**Cardiometabolic statistics for US adults: an open source platform for exploratory analysis**

Byron C. Jaeger, Ligong Chen, Adam P. Bress, <OTHERS FROM JHS HWG WRITING RETREAT>, and Paul Muntner

INTRODUCTION

* Describe the NHANES, quantify its impact:
  + Goal of NHANES
  + Data collected in NHANES
  + Used in many CDC reports but also is publically available.
  + How many papers on hypertension/BP use NHANES data?
  + Healthcare policies/guidelines in the US are informed by NHANES data
    - 2017 ACC/AHA BP guideline (>10,000 citations in < 5 years)
    - Potential impact of the guideline (>900 citations in < 5 years)
* Identify problem: Analyzing NHANES data requires expertise – statistical and familiarity with the data. Challenges include retrieving the correct files from the NHANES database, deriving variables suited to the current analysis, inclusion and exclusion of participants, and carrying out valid statistical analysis accounting for the complex survey design of NHANES.
* Proposed solution: We develop a free online platform that allows users to obtain visual and tabular summaries of the NHANES blood pressure and hypertension data.
* Specific goals:
  + Increase the usage of NHANES data to quantify public health outcomes
  + Improve the validity, consistency, and replicability of results reported from NHANES

METHODS

* The NHANES data and protocols for data collection
* NHANES data used in the current analysis:
  + Which survey participants were included (**Figure S1**) and why
  + NHANES data we included and variables derived from these data (**Table S1**)
* Statistical Analysis
  + Generate tabular or visual summaries of NHANES data that can be saved by the user
  + Pool results over contiguous NHANES cycles or stratify results by cycle
    - If stratifying by cycle, user can adjust for age differences using direct standardization
  + Weights are re-calibrated to match the population size of the US whenever a user requests to estimate the count of US citizens. The re-calibration of weights does not impact the prevalence estimates (b/c they are done separately).
  + Subset the data to be summarized, e.g., by including or excluding adults who report taking antihypertensive medications.
  + Stratify results across subgroups, e.g., compute the mean systolic blood pressure among men and women, separately.
  + Follow CDC recommendations on reliability of results (Ligong can help write this)
    - Define criteria (perhaps in a supplemental table if space is tight)
* Development and validation of the app
  + We used R version 4.1.2 to develop a web application using the “Shiny” R package
  + We disseminate the app publicly online
  + We disseminate code and data from the app in an open-source R Package
  + Tests for validity of the app included replication of summary statistics from
    - From Muntner et al, JAMA 2020
    - From <https://www.cdc.gov/nchs/data/databriefs/db289.pdf>
    - Mention Carey et al but do not indicate that our results match (they do not).
* Application of the app in this analysis
  + Describe how we will report the characteristics of participants in this analysis, and note that this is the only result that the app does not generate.
  + Identify statistics we will compute (mean, quantiles, prevalence, and counts)
  + Identify outcomes we will report (hypertension, antihypertensive medication usage, resistant hypertension, and systolic and diastolic blood pressure)
  + Identify subgroups we will compute outcomes for (overall, adults with hypertension, adults with hypertension and using antihypertensive medication). Note that we do not include pregnant women, but the app allows for that.

RESULTS

* Characteristics of US adults included (**Table 1**)
* Prevalence of hypertension **in the overall sample**
  + By calendar period (**Figure 1**)
  + By age group, sex, and race-ethnicity in 2017-2020 (**Figure S2**)
* Antihypertensive medication use **among those with hypertension**
  + Prevalence of antihypertensive medication use by calendar period (**Figure 2**)
  + Number of antihypertensive medication classes by calendar period (**Figure 3**)
* Prevalence of different antihypertensive medications being taken **among those with hypertension**
  + Identify 3 or 4 specific medications (**1 main figure, 2 or 3 supplemental figures**)
* Prevalence of resistant hypertension among those taking antihypertensive medication
  + By calendar period (**Figure 5**)
  + By age group, sex, and race-ethnicity in 2017-2020 (**Figure S**)
* Mean SBP and DBP
  + SBP By calendar year (**Figure S**)
  + DBP by calendar year (**Figure S**)
  + By age group, sex, race/ethnicity (**Figure S**)
* A supplemental set of “tutorial videos”, one video for each piece of the results (roughly 1-2 minutes per video), where the first author shows how to make all of the main figures using the shiny app?

DISCUSSION

* First paragraph: summary.
  + We made an application that allows anyone to obtain valid summaries of US population statistics for cardiometabolic outcomes.
  + We tested the application by replicating prior work with it
  + Our results show that one can replicate prior NHANES results and extend them
* Second paragraph: rationale for our design
  + Recalibration of weights
  + Adults only
  + Coding of medications
  + Requiring only a single BP to be included
* Third paragraph: Who may benefit from this app
  + In research settings, many population scientists may wish to engage with NHANES data, but do not due to the difficulty of and time required to retrieve the data, derive the relevant variables, and account for the complex survey design of NHANES.
  + In educational settings, many teachers in public health may wish to show recent trends in health outcomes, but do not due to the same obstacles above.
  + Investigators writing grant submissions may wish to cite recent statistics regarding US adults, but it is difficult to identify prior papers providing the specific data they want to report.
* Fourth paragraph: extending this app
  + We have presented BP outcomes but have designed the app to be modular and we plan to add several other cardiometabolic outcomes in future updates.
  + The app can incorporate the new NHANES cycle following their public release
  + The app is flexible to include other data sets.
  + Others can develop or contribute to extensions of this application by collaborating with us on GitHub

Strengths and limitations

* Strengths:
  + We validated our app using prior papers and reports from the CDC
  + We leveraged open-source software to maximize the accessibility of our platform.
  + A multi-disciplinary team with expertise in all areas of BP, statistics, and software collaborated in the design of this platform to increase its accessibility and validity.
  + Platform includes tools that make it easy to save results from exploratory analyses.
* Limitations:
  + App UI is complex
  + App may have bugs that require updates in the future. Notifying all possible users of the app when these fixes occur may be challenging.
  + App comes with no warranty whatsoever, as is common with open-source software.

Conclusion

This application may push scientific reporting further into the online and open-source environment

MAIN TABLES

**Table 1**: characteristics of NHANES participants included in the data for the web application.

MAIN FIGURES

TO BE FILLED IN LATER

SUPPLEMENTAL TABLES

**Table S1**: Variables used in the current analysis

|  |  |  |
| --- | --- | --- |
| Variable | Source | Description |
| Demographics | | |
| Age, years | DEMO | Age in years of the survey participant at the time of screening. Individuals 80 years or older are coded at 80 years of age. |
| Age category, years | Derived | Five categories are derived:  18 to 44 years  45 to 64 years  65 to 74 years  75+ years |
| Race | DEMO | Non-Hispanic White  Non-Hispanic Black  Non-Hispanic Asian  Hispanic  Other |
| Pregnant | DEMO | Yes  No |
| Gender | DEMO | Female  Male |
| Blood Pressure | | |
| SBP, mm Hg | BPX | Three consecutive BP measurements (systolic and diastolic) were taken 60 seconds apart |
| DBP, mm Hg | BPX |
| BP category | Derived | SBP <120 and DBP <80 mm Hg  SBP of 120 to <130 and DBP <80 mm Hg  SBP of 130 to <140 or DBP 80 to <90 mm Hg  SBP of 140 to <160 or DBP 90 to <100 mm Hg  SBP 160+ or DBP 100+ mm Hg |
| BP control |  |  |
| JNC7 definition | Derived | SBP < 140 mm Hg and DBP < 90 mm Hg |
| 2017 ACC/AHA definition | Derived | SBP < 130 mm Hg and DBP < 80 mm Hg |
| Uncontrolled BP |  |  |
| JNC7 definition | Derived | SBP >= 140 mm Hg or DBP >= 90 mm Hg |
| 2017 ACC/AHA definition | Derived | SBP >= 130 mm Hg or DBP >= 80 mm Hg |

BP = blood pressure, SBP = systolic blood pressure, DBP = diastolic blood pressure

SUPPLEMENTAL FIGURES

**Figure S1**: cascade of inclusions, use a similar design as the inclusion/exclusion cascade in the JAMA paper

Figures S2 and above to be filled in later