functional forms of the covariates.

## Methods

### Study population

Data utilized in this paper is sourced from the National Health and Nutrition Examination Survey (NHANES), a comprehensive nationwide survey administered by the National Center for Health Statistics (NCHS) via the Centers for Disease Control and Prevention (CDC). The survey assesses the health and nutrition of the entire non-institutionalized US population, spanning all ages and residing in all 50 states as well as Washington D.C. As such, the survey provides a cross-sectional view of a representative sample of the US population. Further information about NHANES can be found at [www.cdc.gov/nchs/nhanes](http://www.cdc.gov/nchs/nhanes).

---

\*Replication files are available on the author's Github account (<http://github.com/okutse/sleepBP>). **Current version:** December 09, 2023

| Variable  | Name                     | Description  |
|---|--------------------------|--|
| SEQN  | sequence number          | Respondent number  |
| SDMVPSU   | psu                      | Masked variance unit pseudo-PSU variable for variance estimation   |
| WTINT2YR  | weights                  | Full sample 2-year interview weights   |
| SDMVSTRA  | strata                   | Masked variance unit pseudo-stratum variable for variance estimation   |
| RIAGENDR  | gender                   | Respondent's number  |
| RIDAGEYR  | age (yrs)                | Respondent's age in years  |
| DMDMARTL  | marital status           | Marital status   |
| INDFMIN2  | income category          | Total family income (reported as a range value in dollars)   |
| RIDRETH3  | race                     | Recode of reported race and Hispanic origin information, with Non-Hispanic Asian Category  |
| DMDHHSZA  | children <5              | Number of children aged 5 years or younger in the household  |
| DMDDEDUC2   | education level          | What is the highest grade or level of school   |
| {you have/SP has} completed or the highest degree {you have/s/he has} received? |                          |  |
| DMDCITZN  | citizenship status       | {Are you/Is SP} a citizen of the United States? [Information about citizenship is being collected by the U.S. Public Health Serv             |
| SLD012  | sleep                    | Number of hours usually sleep on weekdays or workdays  |
| BMXBMI  | bmi                      | Body mass index  |
| ALQ121  | alcohol use              | In the past 12 months, how often did you drink any type of alcoholic beverage?   |
| LBDHDDSI  | hdl                      | High density lipoprotein   |
| LBDSALSI  | albumin                  | Albumin (g/L)  |
| DIQ010  | diabetes                 | Have you ever been told by a doctor or health professional that you have diabetes or sugar diabetes?   |
| BPQ020  | hypertension             | {Have you/Has SP} ever been told by a doctor or other health professional that {you/s/he} had hypertension, also called high blood pressure? |
| URXCRS  | creatinine               | Creatinine, urine (umol/L)   |
| SLQ040  | snort                    | In the past 12 months, how often did {you/SP} snort, gasp, or stop breathing while {you were/s/he was} asleep?                               |
| LBDTCSI   | total cholesterol levels | Total cholesterol (mmol/L)   |
| LBXHGB  | hemoglobin               | Hemoglobin (g/dL)  |
| LBXSASSI  | AST                      | Aspartate aminotransferase   |
| SMQ040  | smoke                    | Do you now smoke cigarettes?   |
| BPXDI   | blood pressure           | Systolic and diastolic blood pressure taken as the average of the first three measurements   |

## Data

Our current analyses combine the 2015 - 2018 NHANES survey cycles to yield  $n = 19225$  observations on 35 covariates. This sample size was comprised  $n = 9971$  and  $n = 9254$  observations from the 2015/2016 and 2017/2018 survey cycles, respectively. Analyses excluded individuals with missing data on sleep ( $n = 6818$ ), blood pressure (BP) ( $n = 1055$ ), and body mass index (BMI) or those with BMI  $> 25 \text{ kg/m}^2$  (overweight) ( $n = 5521$ ). Individuals that reported being on anti-hypertensive medication were also excluded from further analyses ( $n = 2944$ ).

### Outcome definition:

We defined our outcome as systolic and diastolic blood pressure. These variables are measured by trained examiners using standardized procedures. Given that systolic and diastolic blood pressure measurements are taken at least four times on an individual, our definition of these outcome is based on an average of the first three blood pressure measurements.

### Exposure:

Sleep duration on workdays was evaluated by the questionnaire with the following questions: “Number of hours usually sleep on weekdays or workdays”. Sleep duration was divided into three groups, which were  $< 6$  h, 6–8 h, 8 h respectively, of which 6–8 h was used as the reference group.

### Covariates:

In addition to the exposure and outcome variables, our analyses included the following as covariates: Race divided into four groups as Mexican American, white, black and other race. Alcohol consumption was grouped into drinking, no drinking, not recorded. Smoking status as smoking, not smoking, not recorded. Diabetes was defined as yes, no, borderline, or not recorded. Hypertension was defined as yes, no, or not recorded. Snorting was defined as yes, no, and not recorded. US citizenship status was defined as citizen by birth or naturalization, don't know, not a citizen, or refused to answer. Education level was grouped into four categories including graduate studies, high school, less than grade 12 or some college. Additional covariates included marital status, gender, age, albumin, creatinine, hemoglobin, total cholesterol (TC), aspartate aminotransferase (AST), high-density lipoprotein (HDL), and body mass index (BMI). Details about these variables can be found at <https://wwwn.cdc.gov/nchs/nhanes/search/default.aspx>. Age, albumin, creatinine, hemoglobin, TC, AST, HDL, and BMI were analyzed as continuous variables whereas gender, alcohol consumption, diabetes, smoking, race, hypertension, and snorting were analyzed as categorical variables. Table @ref{tab:tabone} highlights the variable names and descriptions as utilized in this study.

## Statistical modeling

With the purpose of identifying influential variables impacting blood pressure, a variable selection process was conducted. The initial approach involved employing best subset selection with 10-fold cross-validation to ascertain the optimal number of subsets. Notably, for systolic blood pressure, the forward best subset selection method discerned 13 variables, whereas diastolic blood pressure exhibited 19 selected variables. However, a limitation inherent in forward best subset selection is its inability to eliminate previously selected features, potentially disregarding their relevance in light of newly added variables. Consequently, this methodology may yield suboptimal variable selections due to its lack of adaptability.

In an effort to mitigate this limitation and introduce regularization, LASSO regularization was employed. This involved a 10-fold cross-validation process to determine the optimal penalty parameter,  $\lambda$ . For diastolic blood pressure, the  $\lambda$  value associated with the lowest mean squared error (MSE) led to the selection of 25 variables. Conversely, when considering systolic blood pressure, the  $\lambda$  value minimizing the MSE resulted in the exclusion of a sole variable. Subsequently, a  $\lambda$  value was strategically chosen to ensure a negligible increase of no more than 1% in MSE, ultimately leading to the inclusion of 29 variables in the model.

## Results

Table 1 summarizes participant characteristics in the 2015 - 2018 NHANES survey cycles stratified by gender. Among the participants, the proportion of males and females were 44.81% ( $n = 1378$ ) and 55.19% ( $n = 1509$ ), respectively. With ethnicity, the proportion of Mexican American, White and Black were 6.27%, 64.62% and 9.94%, respectively. Overall, the mean (SD) values for age, albumin, creatinine, SBP, DBP, hemoglobin, TC, AST, HDL, BMI were 38.54 (17.72) years, 43.48 (3.58) g/L, 73.21 (19.73)  $\mu\text{mol/L}$ , 115.33 (14.56) mmHg, 68.47 (10.08) mmHg, 14.07(1.44)g/dL, 4.67 (1.01) mmol/L, 23.18 (13.01) IU/L, 1.59 (0.43) mmol/L, and 21.97(2.06) kg/m<sup>2</sup>, respectively. Among the participants, 66.01% were consumed alcohol drinkers, 2.61% were diabetes, 3.16% were hypertension, 6.72% have snort or stop breathing. 17.99% were smokers. Sleep duration was divided into three groups, which were  $< 6$  h, 6–8 h,  $\geq 8$  h, each with a proportion of 6.26%, 40.53%, 53.21%, respectively.

Male and female participants differed significantly on characteristics including

The univariate analysis of potential confounding factors is shown in Table 2.

```
## Warning in styling_latex_scale_down(out, table_info): Longtable cannot be
## resized.
```

Table 1: Unweighted summary characteristics of the participants in the NHANES 2015 - 2018 survey cycles stratified by gender.

| Variable            | Overall, N = 1,977   | Female, N = 1,063   | Male, N = 914        | p-value |
|---------------------|----------------------|---------------------|----------------------|---------|
| <b>BMI</b>          | 22.08 (2.03)         | 21.83 (2.04)        | 22.37 (1.98)         | <0.001  |
| <b>HDL</b>          | 1.60 (0.45)          | 1.73 (0.45)         | 1.45 (0.40)          | <0.001  |
| <b>TC</b>           | 4.81 (1.02)          | 4.87 (1.05)         | 4.73 (0.99)          | 0.005   |
| <b>Hemoglobin</b>   | 13.94 (1.46)         | 13.20 (1.18)        | 14.81 (1.28)         | <0.001  |
| <b>Albumin</b>      | 43.04 (3.59)         | 42.40 (3.36)        | 43.78 (3.70)         | <0.001  |
| <b>AST</b>          | 23.47 (13.17)        | 21.59 (9.78)        | 25.66 (15.98)        | <0.001  |
| <b>Creatinine</b>   | 10,132.60 (7,349.24) | 8,715.09 (6,809.53) | 11,781.19 (7,609.15) | <0.001  |
| <b>Hypertension</b> |                      |                     |                      | 0.002   |
| Don't know          | 4.00 (0.20%)         | 2.00 (0.19%)        | 2.00 (0.22%)         |         |
| No                  | 1,819.00 (92.01%)    | 998.00 (93.89%)     | 821.00 (89.82%)      |         |
| Yes                 | 154.00 (7.79%)       | 63.00 (5.93%)       | 91.00 (9.96%)        |         |
| <b>Diabetes</b>     |                      |                     |                      | 0.010   |
| Borderline          | 33.00 (1.67%)        | 15.00 (1.41%)       | 18.00 (1.97%)        |         |
| No                  | 1,844.00 (93.27%)    | 1,008.00 (94.83%)   | 836.00 (91.47%)      |         |

|                                  |                   |                 |                 |        |
|----------------------------------|-------------------|-----------------|-----------------|--------|
| Yes                              | 100.00 (5.06%)    | 40.00 (3.76%)   | 60.00 (6.56%)   |        |
| <b>Citizenship</b>               |                   |                 |                 | 0.2    |
| Citizen by birth or naturalizati | 1,616.00 (81.74%) | 858.00 (80.71%) | 758.00 (82.93%) |        |
| Don't Know                       | 1.00 (0.05%)      | 1.00 (0.09%)    | 0.00 (0.00%)    |        |
| Not a citizen of the US          | 357.00 (18.06%)   | 201.00 (18.91%) | 156.00 (17.07%) |        |
| Refused                          | 3.00 (0.15%)      | 3.00 (0.28%)    | 0.00 (0.00%)    |        |
| <b>Education</b>                 |                   |                 |                 | <0.001 |
| GraduateStudies                  | 646.00 (32.68%)   | 394.00 (37.06%) | 252.00 (27.57%) |        |
| Highschool                       | 415.00 (20.99%)   | 192.00 (18.06%) | 223.00 (24.40%) |        |
| Less12grade                      | 359.00 (18.16%)   | 143.00 (13.45%) | 216.00 (23.63%) |        |
| someCollege                      | 557.00 (28.17%)   | 334.00 (31.42%) | 223.00 (24.40%) |        |
| <b>Children &gt; 5 yrs</b>       |                   |                 |                 | 0.002  |
| 0                                | 1,547.00 (78.25%) | 798.00 (75.07%) | 749.00 (81.95%) |        |
| 1                                | 274.00 (13.86%)   | 168.00 (15.80%) | 106.00 (11.60%) |        |
| 2                                | 123.00 (6.22%)    | 74.00 (6.96%)   | 49.00 (5.36%)   |        |
| 3 or more                        | 33.00 (1.67%)     | 23.00 (2.16%)   | 10.00 (1.09%)   |        |
| <b>Age (yrs)</b>                 | 43.79 (17.43)     | 42.77 (16.65)   | 44.98 (18.24)   | 0.024  |
| <b>Marital status</b>            |                   |                 |                 | 0.003  |
| Divorced                         | 180.00 (9.10%)    | 111.00 (10.44%) | 69.00 (7.55%)   |        |
| Living with partner              | 197.00 (9.96%)    | 93.00 (8.75%)   | 104.00 (11.38%) |        |
| Married                          | 961.00 (48.61%)   | 530.00 (49.86%) | 431.00 (47.16%) |        |
| Never married                    | 504.00 (25.49%)   | 245.00 (23.05%) | 259.00 (28.34%) |        |
| Separated                        | 54.00 (2.73%)     | 33.00 (3.10%)   | 21.00 (2.30%)   |        |
| Widowed                          | 81.00 (4.10%)     | 51.00 (4.80%)   | 30.00 (3.28%)   |        |
| <b>Survey cycle</b>              | 914.00 (46.23%)   | 502.00 (47.22%) | 412.00 (45.08%) | 0.3    |
| <b>DBP</b>                       | 69.60 (11.18)     | 68.79 (10.70)   | 70.55 (11.65)   | <0.001 |
| <b>SBP</b>                       | 118.61 (16.41)    | 115.78 (16.64)  | 121.90 (15.52)  | <0.001 |
| <b>Sleep</b>                     |                   |                 |                 | 0.005  |
| <6hrs                            | 156.00 (7.89%)    | 73.00 (6.87%)   | 83.00 (9.08%)   |        |
| >8hrs                            | 1,029.00 (52.05%) | 588.00 (55.32%) | 441.00 (48.25%) |        |
| 6-8hrs                           | 792.00 (40.06%)   | 402.00 (37.82%) | 390.00 (42.67%) |        |
| <b>Race</b>                      |                   |                 |                 | <0.001 |
| Black                            | 338.00 (17.10%)   | 143.00 (13.45%) | 195.00 (21.33%) |        |
| Mexican American                 | 181.00 (9.16%)    | 96.00 (9.03%)   | 85.00 (9.30%)   |        |
| Other                            | 768.00 (38.85%)   | 455.00 (42.80%) | 313.00 (34.25%) |        |
| White                            | 690.00 (34.90%)   | 369.00 (34.71%) | 321.00 (35.12%) |        |
| <b>Smoking status</b>            |                   |                 |                 | <0.001 |
| Not recorded                     | 1,229.00 (62.16%) | 763.00 (71.78%) | 466.00 (50.98%) |        |
| Not Smoking                      | 321.00 (16.24%)   | 137.00 (12.89%) | 184.00 (20.13%) |        |
| Smoking                          | 427.00 (21.60%)   | 163.00 (15.33%) | 264.00 (28.88%) |        |
| <b>Snort</b>                     |                   |                 |                 | <0.001 |
| No                               | 1,590.00 (80.42%) | 902.00 (84.85%) | 688.00 (75.27%) |        |
| Not recorded                     | 115.00 (5.82%)    | 49.00 (4.61%)   | 66.00 (7.22%)   |        |
| Yes                              | 272.00 (13.76%)   | 112.00 (10.54%) | 160.00 (17.51%) |        |
| <b>Alcohol</b>                   |                   |                 |                 | <0.001 |
| Drinking                         | 1,457.00 (73.70%) | 737.00 (69.33%) | 720.00 (78.77%) |        |
| No drinking                      | 111.00 (5.61%)    | 52.00 (4.89%)   | 59.00 (6.46%)   |        |
| Not recorded                     | 409.00 (20.69%)   | 274.00 (25.78%) | 135.00 (14.77%) |        |
| <b>Income</b>                    |                   |                 |                 | <0.001 |
| Low income                       | 131.00 (6.63%)    | 64.00 (6.02%)   | 67.00 (7.33%)   |        |

|                     |                 |                 |                 |
|---------------------|-----------------|-----------------|-----------------|
| Lower-middle income | 389.00 (19.68%) | 187.00 (17.59%) | 202.00 (22.10%) |
| Middle income       | 550.00 (27.82%) | 284.00 (26.72%) | 266.00 (29.10%) |
| Unknown/Refused     | 58.00 (2.93%)   | 27.00 (2.54%)   | 31.00 (3.39%)   |
| Upper-middle income | 391.00 (19.78%) | 210.00 (19.76%) | 181.00 (19.80%) |
| Varied/High income  | 458.00 (23.17%) | 291.00 (27.38%) | 167.00 (18.27%) |

<sup>1</sup> Mean (SD); n (%)

<sup>2</sup> Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test

### Foward Stepwise Best Subset

$$\begin{aligned}
\text{sbp} = & 85.65 + 0.47 \times \text{BMI} + 1.49 \times \text{total\_chol} - 0.00 \times \text{creatinine} \\
& + 3.06 \times \text{educ\_levelHighschool} + 2.00 \times \text{educ\_levelLess12grade} + 1.31 \times \text{educ\_levelsomCollege} \\
& + 4.59 \times \text{genderMale} + 0.47 \times \text{age\_yr} + 2.05 \times \text{marital\_statusLiving with partner} \\
& - 0.12 \times \text{marital\_statusMarried} + 3.30 \times \text{marital\_statusNever married} \\
& - 0.99 \times \text{marital\_statusSeparated} + 7.01 \times \text{marital\_statusWidowed} - 1.30 \times \text{sleep>8hrs} \\
& - 2.82 \times \text{sleep6-8hrs} - 4.60 \times \text{raceMexican American} - 3.78 \times \text{raceOther} \\
& - 5.03 \times \text{raceWhite income\_categoryLower-middle income} - 3.02 \times \text{income\_categoryMiddle income} \\
& - 2.88 \times \text{income\_categoryUnknown/Refused} - 2.70 \times \text{income\_categoryUpper-middle income} \\
& - 4.30 \times \text{income\_categoryVaried/High income}
\end{aligned}$$

$$\begin{aligned}
\text{dbp} = & 53.55 + 0.22 \times \text{BMI} + 0.78 \times \text{total\_chol} + 0.85 \times \text{hemoglobin} + 0.02 \times \text{albumin} \\
& + 0.03 \times \text{AST} - 0.89 \times \text{diabetesNo} - 3.37 \times \text{diabetesYes} - 1.06 \times \text{children..51} \\
& - 1.67 \times \text{children..52} - 1.83 \times \text{children..53 or more} - 1.90 \times \text{marital\_statusLiving with partner} \\
& + 0.00 \times \text{marital\_statusMarried} - 2.78 \times \text{marital\_statusNever married} + 0.91 \times \text{marital\_statusSeparated} \\
& + 1.72 \times \text{marital\_statusWidowed} - 2.40 \times \text{sleep>8hrs} - 2.37 \times \text{sleep6-8hrs} - 2.76 \times \text{raceMexican American} \\
& + 0.14 \times \text{raceOther} - 1.98 \times \text{raceWhite} - 2.16 \times \text{snortNot recorded} \\
& - 0.57 \times \text{snortYes} - 2.33 \times \text{alcoholNo drinking} - 1.79 \times \text{alcoholNot recorded}
\end{aligned}$$

## LASSO

$$\begin{aligned} \text{sbp} = & 84.13 \times (\text{Intercept}) + 0.52 \times \text{bmi} + 0.79 \times \text{hdl} + 1.43 \times \text{total\_chol} + 0.04 \times \text{AST} \\ & - 0.17 \times \text{diabetesNo} + 1.88 \times \text{diabetesYes} - 14.02 \times \text{citizenship\_statusDon't Know} \\ & - 0.59 \times \text{citizenship\_statusNot a citizen of the US} - 12.06 \times \text{citizenship\_statusRefused} \\ & + 3.09 \times \text{educ\_levelHighschool} + 2.19 \times \text{educ\_levelLess12grade} + 1.16 \times \text{educ\_levelsCollege} \\ & - 0.99 \times \text{children..51} - 1.17 \times \text{children..52} + 0.90 \times \text{children..53 or more} \\ & + 4.42 \times \text{genderMale} + 0.44 \times \text{age\_yr} + 1.77 \times \text{marital\_statusLiving with partner} \\ & + 0.01 \times \text{marital\_statusMarried} + 2.95 \times \text{marital\_statusNever married} - 1.05 \times \text{marital\_statusSeparated} \\ & + 7.19 \times \text{marital\_statusWidowed} - 1.47 \times \text{sleep>8hrs} - 2.95 \times \text{sleep6-8hrs} \\ & - 4.11 \times \text{raceMexican American} - 3.43 \times \text{raceOther} - 5.14 \times \text{raceWhite} \\ & + 1.75 \times \text{smokeNot Smoking} + 0.22 \times \text{smokeSmoking} - 2.74 \times \text{snortNot recorded} \\ & - 0.61 \times \text{snortYes} - 1.69 \times \text{alcoholNo drinking} - 0.28 \times \text{alcoholNot recorded} \\ & - 1.62 \times \text{income\_categoryLower-middle income} - 3.04 \times \text{income\_categoryMiddle income} \\ & - 2.64 \times \text{income\_categoryUnknown/Refused} - 2.71 \times \text{income\_categoryUpper-middle income} \\ & - 4.54 \times \text{income\_categoryVaried/High income} \end{aligned}$$

$$\begin{aligned} \text{dbp} = & 47.99 + 0.24 \times \text{bmi} + 0.80 \times \text{total\_chol} + 0.58 \times \text{hemoglobin} + 0.17 \times \text{albumin} \\ & + 0.04 \times \text{AST} - 0.42 \times \text{diabetesNo} - 3.18 \times \text{diabetesYes} + 1.36 \times \text{citizenship\_statusDon't Know} \\ & - 0.60 \times \text{citizenship\_statusNot a citizen of the US} - 1.78 \times \text{citizenship\_statusRefused} \\ & - 0.11 \times \text{educ\_levelHighschool} - 0.92 \times \text{educ\_levelLess12grade} - 0.39 \times \text{educ\_levelsCollege} \\ & - 0.99 \times \text{children..51} - 1.39 \times \text{children..52} - 2.08 \times \text{children..53 or more} \\ & + 0.59 \times \text{genderMale} + 0.01 \times \text{age\_yr} - 1.77 \times \text{marital\_statusLiving with partner} \\ & + 0.12 \times \text{marital\_statusMarried} - 2.67 \times \text{marital\_statusNever married} + 0.94 \times \text{marital\_statusSeparated} \\ & + 1.69 \times \text{marital\_statusWidowed} + 2.21 \times \text{cycle} - 2.21 \times \text{sleep>8hrs} \\ & - 2.25 \times \text{sleep6-8hrs} - 2.08 \times \text{raceMexican American} + 0.31 \times \text{raceOther} \\ & - 2.09 \times \text{raceWhite} + 0.44 \times \text{smokeNot Smoking} + 0.75 \times \text{smokeSmoking} \\ & - 2.15 \times \text{snortNot recorded} - 0.50 \times \text{snortYes} - 1.28 \times \text{alcoholNo drinking} \\ & - 1.17 \times \text{alcoholNot recorded} \end{aligned}$$

## Discussion and conclusion

## References