

Project 1: Epidemiology and Lifestyle Determinants of High Blood Pressure in the United States

Introduction to Statistical Thinking and Data Analysis

MSc in Epidemiology and MSc in Health Data Analytics, Imperial College London

17 October 2022

Background

High blood pressure (hypertension) is one of the leading risk factors for stroke, heart disease, and kidney disease.¹ Globally, high blood pressure is estimated to be the risk factor responsible for preventable morbidity and mortality in 2017.²

Averting the negative health impacts of cardiovascular and kidney disease requires identifying populations at risk of high blood pressure who would benefit from public health interventions to treat and prevent high blood pressure.

Though effective medicines for high blood pressure exist (ACE inhibitors, ARBs, calcium channel blockers),³ lifelong treatment to reduce blood pressure are expensive for the health system and may be undesirable for patients. Previous research has suggested several mutable lifestyle factors that are associated with high blood pressure and may be targets for intervention for persons wishing to reduce their blood pressure. Targeting public health promotion required identification of lifestyle factors that are most strongly and directly related to high blood pressure in the American population.

Aims

1. Describe the epidemiology of blood pressure among adults in the United States, including estimating mean blood pressure and identifying demographic groups with higher average blood pressure.
2. Identify lifestyle risk factors that could be targeted in public health campaigns to reduce the prevalence of high blood pressure.

Data

The *National Health and Nutrition Examination Survey* (NHANES) is a nationally representative survey to assess the health and nutrition of adults and children in the United States. The survey was first conducted in the 1960s and has been conducted continuously since 1999 with around 5000 respondents sampled and interviewed in their homes every year. The survey consists of a combination of questionnaire responses and physical and biomarker measurements. The NHANES interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests administered by highly trained medical personnel.

Findings from the survey are used to determine the prevalence of major diseases and risk factors for diseases. Information are used to assess nutritional status and its association with health promotion and disease

prevention. NHANES findings are also the basis for national standards for such measurements as height, weight, and blood pressure. Data from this survey are used in epidemiological studies and health sciences research, which help develop sound public health policy, direct and design health programs and services, and expand the health knowledge.

More information about the survey objectives, design, and procedures are available from the website: https://www.cdc.gov/nchs/nhanes/about_nhanes.htm

For this project, we will use an extract of the NHANES dataset consisting of an extract of 75 variables about 10,000 respondents to NHANES between 2009 and 2012 abstracted for educational purposes. The actual NHANES survey datasets include sampling weights to account for non-equal sampling probability of certain population groups to increase the statistical efficiency of the survey, which is not covered in this course. The dataset of 10,000 respondents in the NHANES R package has been constructed such that the dataset can be analysed as if it were a simple random sample from the American population. See the package documentation for information and links about accessing and analysing the actual NHANES data for research purposes; there are other R packages available on CRAN to assist with accessing and processing the actual NHANES survey data.

Access the dataset in R and review the dataset documentation via the following commands:

```
install.packages("NHANES")
library(NHANES)
data(NHANES)
?NHANES
```

Notes

- Typically, **systolic blood pressure** is monitored most closely as research has found that greater risk of stroke and heart disease is most strongly related to elevated systolic blood pressure. The variable **BPSysAve** in the NHANES dataset reports a consensus systolic blood pressure measurement for each participant combined over three readings.
- Make sure that your presentation explicitly addresses both of the Aims 1 and 2. For each aim define and test **one or two hypotheses** based on previous research or hypotheses about correlates and determinants of high blood pressure.
 - For Aim 1: examples of relevant demographic groups across which to compare blood pressure could be defined based on gender, age, race, education level, or income level (see the ‘Demographic Variables’ section in the help at `?NHANES`).
 - For Aim 2: risk factors of interest may include physical activity, BMI, alcohol use, or smoking status (see the ‘Lifestyle Variables’ section in the help at `?NHANES`).
- You are *not* expected or encouraged to do substantial research on blood pressure or risk factors beyond the information presented herein. Focus effort on defining testable hypotheses and clearly reporting your analysis to address the hypotheses.

Project Outline

Week 1 (17 October): Develop an analysis plan

- Based on the Aims described above, specify research questions that you will address.
 - Identify 2-3 specific hypotheses that you will test.

- Define the dataset and any inclusion or exclusion criteria for the dataset to be analysed.
- Identify the variables that you will use in your main analyses, including the primary outcome variable(s), main exposure variable(s), and any other covariates you will consider. Define any calculated variables required for your analysis.
 - This may include exploratory data analysis to understand the variables in the dataset, for example identifying quality of measurement or amount of missing data for certain variables.
- Agree the statistical methods you will use for descriptive analyses and to address each of your hypotheses. Define table shells for reporting the results of your analysis.

After 45 minutes: Each group presents one slide stating your hypotheses. **15 minutes before end of session:** Each group present one slide stating the main outcome and exposure variables for each hypothesis and any exclusion criteria.

Week 2 (24 October): Statistical analysis and interpretation

- Carry out your statistical analysis plan developed in Week 1.
- Interpret the results of your analysis with respect to each of your research questions and hypotheses.
- Check any important assumptions underpinning your analysis and conclusions and conduct any sensitivity analyses to confirm the robustness of your results. Enumerate any important limitations to your results or interpretations.
- Contextualize your findings compared to other existing evidence.
- Identify key recommendations from your study for advanced scientific understanding, public health policy or practice, and future research priorities.
- Prepare a 10-minute presentation of your results, including:
 - Background, research question, and hypotheses.
 - Methods: variables used and statistical analyses.
 - Results
 - Interpretation and conclusions

After 1 hour: Each group present one to three slides of your descriptive analysis (Table 1 or key figures).

Week 3 (31 October): Present results

There will be **10 minutes** at the start of the session for any final coordination of your presentation. Following this, each group will give a 10 minute presentation of their results followed by three minutes for questions and discussion from the audience.

After the presentations have completed, each group will debrief Project 1 by identifying:

- One thing your group did well.
- One thing you would do differently next time.
- One thing you particularly liked about a different group's analysis or presentation.

References

1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19.1 million participants. *Lancet* 2017; 389:37-55.

2. GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; 392:1923-1994.
3. <https://www.nhs.uk/conditions/high-blood-pressure-hypertension/treatment/>