Cardiometabolic statistics for US adults: an open source web application for exploratory analysis

Byron C. Jaeger, PhD,1 Ligong Chen, PhD,2 Adam P. Bress, PharmD, MS,3,4 YOUR NAME, DEGREE,x Lisandro D. Colantonio, PhD,2 Paul Muntner, PhD,2

1Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, NC. 2Department of Epidemiology, University of Alabama at Birmingham. 3Informatics, Decision-Enhancement, and Analytic Sciences (IDEAS) Center, Veterans Affairs, Salt Lake City Health Care System, Salt Lake City, UT. 4Department of Population Health Sciences, University of Utah School of Medicine, Salt Lake City, UT. x YOUR AFFILIATION

**ABSTRACT**

The National Health and Nutrition Examination Survey (NHANES) is a program conducted by the US National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC) and is designed to assess the health and nutritional status of the US population. The NHANES program was initiated in the early 1960s and beginning in 1999 has been conducted continuously, in two-year cycles. NHANES data are collected through participant interviews and physical examinations, each conducted following standardized protocols. The NHANES interview includes questions about prior diagnoses of hypertension and self-reported antihypertensive medication use and a medication inventory where containers for prescription medications are reviewed and recorded. The physical examination includes blood pressure (BP) measurements by trained and certified physicians.

NHANES data have been used to provide estimates of the proportion of US adults with hypertension, aware they had hypertension, taking antihypertensive medication and with controlled BP. Data from NHANES have been used to estimate the impact of the lower BP levels used to define hypertension and controlled BP in the 2017 American College of Cardiology/American Heart Association (ACC/AHA) BP guideline versus the Seventh Joint National Committee (JNC7) BP guideline on the proportion and number of US adults with hypertension, recommended antihypertensive medication and with uncontrolled BP. Additionally, NHANES data have been used to track the proportion of US adults with controlled BP, which were included in a Call-to-Action to Control BP from the US Surgeon General.

NHANES data are publically available and easily accessible through the CDC website. However, analyzing NHANES data may be challenging for many people due to the complexity of its study design. The NHANES program uses a multi-stage sampling design to select participants for enrollment and data analysis requires specific programming commands to provide valid statistical estimates. Also, each NHANES cycle includes over 50 data files, each of which has detailed informational material. NHANES data collection protocols have changed over time for certain variables, and these differences should be accounted for when analyzing data. To address the barriers to analyzing NHANES data, we developed an open-access web-based application that allows users to obtain BP and hypertension statistics for US adults using NHANES data. The goal of the application is to increase the usage of NHANES data for BP and hypertension research and policy decision making with a focus on ensuring results are valid and transparent. In this manuscript, we review the design of the application, show some results of NHANES analyses using the application, and provide a tutorial on how people can use the application.

**METHODS**

The open-access web-based application uses data from 10 NHANES cycles, specifically NHANES 1999-2000 through NHANES 2017- March 2020, prior to the COVID-19 pandemic. In each NHANES cycle, potential participants were identified using a multi-stage sampling process. The protocols for each NHANES cycle were approved by the NCHS Institutional Review Board. Written informed consent was obtained from each participant.

NHANES data were collected through in-home interviews and study examinations conducted at mobile examination centers. Data elements included in the application, along with their definitions, are listed in Supplemental Table 1. Information on age, sex, race/ethnicity, pregnancy status for women, cigarette smoking, a prior diagnosis of diabetes, use of insulin or oral glucose-lowering medication, a prior diagnosis of hypertension, self-reported antihypertensive medication use, and history of coronary heart disease, myocardial infarction, stroke, and heart failure were obtained by self-report during the in-home interview. During the in-home interview, the labels of medications that participants reported taking in the preceding 30 days were recorded. For the application, we categorized antihypertensive medications by class as defined in the 2017 ACC/AHA BP guideline. Data on medication dosages are not available in NHANES.

During the study examination, height and weight were measured and used to calculate body mass index. Systolic and diastolic BP (SBP and DBP, respectively) were measured three times, and blood and spot urine samples were collected. BP was measured using a mercury sphygmomanometer from 1999-2000 through 2015-2016 and using an oscillometric device in 2017-2020. The oscillometric BP values were calibrated to the mercury device as described previously. Blood was used to measure total and high-density lipoprotein (HDL) cholesterol, glycated hemoglobin and serum creatinine. Spot urine was used to measure albumin and creatinine and to conduct a pregnancy test.

Overall, 107,622 US individuals participated in NHANES 1999-2000 to 2017- March 2020. As statistics for hypertension and SBP and DBP levels in children and adolescents are markedly different than for adults, we restricted the data in the application to adults ≥ 18 years of age. We further restricted the application to participants who completed the in-home interview and study examination, with one or more SBP and DBP measurement and who had data on self-reported antihypertensive medication use. After these exclusions were applied, the application included data from 56,035 participants (Figure S1).

*Statistical analysis*

The application was created using Shiny, an open-source software package for creating web applications using the R programming language. Shiny translates R programming code into HTML, CSS, or JavaScript commands and creates a website interface. The statistical analyses that are used on the application are conducted with R programming language packages that account for the non-random selection of NHANES participants. All statistical analyses on the application are weighted to provide estimates for the non-institutionalized US population. Following CDC recommendations, the application automatically suppresses statistics which are unreliable due to being based on a small sample size.

The application can be used to estimate characteristics of the population which are relevant when investigating BP and hypertension data (Table 1, top panel). The application also allows users to estimate statistics for a number of hypertension-related variables (Table 1, bottom panel).

The application was designed with a number of options so that analyses can be customized to the user’s specifications. Results can be presented as figures or in tables. Figures can be made with the data represented as bars or points. The 10 NHANES cycles can be analyzed separately or multiple cycles can be pooled together, which provides a larger sample size leading to more stable statistical estimates. The analyses can be done crude or with age-adjustment using direct standardization in four age groups (18-44, 45-64, 65-74 and ≥75 years of age). Users can select the age distribution to be used as the standard population. Analyses can be done in sub-groups and stratified by participant characteristics.

To demonstrate the features and utility of the application, we conducted a series of statistical analyses. We estimated the crude and age-adjusted mean SBP and prevalence of hypertension for US adults by NHANES cycle from 1999-2000 through 2017-2020. We estimated the age-adjusted prevalence of hypertension by NHANES cycle for individuals with and without chronic kidney disease, separately. For age adjustment in these analyses, we used the estimated age distribution of US adults from 1999 to 2020 as the standard (49.3%, 33.6%, 10.1% and 7.0% being 18 to 44, 45 to 64, 65 to 74 and ≥75 years of age, respectively). Also, we estimated the absolute number of US adults with hypertension from 1999-2000 through 2017-2020. For the estimation of absolute counts of US adults, the application recalibrates the statistical weights to account for participants being excluded due to missing information on SBP, DBP or self-reported antihypertensive medication use. Weights are recalibrated within age, sex, race/ethnicity groups and NHANES cycle.

The age-adjusted distribution of the number of antihypertensive medication classes being taken was estimated for US adults with hypertension in each NHANES cycle, overall and among those who self-reported taking antihypertensive medication and had at least one antihypertensive drug class identified in their medication inventory. The age-adjusted proportion of US adults with controlled BP, defined using the 2017 ACC/AHA BP guideline, was estimated by NHANES cycle among US adults with hypertension and who self-reported taking antihypertensive medication. The age-adjusted prevalence of resistant hypertension was estimated by NHANES cycle for US adults with hypertension taking antihypertensive medication and among those with hypertension taking ≥ 3 classes of antihypertensive medication. For the analyses of number of antihypertensive medication classes being taken, resistant hypertension and BP control, we set the age distribution for the standard population to represent US adults with hypertension from 1999 to 2020: 26.4%, 43.4%, 17.0% and 13.2% being 18 to 44 years, 45 to 64, 65 to 74, and ≥75 years of age, respectively. Using the stratification option in the application, we estimated the prevalence of hypertension in NHANES 2017-2020 by age group, sex, and race-ethnicity. Using the data pooling option in the application, we grouped the NHANES 2011-2020 cycles to estimate the proportion of pregnant women who had uncontrolled BP according to the 2017 ACC/AHA BP guideline, SBP ≥ 130 mm Hg or DBP ≥ 80 mm Hg.

**RESULTS**

The estimated crude mean SBP for all US adults was 123 mm Hg in 1999-2000, 120 mm Hg in 2009-2010, and 123 mm Hg in 2017-2020 (Figure 1; panel A). The estimated age-adjusted mean SBP was 124, 120, and 122 mm Hg in 1999-2000, 2009-2010 and 2017-2020, respectively (Figure 1; panel B). The estimated crude prevalence of hypertension among US adults was lowest in 2009-2010 (43.3%) and highest in 2017-2020 (46.6%) (Figure 2; panel A). After age-adjustment, the prevalence of hypertension was highest in 1999-2000 (47.9%), lowest in 2009-2010 and 2013-2014 (43.0%) and was 44.7% in 2017-2020. (Figure 2; panel B). The number of US adults with hypertension increased from 89.8 million in 1999-2000 to 115.3 million US adults in 2017-2020 (Figure 3). The age adjusted mean SBP was 140 in 1999-2000, 134 in 2009-2010, and 136 in 2017-2020 among US adults with hypertension and 117 in 1999-2000, 115 in 2009-2010, and 117 in 2017-2020 among US adults without hypertension (Figure S2). In each NHANES cycle, the age-adjusted estimated mean SBP was 18 to 23 mm Hg higher among US adults with versus without hypertension. In 2017-2020, a higher percentage of US adults with versus without hypertension were non-Hispanic Black (13.5% versus 9.3%) while a lower percentage of US adults with versus without hypertension were Hispanic (12.3% versus 18.1%) (Table 1). In each NHANES cycle, the estimated age-adjusted prevalence of hypertension among US adults with CKD was higher versus their counterparts without CKD (Figure S3).

Among US adults with hypertension, 61.5% were not taking antihypertensive medication in 1999-2020 compared with 44.6% in 2017-2020. (Figure 4; panel A). Among US adults who self-reported taking antihypertensive medication, the percentage taking four or more classes of antihypertensive medication increased from 3.5% in 1999-2000 to 7.6% in 2017-2020 (Figure 4; panel B). The age-adjusted estimated prevalence of BP control among US adults with hypertension was 10.2%, 25.8%, and 23.8% in 1999-2000, 2009-2010, and 2017-2020, respectively (Figure 5; panel A). For adults who self-report taking antihypertensive medication, 26.2%, 44.1%, and 43.3% had controlled BP in 1999-2000, 2009-2010, and 2017-2020, respectively (Figure 5; panel B).

Among US adults with hypertension who self-reported taking antihypertensive medication, the prevalence of resistant hypertension was 13.7% in 1999-2000, 17.7% in 2009-2010, and 16.2% in 2017-2020 (Figure S4; panel A). Among those who were taking three or more classes of antihypertensive medication, the prevalence of resistant hypertension was 73.5%, 64.1%, and 67.9% in 1999-2000, 2009-2010, and 2017-2020, respectively (Figure S4; panel B).

In 2017-2020, the prevalence of hypertension was higher at progressively older age (Figure S5), among men versus women (Figure S6), and among non-Hispanic Black adults versus non-Hispanic White adults, non-Hispanic Asian adults, Hispanic adults and adults from other race/ethnic groups (Figure S7).

**DISCUSSION**

The web application has a number of complementary features that allow users to complete various analysis tasks. In particular, users can generate descriptive summaries of the US population spanning over 20 years, including summaries of characteristics and cardiometabolic outcomes. The flexibility of inclusion criteria paired with the large set of outcomes and stratifying variables available allow an enormous number of hypotheses to be explored. Outputs from the app can be included in scientific proposals, presentations, or articles, citing the application’s publicly available documentation to contextualize results. With the ability to download summary data from the application, users can visualize or tabulate results with aesthetics and software of their choosing. The web application can also be used for educational purposes, as teachers can use it to analyze health outcomes for US adults interactively with students.

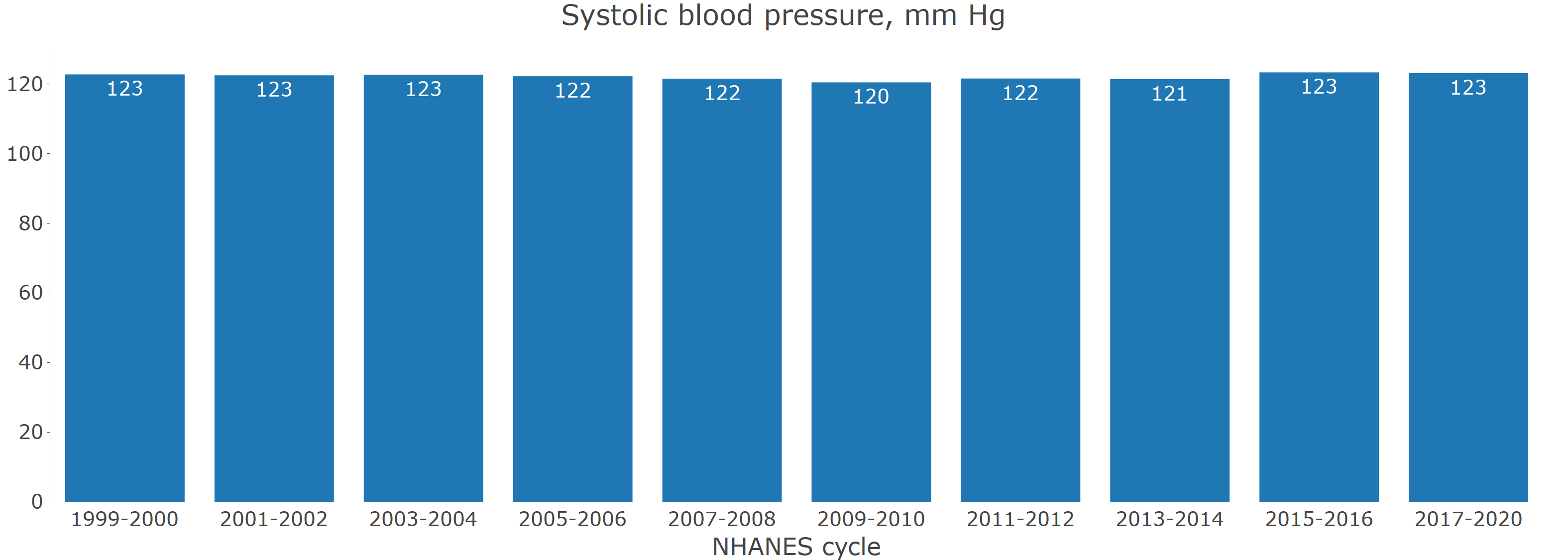
With an open-source software license and publicly available documentation and data, the web application we developed is designed to be extended. The Shiny interface and R codes supporting this application can support general sub-populations of NHANES participants, such as a sub-population of participants with cholesterol data or a sub-population of pediatric survey participants. As the current analysis demonstrates creation and analysis of a sub-population of participants with BP measurements, others can use it as a template for developing extensions of the web application that engage with new sub-populations. In addition to including new sub-populations of NHANES participants, the application may be extended to include inferential statistics and hypothesis testing, and may also incorporate new cycles of NHANES data as they become available.

**Table 1**: Race/ethnicity distribution of the population with and without hypertension in 2017-2020

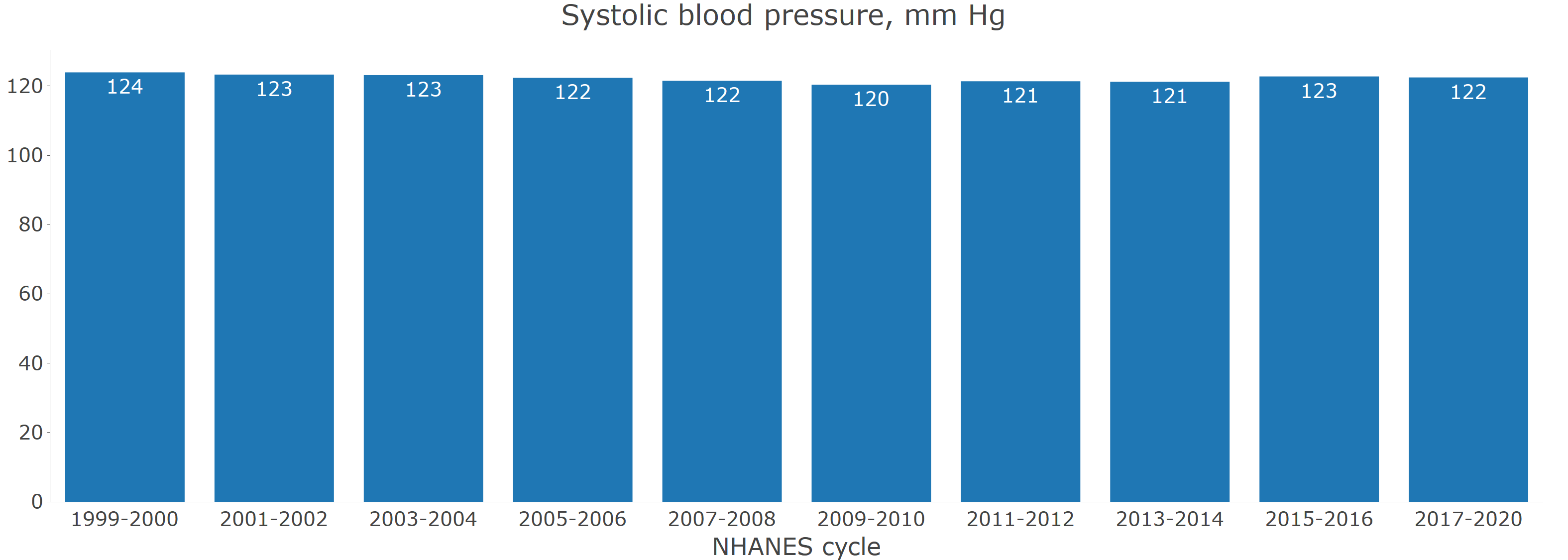
| **svy\_year** | **htn\_accaha** | **demo\_race** | **statistic** | **estimate** | **std\_error** | **ci\_lower** | **ci\_upper** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 2017-2020 | No | Hispanic | percentage | 18.1 | 1.84 | 14.5 | 21.7 |
| 2017-2020 | No | Non-Hispanic Asian | percentage | 5.90 | 0.88 | 4.18 | 7.62 |
| 2017-2020 | No | Non-Hispanic Black | percentage | 9.26 | 1.15 | 7.00 | 11.5 |
| 2017-2020 | No | Non-Hispanic White | percentage | 62.8 | 2.35 | 58.2 | 67.4 |
| 2017-2020 | No | Other | percentage | 3.96 | 0.38 | 3.21 | 4.71 |
| 2017-2020 | Yes | Hispanic | percentage | 12.3 | 1.21 | 9.95 | 14.7 |
| 2017-2020 | Yes | Non-Hispanic Asian | percentage | 5.06 | 0.79 | 3.50 | 6.61 |
| 2017-2020 | Yes | Non-Hispanic Black | percentage | 13.5 | 1.84 | 9.88 | 17.1 |
| 2017-2020 | Yes | Non-Hispanic White | percentage | 64.8 | 2.80 | 59.4 | 70.3 |
| 2017-2020 | Yes | Other | percentage | 4.30 | 0.58 | 3.17 | 5.43 |

**Figure 1** Mean systolic blood pressure for US adults by calendar year

A. Without age adjustment

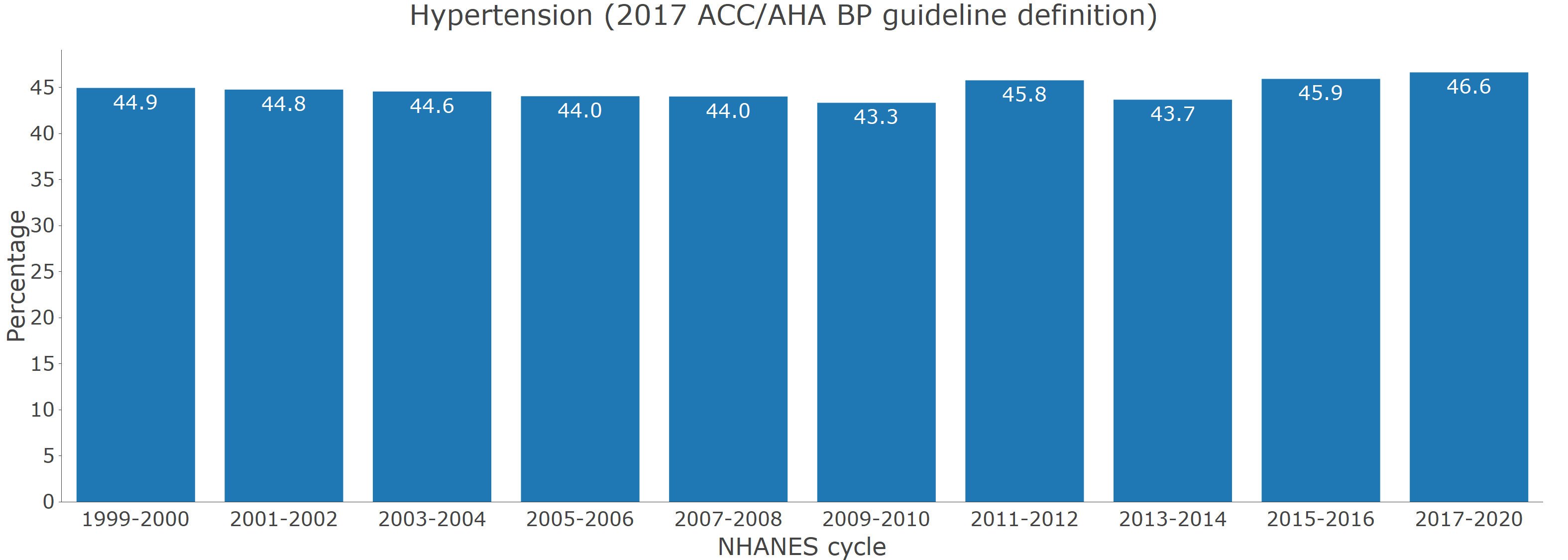


B. With age adjustment through direct standardization.

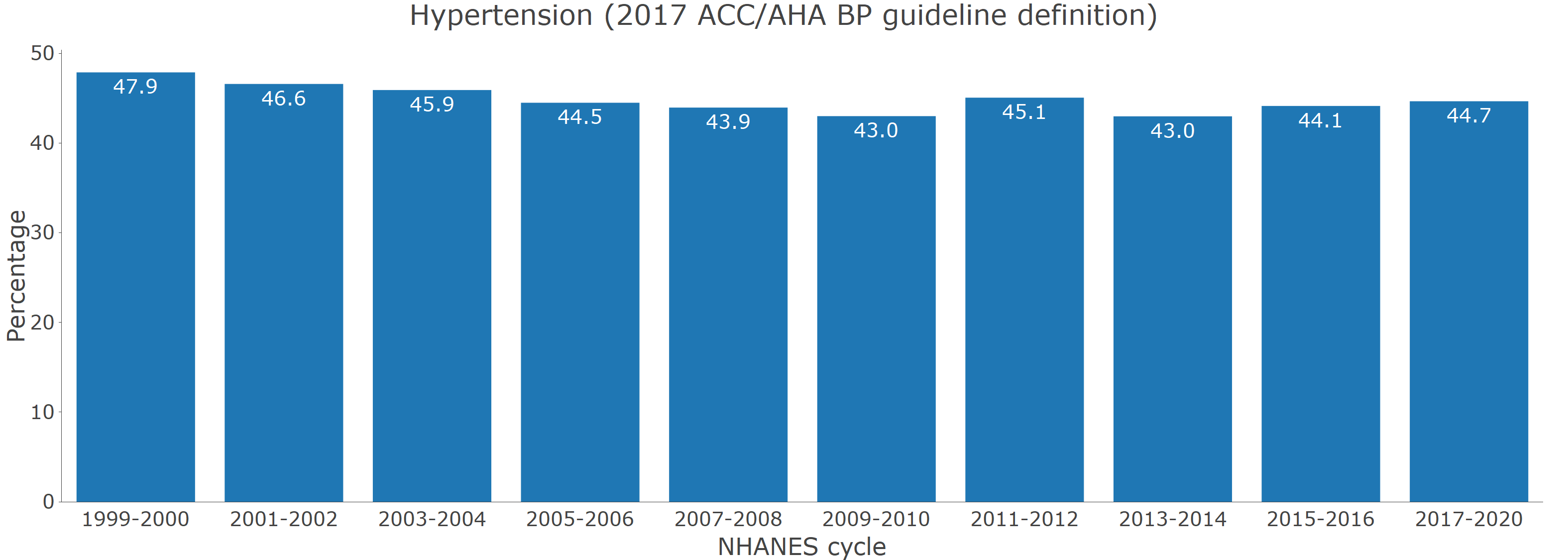


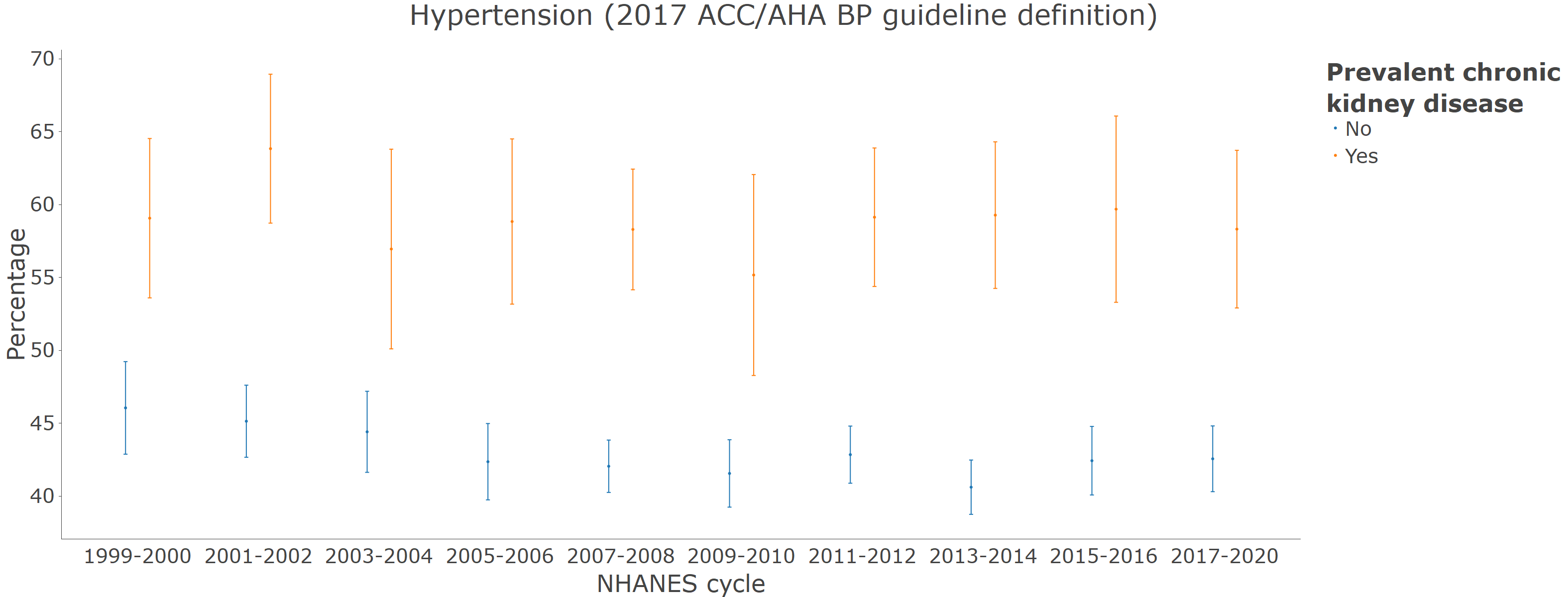
**Figure 2** Prevalence of hypertension for US adults by calendar year

A. Without age adjustment

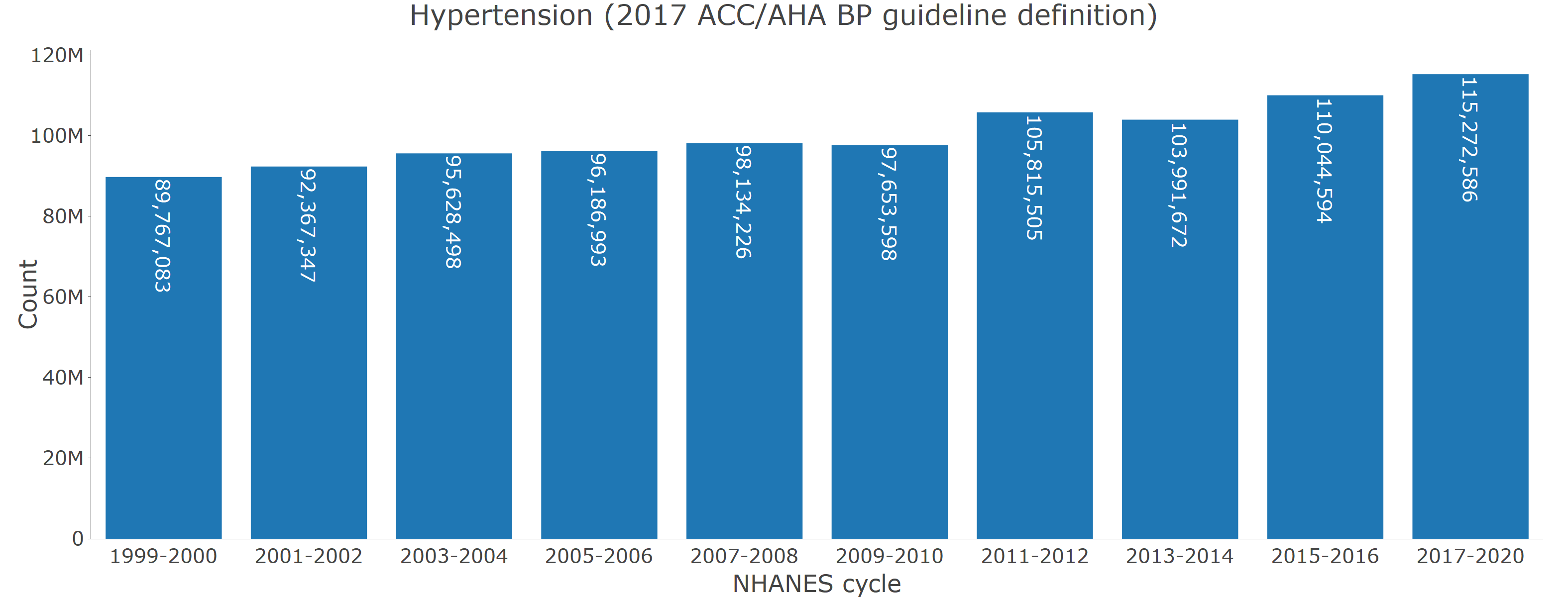


B. With age adjustment through direct standardization.



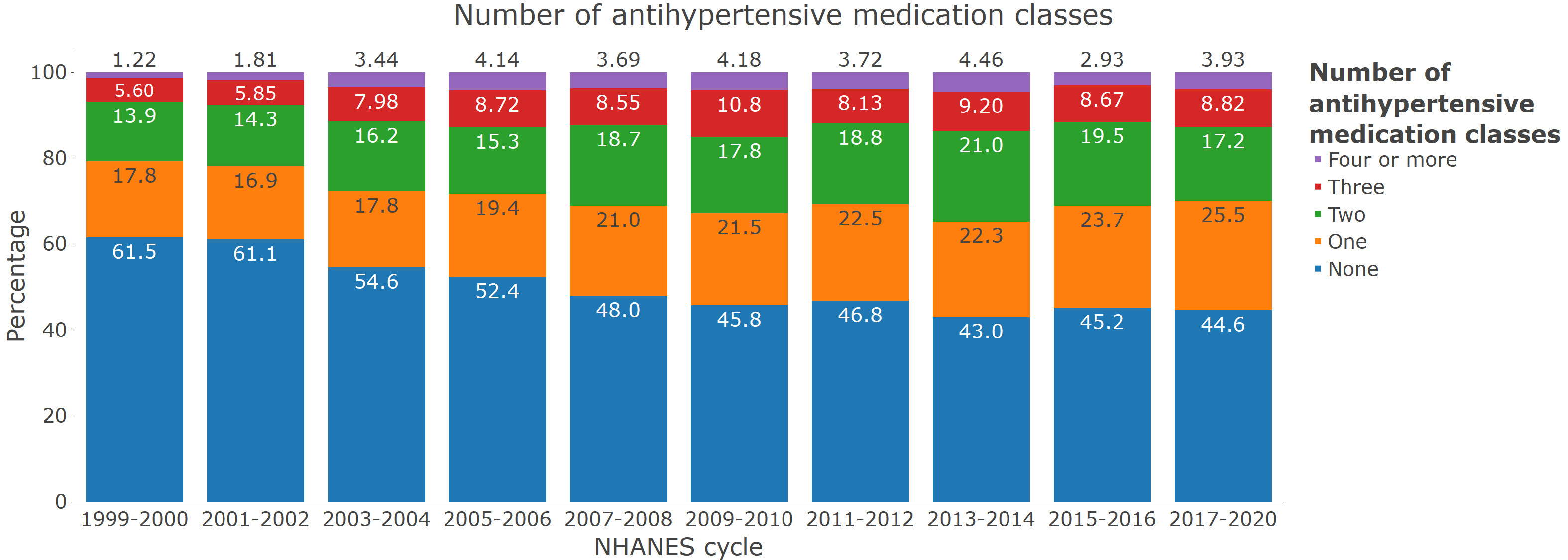


**Figure 3** Number of US adults with hypertension

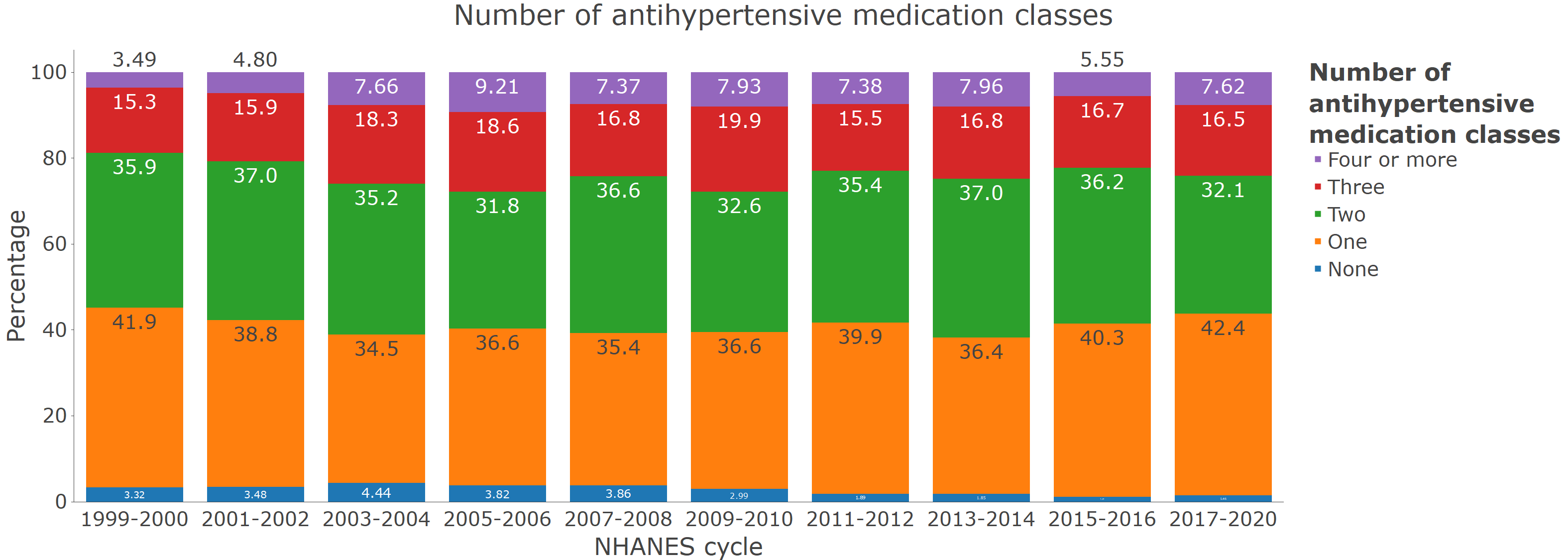


**Figure 4** Distribution of the number of antihypertensive medication classes being taken among those with hypertension by calendar period

A. Among those with hypertension

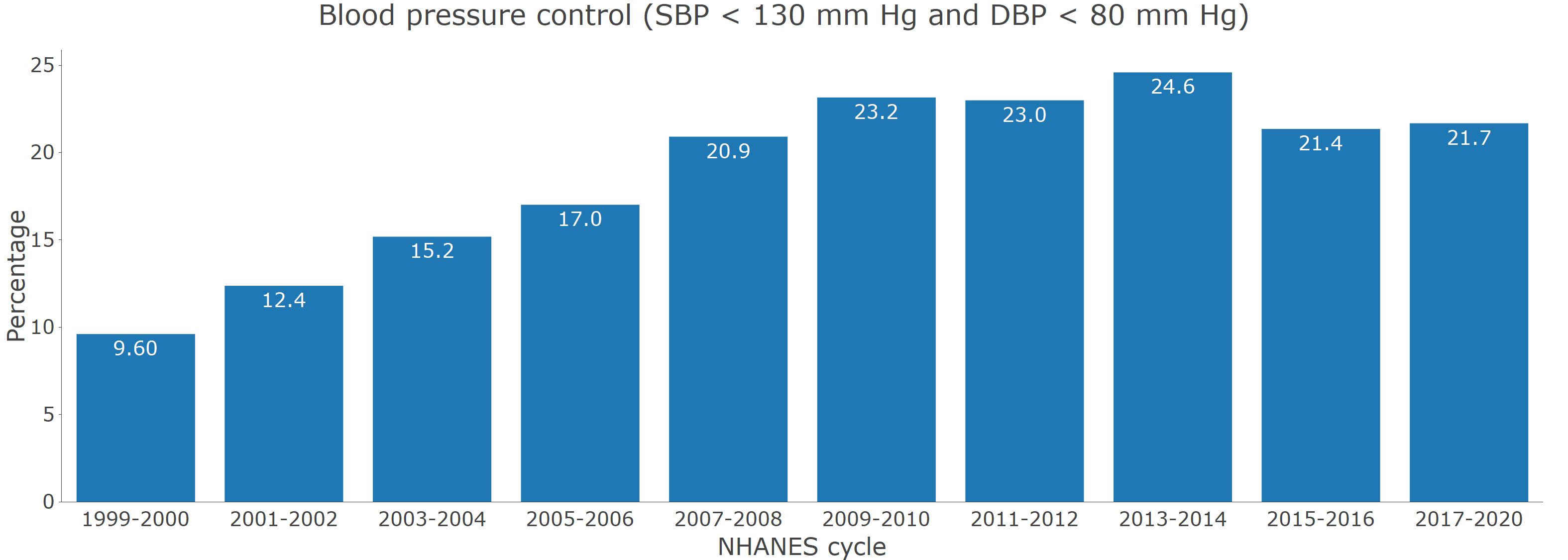


B. Among those with hypertension who self-report taking antihypertensive medication

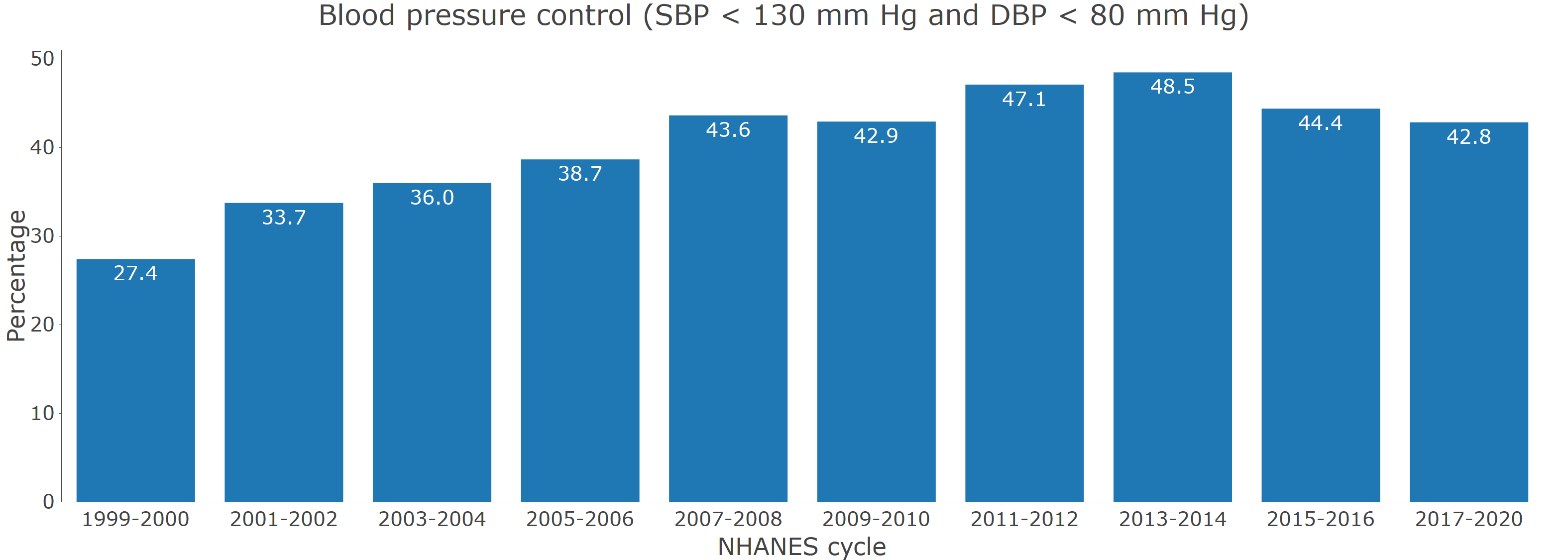


**Figure 5** Blood pressure control by calendar year

A. Age-adjusted among those with hypertension



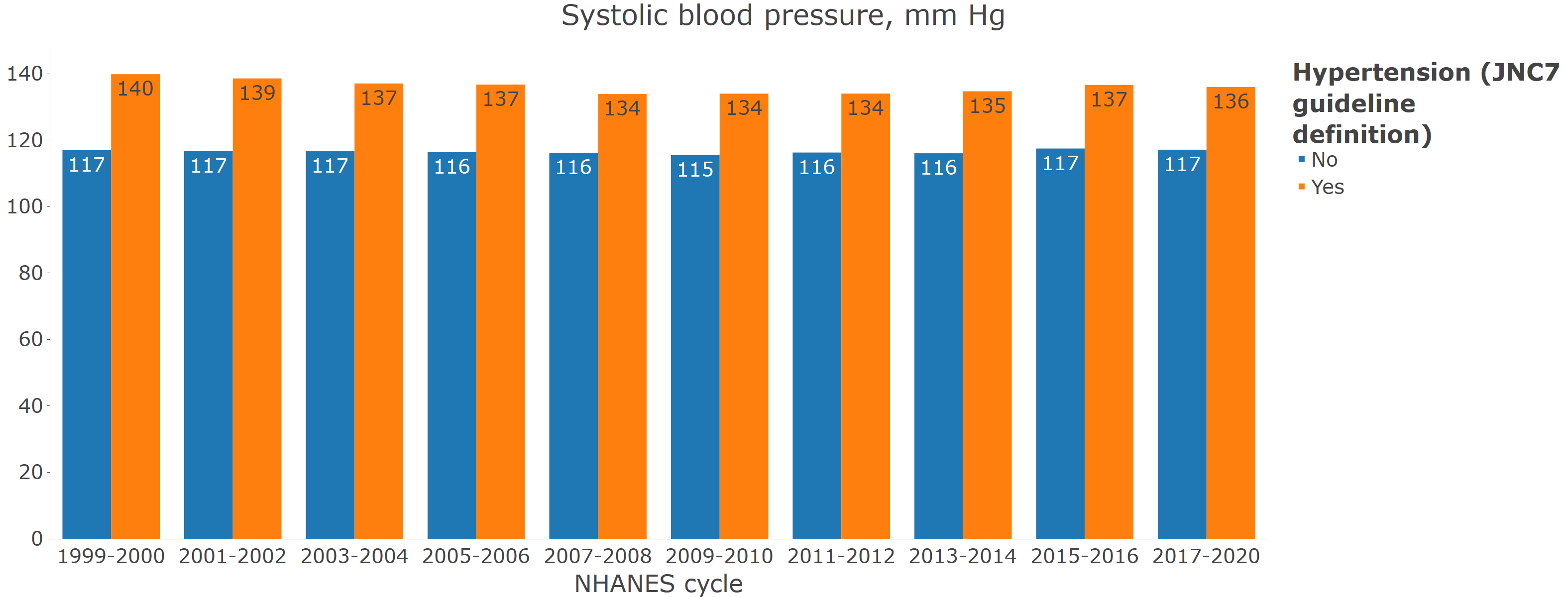
B. Age-adjusted among those who self-report taking antihypertensive medication



**Figure S1**

|  | **Overall** | **1999-2000** | **2001-2002** | **2003-2004** | **2005-2006** | **2007-2008** | **2009-2010** | **2011-2012** | **2013-2014** | **2015-2016** | **2017-2020** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Participants | 107,622 | 9,965 | 11,039 | 10,122 | 10,348 | 10,149 | 10,537 | 9,756 | 10,175 | 9,971 | 15,560 |
| ≥ 18 years old | 63,041 | 5,448 | 5,993 | 5,620 | 5,563 | 6,228 | 6,527 | 5,864 | 6,113 | 5,992 | 9,693 |
| Completed interview and examination | 59,799 | 4,976 | 5,592 | 5,303 | 5,334 | 5,995 | 6,360 | 5,615 | 5,924 | 5,735 | 8,965 |
| Had SBP and DBP measurements | 56,286 | 4,755 | 5,251 | 4,902 | 5,028 | 5,670 | 6,053 | 5,436 | 5,700 | 5,557 | 8,024 |
| Had self-reported information on antihypertensive medication | 56,035 | 4,694 | 5,184 | 4,838 | 5,015 | 5,665 | 6,043 | 5,337 | 5,694 | 5,552 | 8,013 |
| Final Sample size | 56,035 | 4,694 | 5,184 | 4,838 | 5,015 | 5,665 | 6,043 | 5,337 | 5,694 | 5,552 | 8,013 |

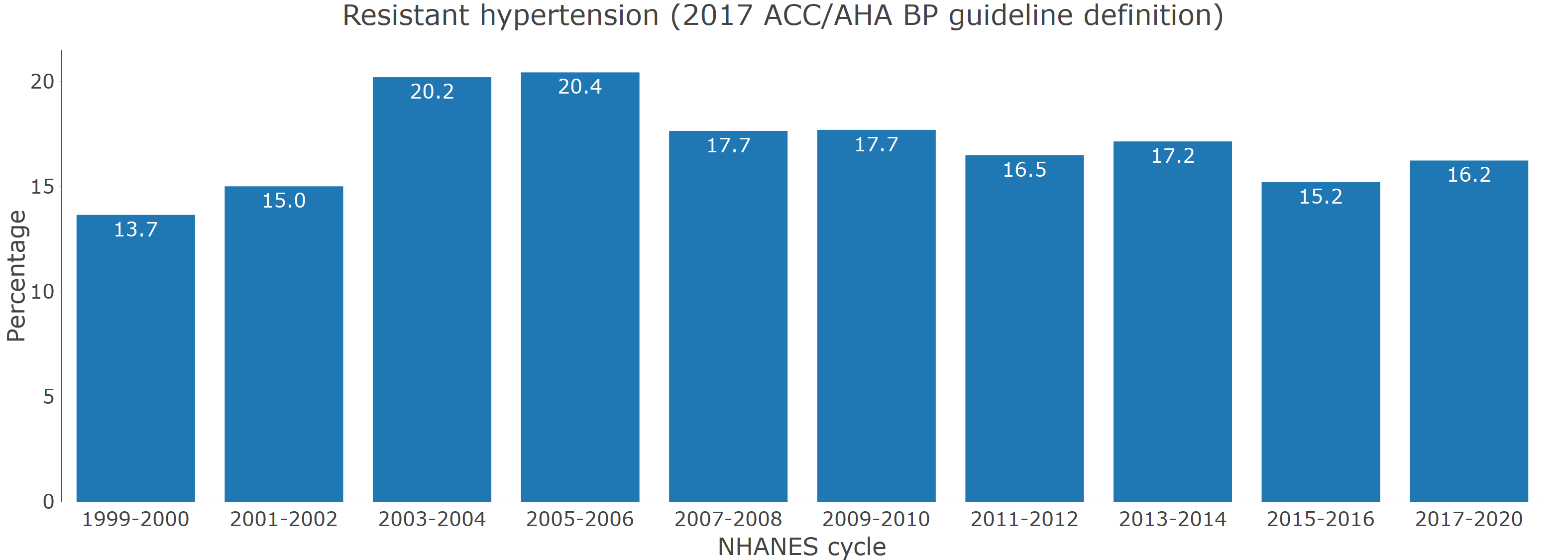
**Figure S2**: Age-adjusted mean systolic blood pressure among people with and without hypertension



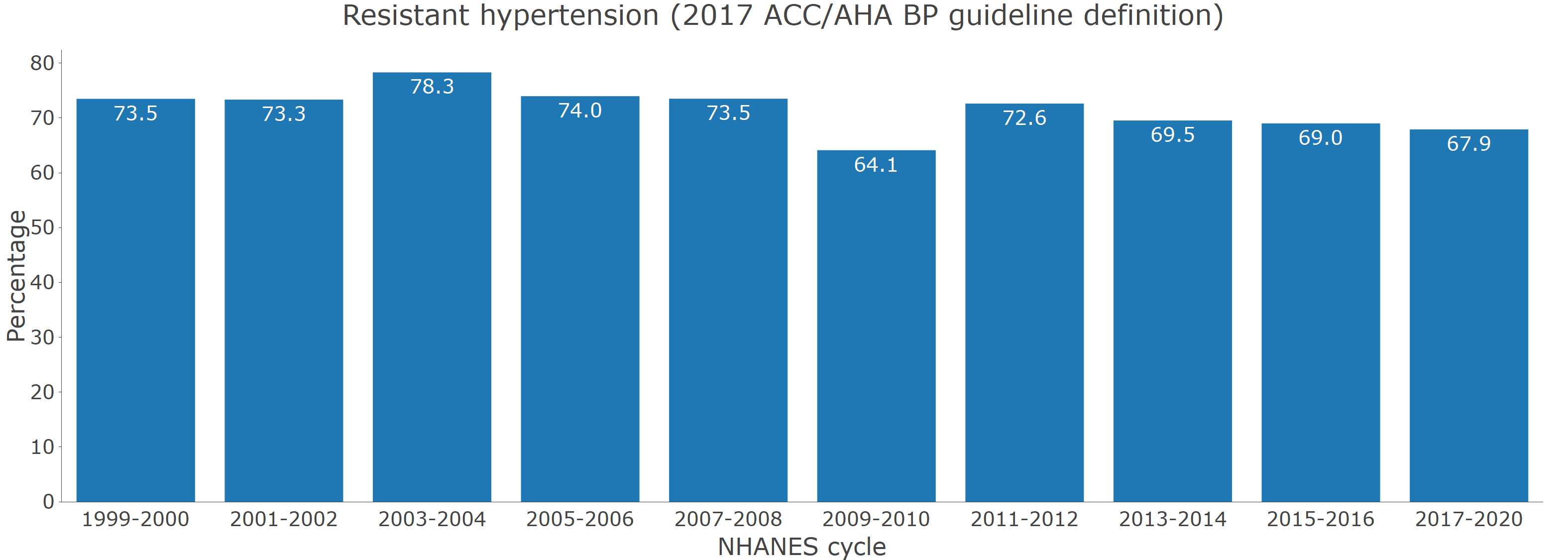
**Figure S3**: Age-adjusted prevalence of hypertension among US adults with and without chronic kidney disease

**Figure S4**: Prevalence of resistant hypertension by calendar year

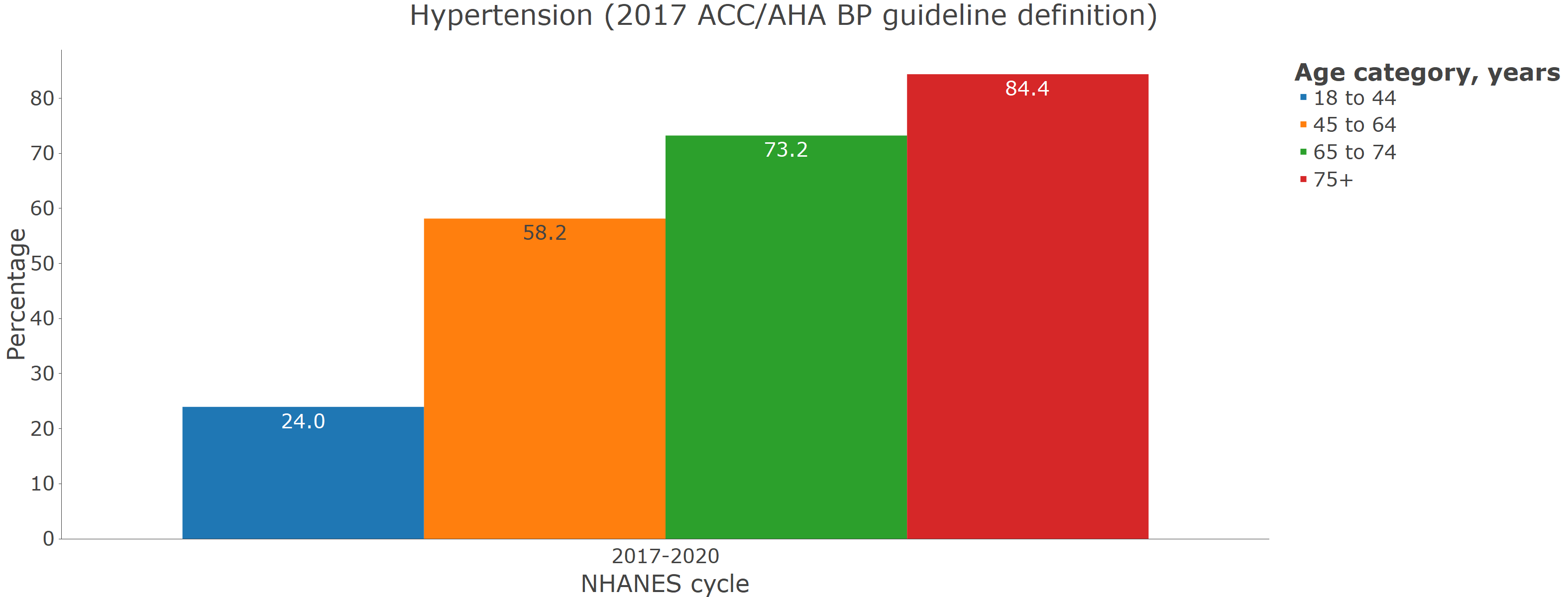
A. Among those with hypertension who self report taking antihypertensive medication



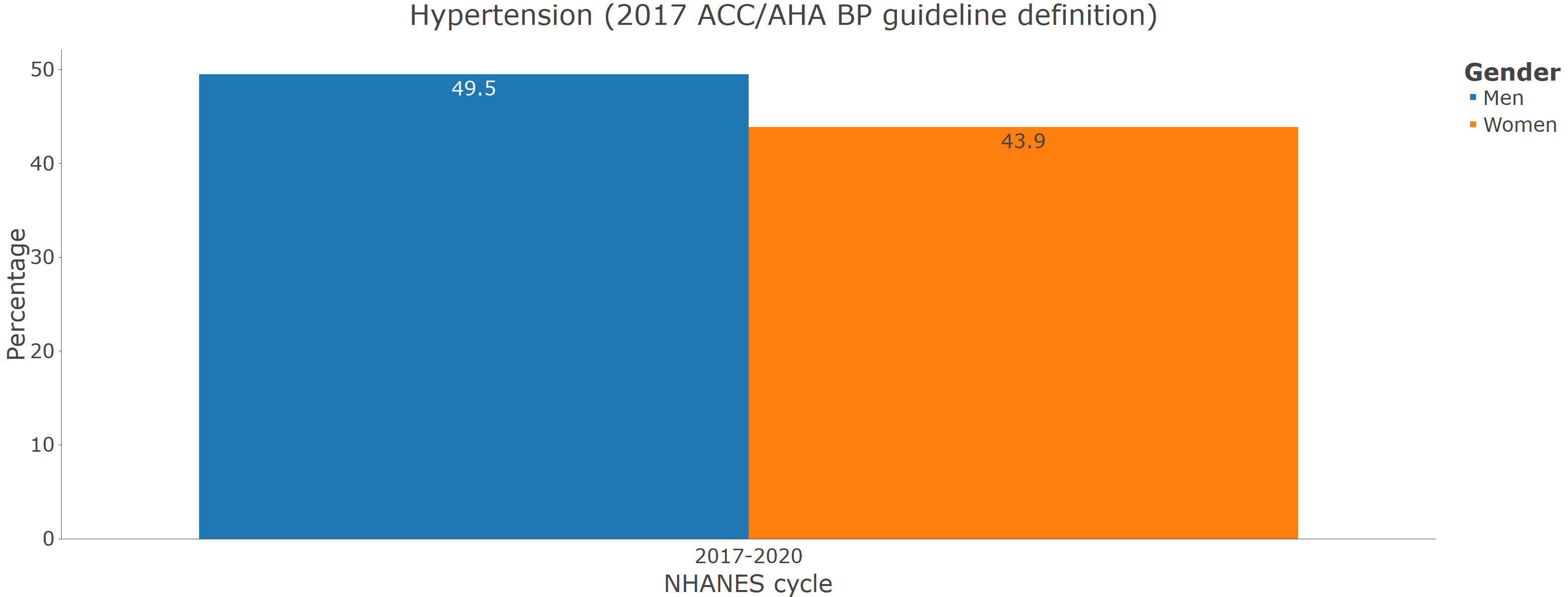
B. Among those with hypertension who self report and are taking 3+ classes of antihypertensive medication



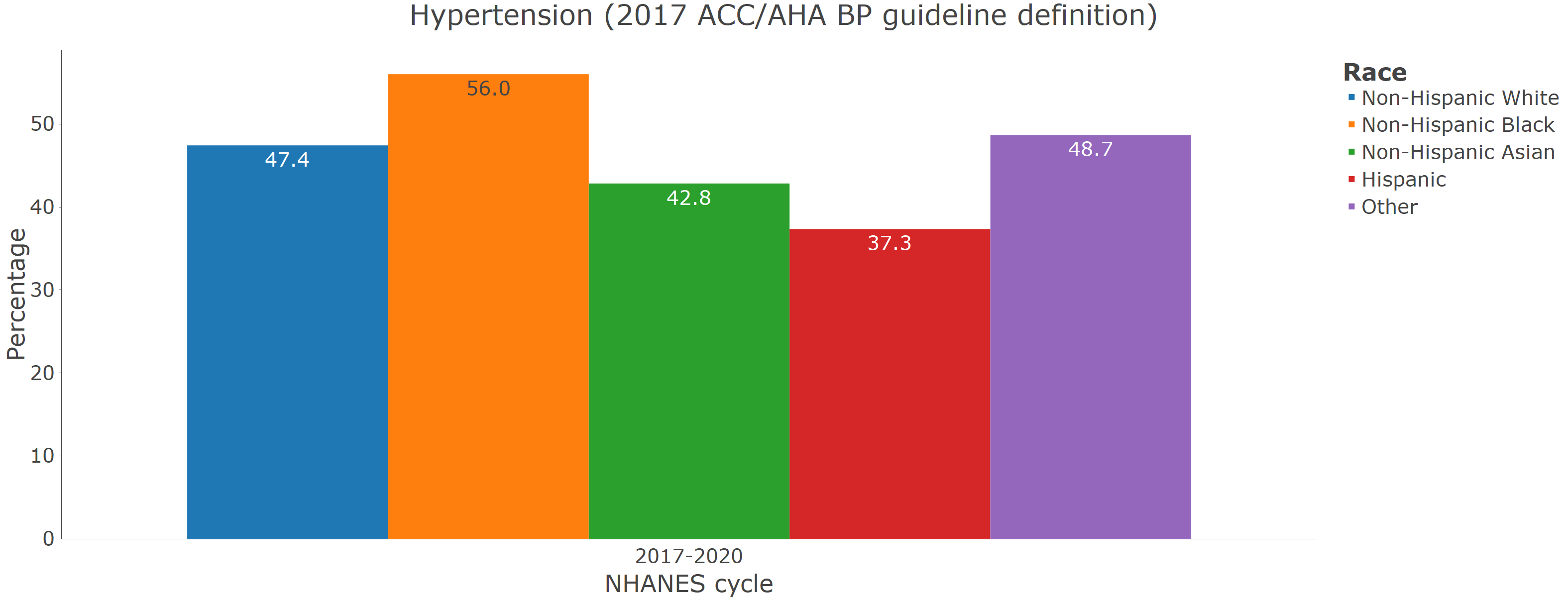
**Figure S5** Prevalence of hypertension by age categories in 2017-2020



**Figure S6** Prevalence of hypertension by sex in 2017-2020



**Figure S7** Prevalence of hypertension by race/ethnicity in 2017-2020



**Figure S8** Prevalence of uncontrolled blood pressure among pregnant women from 2011-2012 through 2017-2020.

