HHS Public Access

Author manuscript

Am J Hypertens. Author manuscript; available in PMC 2024 February 27.

Published in final edited form as:

Am J Hypertens. 2023 November 15; 36(12): 677–685. doi:10.1093/ajh/hpad081.

Leveraging Electronic Health Records to Construct a Phenotype for Hypertension Surveillance in the United States

Siran He[1,*], **Soyoun Park**[1], **Elena Kuklina**[1], **Nicole L. Therrien**[1], **Elizabeth A. Lundeen**[2], **Hilary K. Wall**[1], **Katrice Lampley**[1,3], **Lyudmyla Kompaniyets**[4], **Samantha L. Pierce**[4], **Laurence Sperling**[1], **Sandra L. Jackson**[1]

[1]Division for Heart Disease and Stroke Prevention, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA

[2]Division of Diabetes Translation, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA

[3]ASRT, INC, Smyrna, GA, USA

[4]Division of Nutrition, Physical Activity, and Obesity, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA

Abstract

BACKGROUND—Hypertension is an important risk factor for cardiovascular diseases. Electronic health records (EHRs) may augment chronic disease surveillance. We aimed to develop an electronic phenotype (e-phenotype) for hypertension surveillance.

METHODS—We included 11,031,368 eligible adults from the 2019 IQVIA Ambulatory Electronic Medical Records-US (AEMR-US) dataset. We identified hypertension using three criteria, alone or in combination: diagnosis codes, blood pressure (BP) measurements, and antihypertensive medications. We compared AEMR-US estimates of hypertension prevalence and control against those from the National Health and Nutrition Examination Survey (NHANES) 2017–18, which defined hypertension as BP ≥130/80 mm Hg or ≥1 antihypertensive medication.

RESULTS—The study population had a mean (SD) age of 52.3 (6.7) years, and 56.7% were women. The selected three-criteria e-phenotype (≥1 diagnosis code, ≥2 BP measurements of ≥130/80 mm Hg, or ≥1 antihypertensive medication) yielded similar trends in hypertension prevalence as NHANES: 42.2% (AEMR-US) vs. 44.9% (NHANES) overall, 39.0% vs. 38.7% among women, and 46.5% vs. 50.9% among men. The pattern of age-related increase in hypertension prevalence was similar between AEMR-US and NHANES. The prevalence of hypertension control in AEMR-US was 31.5% using the three-criteria e-phenotype, which was higher than NHANES (14.5%).

*Corresponding author: Siran He (SiranHe@cdc.gov).

Disclosure

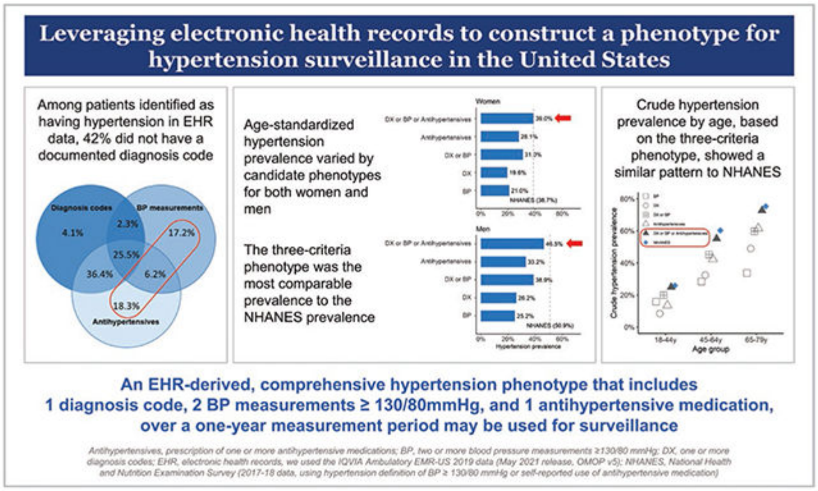
The authors have no conflict of interest to declare.

Supplementary Data

Supplementary materials are available at *American Journal of Hypertension* (http://ajh.oxfordjournals.org).

CONCLUSIONS—Using an EHR dataset of 11 million adults, we constructed a hypertension e-phenotype using three criteria, which can be used for surveillance of hypertension prevalence and control.

Graphical Abstract



Keywords

blood pressure; chronic disease; electronic health record; hypertension; phenotype; surveillance

---

Nearly half of American adults have hypertension, which contributed primarily or as a factor to 516,955 deaths in 2019 in the United States.[1,2] Improving hypertension prevention and control has the potential to avert substantial morbidity and mortality.[3] For example, a 5 mm Hg reduction in systolic blood pressure (BP) is associated with a 10% lower risk of major cardiovascular events.[4] To measure hypertension burden and evaluate progress in hypertension prevention and control, it is important to monitor hypertension trends at the population level.[5,6]

Hypertension surveillance can guide public health and clinical efforts to lower BP.[7] Traditional chronic disease surveillance systems are considered the gold standard, such as the National Health and Nutrition Examination Survey (NHANES).[8] However, these systems are resource-intensive, cross-sectional, and often have delays in results dissemination.[9] Electronic health record (EHR) data could be leveraged to augment public health surveillance.[9] EHR data are timely and rich in clinical detail and objective measures, which are ideal characteristics for surveillance.[7,10] EHR data may also help identify gaps in health services, as they allow for near-real-time analyses.[11]

Many decisions are required to define an EHR-based hypertension electronic phenotype (e-phenotype), which can yield different prevalence estimates. Such decisions include which phenotyping approach to use; which data elements to include; which BP cutoffs to apply; and the timeframe for analysis.[12–14] Despite promising findings at the state- and local-level using EHR phenotypes, such algorithms have yet to be implemented at the national level for surveillance purposes.[15,16] The overarching goal of this work, therefore, is to construct an EHR-based e-phenotype to augment hypertension surveillance in the United States. The three aims are: (i) to develop candidate e-phenotypes for hypertension; (ii) to guide the selection of an e-phenotype for hypertension surveillance based on NHANES estimates; and (iii) to report the prevalences of hypertension and control.

METHODS

Data and study populations

EHR data.—We used the IQVIA Ambulatory Electronic Medical Records-US (AEMR-US) dataset (May 2021 release), in the Observational Medical Outcomes Partnership (OMOP v5) format. The OMOP common data model serves to standardize the structure and content of observational data.[17] Data were extracted using the IQVIA E360 Platform. The AEMR-US data are collected from over 100,000 healthcare providers, with information from over 76 million patients who had face-to-face outpatient visits since 2006. Approximately 40% of the outpatient facilities were primary care, and the rest were specialists. All data are de-identified.

There were 18,185,788 unique patients in AEMR-US 2019 (not weighted to be representative of the US population). After removing 0.5% patients with “mis-bridged” records (i.e., discordant sexes, states, or years of birth) and restricting age to 18–79 years, 14,091,431 records were screened for exclusion (Figure 1). Consistent with the electronic clinical quality measure CMS165v10, we excluded patients with end-stage renal disease (0.2%), those with pregnancy-related events (9.3% of females aged 18–44 years), and those in hospice, palliative, or long-term care (0.1%).[18] We also excluded 20.0% patients without one BP (one systolic and one diastolic BP measured on the same day). Patients without BP measurements could dilute the denominator, e.g., visits to specialists instead of primary care (Supplementary Table S1).[19] The final analytical sample was 11,031,368.

Hypertension in AEMR-US is described in the “phenotyping hypertension” section, and cases identified through any of the candidate e-phenotypes are considered “suspected hypertension”. Among patients with suspected hypertension with at least one BP in 2019, hypertension control was based on the most recent BP measurement.[18] We examined control at both BP <130/80 mm Hg and <140/90 mm Hg cutoffs.

NHANES.—NHANES is a survey designed by the National Center for Health Statistics, Centers for Disease Control and Prevention (CDC).[20] The NHANES uses a stratified, multi-stage, probability cluster sampling design to collect data from a representative sample of the US population. Detailed methods are published elsewhere.[20] We used data from the NHANES 2017–18 cycle, and used mobile examination center weights and design variables for the analysis.

We screened the records of 8,704 participants for eligibility. We sequentially excluded 3,553 participants aged <18 years or >79 years, 55 pregnant women, 247 without BP measurement, and retained 4,849 participants (weighted mean age 46.1 years; 50.6% women) (Supplementary Table S2).

In NHANES, hypertension was defined as BP ≥ the cutoffs (130/80 or 140/90 mm Hg) based on an average of up to three measurements using a standardized protocol, or self-reported antihypertensive medication use. Hypertension control was defined as BP <130/80 mm Hg or <140/90 mm Hg (using same-day average BP).

Phenotyping hypertension

We used three criteria to identify suspected hypertension: diagnosis codes, BP measurements, and prescription of antihypertensive medications.

Diagnosis codes.—We imported hypertension diagnosis codes in International Classification of Diseases, Clinical Modification (ICD-9-CM 401.x to 405.x, and ICD-10-CM I10 to I15) into the E360 platform. Because we used the AMER-US data in the OMOP format, all source ICD codes were mapped to OMOP standard codes to identify eligible patients. Each standard code had an associated standard name, and we incorporated key standard names into an enhanced search. Guided by clinician input, we manually cleaned the exported standard codes (Supplementary Table S3). We finalized 113 standard codes to identify hypertension based on diagnosis codes.

BP measurements.—A patient must have high BP measurements on two or more days to be classified as having hypertension. We defined high BP based on the 2017 American College of Cardiology/American Heart Association (ACC/AHA) cutoffs of ≥130/80 mm Hg.[21] We conducted additional analyses using the 2003 Joint Committee (JNC 7) cutoffs of ≥140/90 mm Hg.[22] Biologically implausible values were excluded: systolic BP <30 or >300 mm Hg, and diastolic BP <20 or >150 mm Hg.

Medications.—A list of antihypertensive medications was developed to align with the 2017 ACC/AHA guideline.[21] We imported names of the single agents into the E360 platform and mapped them to RxNorm codes to create standard code lists. We manually cleaned the lists to remove products not approved for hypertension (see Supplementary Methods and Table S4 for details). The final code list had 3599 RxNorm codes.

Candidate e-phenotypes.—We built candidate e-phenotypes using one or more of the criteria: ≥1 diagnosis code, high BP measurements on ≥2 days, ≥1 antihypertensive medication, or a combination thereof. The selection of an e-phenotype was guided by comparing hypertension prevalence estimates in AEMR-US to those in NHANES. Because the care-seeking patient population in AEMR-US likely differs from the NHANES population in important ways, we did not expect a perfect match. The purpose of comparing these two datasets was to assess how each e-phenotype compared against existing national estimates for hypertension.

Covariates

Covariates included age, sex, race, and ethnicity, body mass index (BMI, cleaned via the R package “growthcleanr” based on the Daymont *et al.*[23] algorithm), smoking status (current smokers), and diabetes status (adapted from the SUPREME-DM algorithm).[24] Details are in Supplementary Methods.

Statistical analysis

We performed descriptive analyses for selected e-phenotypes and covariates. We first investigated the overlaps of the suspected hypertension as identified by various criteria, which was based on an "AND" relationship among criteria. We then calculated hypertension prevalence for each candidate e-phenotype using an "OR" relationship among criteria. We compared crude and age-standardized hypertension and control prevalences in AMER-US against those in NHANES. Age-standardization was conducted using the direct method based on in the 2010 US census.

The *main* measurement period covered January 1 through December 31, 2019. The denominator included eligible patients in 2019, and the numerator included suspected hypertension in 2019. Because hypertension is a chronic condition, suspected cases may not be captured within a 1-year period. We conducted analysis for an *extended* period, which spanned from January 1, 2018 through December 31, 2019. The denominator was still eligible patients from 2019, but the numerator included suspected hypertension in 2018 or 2019 (Figure 1).

We conducted a set of sensitivity analyses using the subset of NHANES adult participants who had at least one healthcare visit in the past year. The overall NHANES population is representative of the US population, which may not be the case for the care seeking sub-population.

We used SAS (v9.4, Cary, NC, USA) to analyze AMER-US data; SAS SUDAAN (v9.4, Cary, NC, USA) for NHANES data; R (v4.1.2, Vienna, Austria) and Lucidchart (South Jordan, UT, USA) for data visualization. This activity was reviewed by CDC and conducted consistent with applicable federal law and CDC policy.[1]

Data availability

The data underlying this article are proprietary and cannot be shared publicly. However, standard codes for the e-phenotype will be shared upon request.

RESULTS

The mean (SD) age of the study population was 52.3 (16.7) years, and 56.7% were women (Table 1). Over 65% were White, 8% Black, and about 2% were Asian. Approximately 10.3% adults were current smokers, and 14.7% had diabetes. The mean (SD) BMI was 29.9 (7.1) kg/m^2, which is at the upper range of overweight for adults. The number of patients who were identified using single hypertension criterion varied between 1,103,481 (10.0%, 2 BP ≥140/90 mm Hg) to 4,257,341 (38.6%, antihypertensives). About 90% of adults with one or two hypertension diagnosis codes were prescribed at least 1 antihypertensive medication.

Of all patients with suspected hypertension, 25.5% were identified by all three criteria simultaneously in the main measurement period using BP ≥130/80 mm Hg (Table 2). Patients meeting two criteria were more common than those meeting all three: 61.9% had a diagnosis code and an antihypertensive medication, 31.7% had ≥2 high BP and an antihypertensive medication, and 27.8% had a diagnosis code and ≥2 high BP.

Approximately 41.7% of patients with suspected hypertension did not have a diagnosis code in the main study period. The proportion of patients identified by all three criteria increased in the extended period.

Crude hypertension prevalence increased with age (Figure 2). Using the three-criteria e-phenotype with BP ≥130/80 mm Hg, the prevalence steadily increased: 24.9%, 55.2%, and 72.7% for age groups 18–44, 45–64, and 65–79 years, respectively, which followed a similar pattern to those in NHANES (25.9%, 60.3%, and 75.3% for the corresponding groups) (Figure 2A–C; Supplementary Table S5). Patterns in hypertension prevalence identified through the three-criteria definition were the most similar to NHANES, despite slightly lower estimates in the 45–64 years age group.

Age-standardized hypertension prevalence was higher when using combinations of criteria than a single criterion, and the highest estimate was observed using the three-criteria e-phenotype with BP ≥130/80 mm Hg in AEMR-US (42.2%), which was similar to NHANES (44.9%) (Figure 3A). The prevalence for women was 39.0% in AEMR-US and 38.7% in NHANES (Figure 3B), and for men, 46.5% in AEMR-US and 50.9% in NHANES (Figure 3C). Hypertension prevalence was consistently higher when using an extended period compared to the main measurement period (Supplementary Table S6).

Patients in AEMR-US had higher age-standardized BP control prevalence (31.5%) than those in NHANES (14.5%) when using the BP <130/80 mm Hg cutoffs (Table 3). In AEMR-US, crude BP control estimates was the highest in 65- to 79-year-old patients (35.7%), which was slightly higher than NHANES participants in the same age group (30.8%).

In sensitivity analyses examining BP cutoffs of ≥140/90 mm Hg, compared to estimates using BP cutoffs of ≥130/80 mm Hg, hypertension prevalence was consistently lower across candidate e-phenotypes that contained BP measurement as a criterion, as expected (Figures 2D–F and 3D–F). In the sensitivity analysis using the subset of the NHANES population with healthcare visit (mean ± SD age 49.4 ± 7.4 years; 54.5% women), hypertension prevalence estimates were consistently higher in the care-seeking subgroup than in the original NHANES population (Supplementary Table S7).

DISCUSSION

In this EHR study of 11 million adults, we developed an e-phenotype for hypertension surveillance. We adopted an iterative, rule-based process to construct each component of the e-phenotype. Guided by NHANES estimates, we selected a three-criteria e-phenotype for hypertension: ≥1 diagnosis code, ≥2 BP measurements of 130/80 mm Hg or higher, or ≥1 antihypertensive medication within a 1-year measurement period.

Researchers in the UK conducted an equivalent analysis. They selected the combination of 1 diagnosis code and 2 high BP (≥140/90 mm Hg).[13] In our study, we added antihypertensive medications to the e-phenotype, which enhanced its performance when evaluated against NHANES across age and sex strata. The difference between our e-phenotypes could be due to different data sources, exclusion criteria, methodologies (e.g., age-standardization in our study), and vocabularies of medical codes. Despite these differences, both studies demonstrated the potential of leveraging EHR data for national hypertension surveillance. A US study using de-identified EHR data reported that the best hypertension algorithm without narrative fields was derived from normalized ICD codes, medications, and BP measurements, consistent with our three-criteria e-phenotype selection.[25]

Because we built an enhanced search for hypertension diagnosis codes, we retrieved more patients than merely importing ICD codes into AEMR-US. Nevertheless, among all patients with suspected hypertension, over 40% did not have a diagnosis code and may have been "hiding in plain sight" (in care but undiagnosed).[26] Banerjee and colleagues[27] reported that in the outpatient setting in the United States, hypertension diagnosis rates are suboptimal. In addition, some patients may have been diagnosed previously, but the code was not captured even within the extended period. It is also possible that the diagnosis was documented in the unstructured fields (e.g., clinical notes) but not in the structured elements, though this phenomenon may vary across facilities.[28,29]

Using our selected e-phenotype, we observed that only one third of patients in AEMR-US had controlled blood pressure (BP <130/80 mm Hg), which is suboptimal. However, this is a higher proportion than in NHANES, which may be explained by several factors. First, AEMR-US is a care-seeking population, whereas, NHANES includes participants in and outside of the healthcare system. This potential selection bias in AEMR-US may lead to the observation of better BP control. Second, we only examined BP control among patients with at least one BP in 2019 in AEMR-US. These patients may have had more clinical visits and better hypertension management than those without BP measurements. Third, the timepoints of BP measurements differed: In AEMR-US, we identified hypertension using high BP on two or more days, and determined control using the most recent BP; in NHANES, hypertension and control were both based on the same-day average of BP measurements. To be assessed as having hypertension in NHANES, patients necessarily either had uncontrolled BP or a medication; there was no opportunity for a patient to have had high BP and then to have had their BP controlled, due to the same-day assessment. Anyone with transient or chronic high BP could have been classified as both having hypertension and uncontrolled hypertension in NHANES.

We evaluated an EHR-based e-phenotype against a nationally representative survey; such comparisons must be interpreted with caution. Patients in an EHR dataset are a "care seeking" population, who differ from the general population in health status and healthcare access. Selection bias is a known issue in EHR-based research, and it is likely that we captured older, sicker, and wealthier patients than the national average.[30] In addition, each criterion of the hypertension definition in AMER-US and NHANES may not align exactly. The AEMR-US dataset has hypertension diagnoses codes, which were not available in NHANES. High BP was identified by different-day BP measurements in AEMR-US, vs. a same-day average in NHANES. In addition, the self-reported medications in NHANES were ascertained specifically for hypertension treatment, whereas in AEMR-US medications had lower specificity (i.e., may have been taken for non-hypertension purposes), and there may be a high prevalence of non-adherence to prescribed medications.[31] Because diagnosis codes were likely to be the most reliable component of a hypertension e-phenotype in EHR data, followed by BP measurement, we only examined the two-way candidate e-phenotype that included these two criteria. The exclusion criteria also differed between AEMR-US and NHANES in our analyses.

This study had several limitations due to the nature of EHR data, which are generated primarily for clinical billing and documentation. First, biases, data quality issues (e.g., the quality of BP measurements may vary across clinics), and missingness are general concerns when using EHR data.[10,30,32,33] The potential issue of low representativeness could be partly observed in the lack of information from Hispanic patients in AEMR-US. Given these potential biases and the high degree of missingness of race and ethnicity data, it is challenging to use EHR data to assess health disparities. Development of reliable and contextualized imputation and weighting methods could partially address this limitation. Second, AEMR-US data only have EHR-level information, and we do not know whether a medication was filled, picked up at a pharmacy, or taken by the patient. Third, assessing hypertension based on antihypertensive medications could also incur false positives and artificially inflate the apparent hypertension prevalence, as some patients without hypertension may be prescribed these medications for other conditions such as heart failure. Fourth, we could not strictly follow the CMS165v10 specifications, particularly in terms of frailty, for which we provided detailed explanation in the Supplementary Methods.

This work has notable strengths. To our knowledge, despite extensive efforts in EHR-based phenotyping in the clinical research setting, this is the first hypertension e-phenotype constructed specifically for public health surveillance at the national level in the United States.[15,16,25,34,35] Each criterion within the e-phenotype was developed based on an iterative review process by a team of clinicians, epidemiologists, health services researchers, and pharmacists. We also aligned the exclusion criteria with a widely used clinical quality measure.[18] There are also advantages inherent to EHR data, including a large patient population with high national coverage (albeit not nationally representative), a far-reaching and cross-cutting common data model that allows the e-phenotype to be applied to other OMOP databases, as well as opportunities for longitudinal analyses.[10,16]

Through this study, we demonstrated that a viable hypertension e-phenotype can be developed and applied to EHR data in the United States for surveillance purposes. This research shows that public health researchers and practitioners, health services researchers, and health care professionals can help identify, validate, and promote standard definitions for health conditions using EHR data. Despite the challenges associated with secondary use of EHR data for public health purposes, improving EHR data utilization and interpretation is immensely valuable as part of an effort to modernize chronic disease surveillance systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Ms. Nilka Rios Burrows for her help in identifying source codes for end-stage renal disease. Source of funding: none.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

REFERENCES

1. Hypertension Cascade: Hypertension Prevalence. Treatment and Control Estimates Among US Adults Aged 18 Years and Older Applying the Criteria From the American College of Cardiology and American Heart Association's 2017 Hypertension Guideline—NHANES 2015–2018. US Department of Health and Human Services; 2021, Centers for Disease Control and Prevention (CDC): Atlanta, GA, 2021.
2. Multiple Cause of Death, 1999–2020 Request. https://wonder.cdc.gov/mcd-icd10.html. Accessed 20 July 2022.
3. Farley TA, Dalal MA, Mostashari F, Frieden TR, Frieden TR. Deaths preventable in the U.S. by improvements in use of clinical preventive services. Am J Prev Med 2010; 38(6):600–609. [PubMed: 20494236]
4. Rahimi K, Bidel Z, Nazarzadeh M, Copland E, Canoy D, Ramakrishnan R, Pinho-Gomes A-C, Woodward M, Adler A, Agodoa L, Algra A, Asselbergs FW, Beckett NS, Berge E, Black H, Brouwers FPJ, Brown M, Bulpitt CJ, Byington RP, Cushman WC, Cutler J, Devereaux RB, Dwyer J, Estacio R, Fagard R, Fox K, Fukui T, Gupta AK, Holman RR, Imai Y, Ishii M, Julius S, Kanno Y, Kjeldsen SE, Kostis J, Kuramoto K, Lanke J, Lewis E, Lewis JB, Lievre M, Lindholm LH, Lueders S, MacMahon S, Mancia G, Matsuzaki M, Mehlum MH, Nissen S, Ogawa H, Ogihara T, Ohkubo T, Palmer CR, Patel A, Pfeffer MA, Pitt B, Poulter NR, Rakugi H, Reboldi G, Reid C, Remuzzi G, Ruggenenti P, Saruta T, Schrader J, Schrier R, Sever P, Sleight P, Staessen JA, Suzuki H, Thijs L, Ueshima K, Umemoto S, van Gilst WH, Verdecchia P, Wachtell K, Whelton P, Wing L, Yui Y, Yusuf S, Zanchetti A, Zhang Z-Y, Anderson C, Baigent C, Brenner BM, Collins R, de Zeeuw D, Lubsen J, Malacco E, Neal B, Perkovic V, Rodgers A, Rothwell P, Salimi-Khorshidi G, Sundström J, Turnbull F, Viberti G, Wang J, Chalmers J, Teo KK, Pepine CJ, Davis BR. Pharmacological blood pressure lowering for primary and secondary prevention of cardiovascular disease across different levels of blood pressure: an individual participant-level data meta-analysis. Lancet 2021; 397(10285):1625–1636. [PubMed: 33933205]
5. Washington, DC: U.S. Department of Health and Human Services, Office of the Surgeon General. The Surgeon General's Call to Action to Control Hypertension. U.S. Department of Health and Human Services: Washington, DC, 2020.
6. Institute of Medicine (US) Committee on Public Health Priorities to Reduce and Control Hypertension. A Population-Based Policy and Systems Change Approach to Prevent and Control Hypertension. Washington (DC): National Academies Press (US); 2010. Available from: https://www.ncbi.nlm.nih.gov/books/NBK220087/ doi:10.17226/12819.
7. Garies S, Youngson E, Soos B, Forst B, Duerksen K, Manca D, McBrien K, Drummond N, Quan H, Williamson T. Primary care EMR and administrative data linkage in Alberta, Canada: describing the suitability for hypertension surveillance. BMJ Health Care Inform 2020; 27(3):e100161. doi:10.1136/bmjhci-2020-100161
8. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention: Hyattsville, MD.
9. Heisey-Grove D, Wall HK, Helwig A, Wright JS; Centers for Disease Control and Prevention (CDC). Using electronic clinical quality measure reporting for public health surveillance. Morb Mortal Wkly Rep 2015; 64(16):439–442.
10. Williams BA, Voyce S, Sidney S, Roger VL, Plante TB, Larson S, LaMonte MJ, Labarthe DR, DeBarmore BM, Chang AR, Chamberlain AM, Benziger CP. Establishing a National Cardiovascular Disease Surveillance System in the United States using electronic health record data: key strengths and limitations. J Am Heart Assoc 2022; 11:e024409. [PubMed: 35411783]
11. Lu Y, Huang C, Mahajan S, Schulz WL, Nasir K, Spatz ES, Krumholz HM. Leveraging the electronic health records for population health: a case study of patients with markedly elevated blood pressure. J Am Heart Assoc 2020; 9(7):e015033. [PubMed: 32200730]

12. Lee S, Doktorchik C, Martin EA, D'Souza AG, Eastwood C, Shaheen AA, Naugler C, Lee J, Quan H. Electronic medical record-based case phenotyping for the Charlson Conditions: scoping review. JMIR Med Inform 2021; 9(2):e23934. [PubMed: 33522976]
13. Peng M, Chen G, Kaplan GG, Lix LM, Drummond N, Lucyk K, Garies S, Lowerison M, Weibe S, Quan H. Methods of defining hypertension in electronic medical records: validation against national survey data. J Public Health 2016; 38(3):e392–e399.
14. Zhang Y, Cai T, Yu S, Cho K, Hong C, Sun J, Huang J, Ho Y-L, Ananthakrishnan AN, Xia Z, Shaw SY, Gainer V, Castro V, Link N, Honerlaw J, Huang S, Gagnon D, Karlson EW, Plenge RM, Szolovits P, Savova G, Churchill S, O'Donnell C, Murphy SN, Gaziano JM, Kohane I, Cai T, Liao KP. High-throughput phenotyping with electronic medical record data using a common semi-supervised approach (PheCAP). Nat Protoc 2019; 14(12):3426–3444. [PubMed: 31748751]
15. Kim RS, Shankar V. Prevalence estimation by joint use of big data and health survey: a demonstration study using electronic health records in New York city. BMC Med Res Methodol 2020; 20:1–10.
16. Klompas M, Cocoros NM, Menchaca JT, Erani D, Hafer E, Herrick B, Josephson M, Lee M, Payne Weiss MD, Zambarano B, Eberhardt KR, Malenfant J, Nasuti L, Land T. State and local chronic disease surveillance using electronic health record systems. Am J Public Health 2017; 107(9):1406–1412. [PubMed: 28727539]
17. Data Standardization—OHDSI. https://www.ohdsi.org/data-standardization/. Accessed 19 January 2023.
18. Controlling High Blood Pressure. https://ecqi.healthit.gov/ecqm/ec/2022/cms165v10. Accessed 19 July 2022.
19. Billups SJ, Saseen JJ, Vande Griend JP, Schilling LM. Blood pressure control rates measured in specialty vs primary care practices within a large integrated health system. J Clin Hypertens 2018; 20(9):1253–1259.
20. CDC. National Health and Nutrition Examination Survey. https://www.cdc.gov/nchs/nhanes/index.htm. Accessed 7 November 2022.
21. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Himmelfarb CD, DePalma SM, Gidding S, Jamerson KA, Jones DW, MacLaughlin EJ, Muntner P, Ovbiagele B, Smith SC, Spencer CC, Stafford RS, Taler SJ, Thomas RJ, Williams KA, Williamson JD, Wright JT. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018; 71:e44–e45. doi:10.1161/hyp.0000000000000065
22. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL Jr, Jones DW, Materson BJ, Oparil S, Wright JT Jr, Roccella EJ, National Heart, Lung, and Blood Institute Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure; National High Blood Pressure Education Program Coordinating Committee. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. JAMA 2003; 289(19):2560–2572. doi:10.1001/jama.289.19.2560. Epub 2003 May 14. Erratum in: JAMA. 2003 Jul 9;290(2):197. [PubMed: 12748199]
23. Daymont C, Ross ME, Russell Localio A, Fiks AG, Wasserman RC, Grundmeier RW. Automated identification of implausible values in growth data from pediatric electronic health records. J Am Med Inform Assoc 2017; 24(6):1080–1087. [PubMed: 28453637]
24. Nichols GA, Schroeder EB, Karter AJ, Gregg EW, Desai J, Lawrence JM, O'Connor PJ, Xu S, Newton KM, Raebel MA, Pathak RD, Waitzfelder B, Segal J, Lafata JE, Butler MG, Lester H, Thomas A, Steiner JF. Trends in diabetes incidence among 7 million insured adults, 2006–2011: the SUPREME-DM project. Am J Epidemiol 2014; 181(1):32–39. [PubMed: 25515167]
25. Teixeira PL, Wei W-Q, Cronin RM, Mo H, VanHouten JP, Carroll RJ, LaRose E, Bastarache LA, Rosenbloom ST, Edwards TL, Roden DM, Lasko TA, Dart RA, Nikolai AM, Peissig PL, Denny JC. Evaluating electronic health record data sources and algorithmic approaches to identify hypertensive individuals. J Am Med Inform Assoc 2017; 24(1):162–171. [PubMed: 27497800]
26. Wall HK, Hannan JA, Wright JS. Patients with undiagnosed hypertension: hiding in plain sight. JAMA 2014; 312(19):1973–1974. [PubMed: 25399269]

27. Banerjee D, Chung S, Wong EC, Wang EJ, Stafford RS, Palaniappan LP. Underdiagnosis of hypertension using electronic health records. Am J Hypertens 2012; 25(1):97–102. [PubMed: 22031453]
28. Schulz WL, Young HP, Coppi A, Mortazavi BJ, Lin Z, Jean RA, Krumholz HM. Temporal relationship of computed and structured diagnoses in electronic health record data. BMC Med Inform Decis Mak 2021; 21:61. [PubMed: 33596898]
29. Rakotz MK, Ewigman BG, Sarav M, Ross RE, Robicsek A, Konchak CW, Gavagan TF, Baker DW, Hyman DJ, Anderson KP, Masi CM. A technology-based quality innovation to identify undiagnosed hypertension among active primary care patients. Ann Fam Med 2014; 12(4):352–358. [PubMed: 25024244]
30. Bower JK, Patel S, Rudy JE, Felix AS. Addressing bias in electronic health record-based surveillance of cardiovascular disease risk: finding the signal through the noise. Curr Epidemiol Rep 2017; 4:346–352. [PubMed: 31223556]
31. Bautista LE. Predictors of persistence with antihypertensive therapy: results from the NHANES. Am J Hypertens 2008; 21(2):183–188. [PubMed: 18188161]
32. Hwang KO, Aigbe A, Ju H-H, Jackson VC, Sedlock EW. Barriers to accurate blood pressure measurement in the Medical Office. J Prim Care Community Health 2018; 9:2150132718816929. [PubMed: 30764705]
33. Everson J, Barker W, Patel V. Electronic health record developer market segmentation contributes to divide in physician interoperable exchange. J Am Med Inform Assoc 2022; 29(7):1200–1207. [PubMed: 35442438]
34. Min L, Ha J-K, Hofer TP, Sussman J, Langa K, Cushman WC, Tinetti M, Kim HM, Maciejewski ML, Gillon L, Larkin A, Chan C-L, Kerr E. Validation of a health system measure to capture intensive medication treatment of hypertension in the Veterans Health Administration. JAMA Netw Open 2020; 3(7):e205417. [PubMed: 32729919]
35. McDonough CW, Babcock K, Chucri K, Crawford DC, Bian J, Modave F, Cooper-DeHoff RM, Hogan WR. Optimizing identification of resistant hypertension: computable phenotype development and validation. Pharmacoepidemiol Drug Saf 2020; 29(11):1393–1401. [PubMed: 32844549]

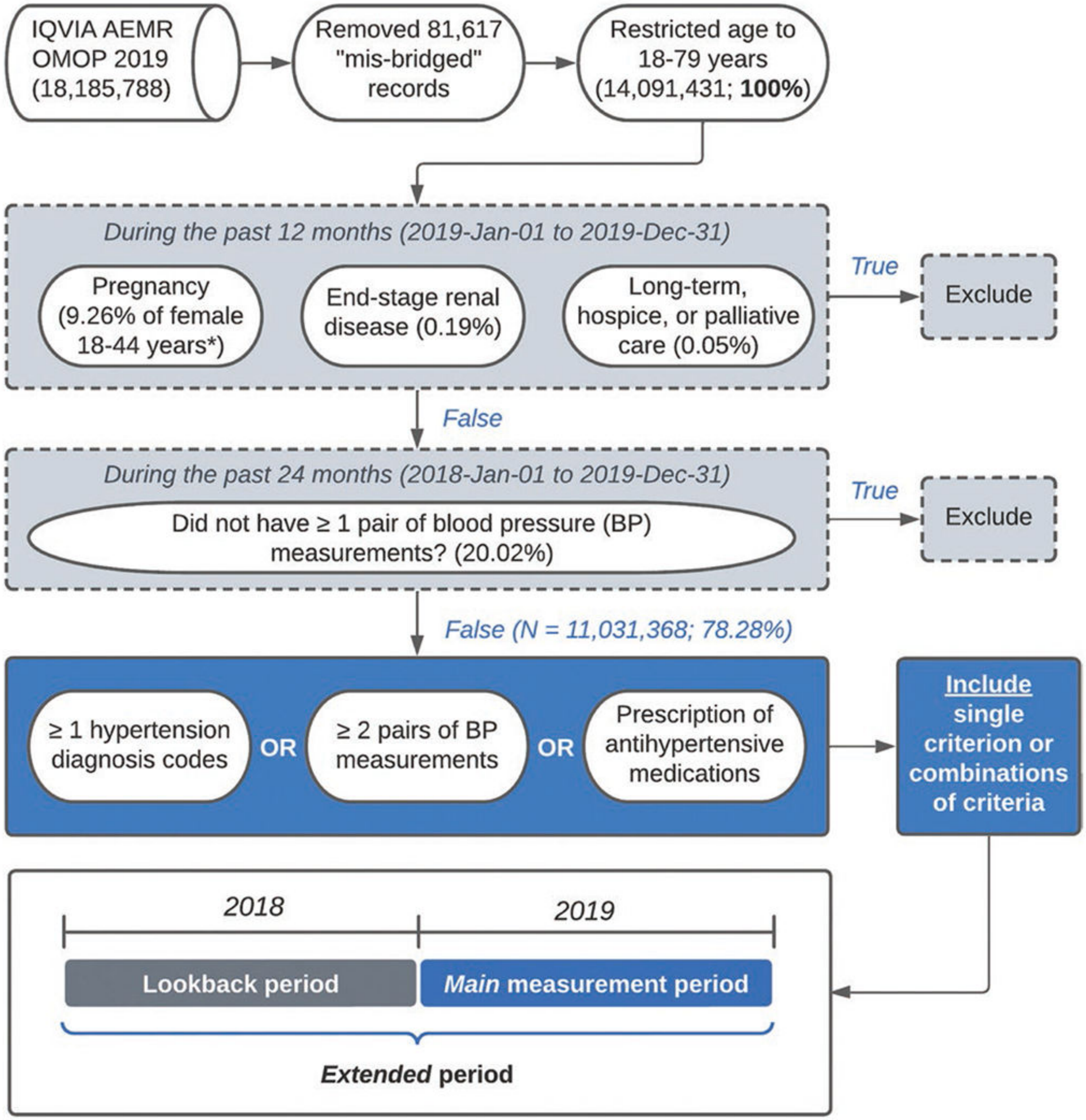


Figure 1.

Constructing hypertension e-phenotypes in AEMR-US. *The denominator for pregnancy: females aged 18–44 years (N = 2,876,436); All other items used the denominator noted as “100%”. Abbreviation: IQVIA AEMR OMOP, IQVIA Ambulatory Electronic Medical Record-US dataset, in Observational Medical Outcomes Partnership format.

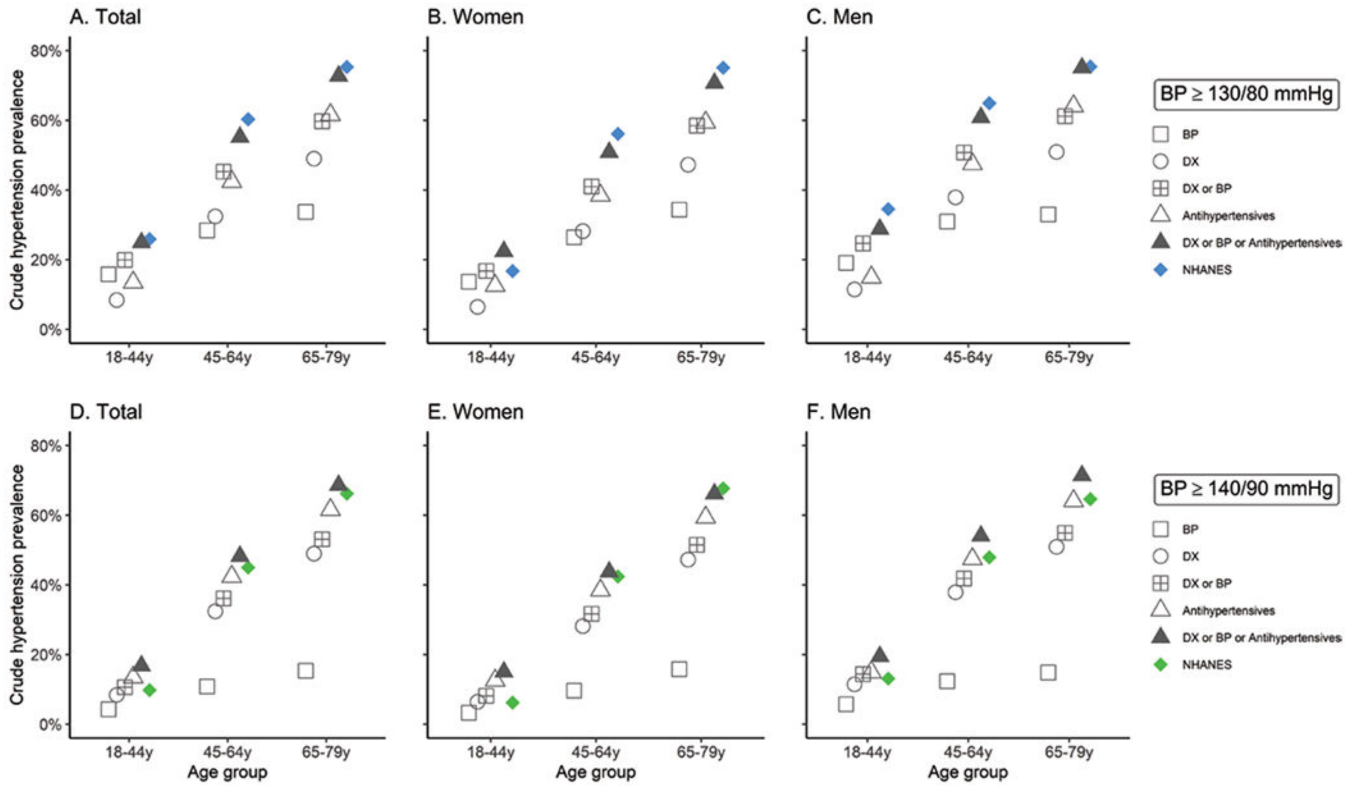


Figure 2.

Crude hypertension prevalence by age in AEMR-US 2019 based on selected criteria, compared with NHANES 2017–18. Main measurement period, 2019. Abbreviations: AEMR-US, IQVIA Ambulatory Electronic Medical Record-US dataset (2019); BP, blood pressure, ≥2 BP measurements; DX, ≥1 diagnosis code; Antihypertensives, ≥1 antihypertensive medication; NHANES, National Health and Nutrition Examination Survey (2017–18); Hypertension in NHANES: BP ≥140/90 or 130/80, or self-reported antihypertensive medication use.

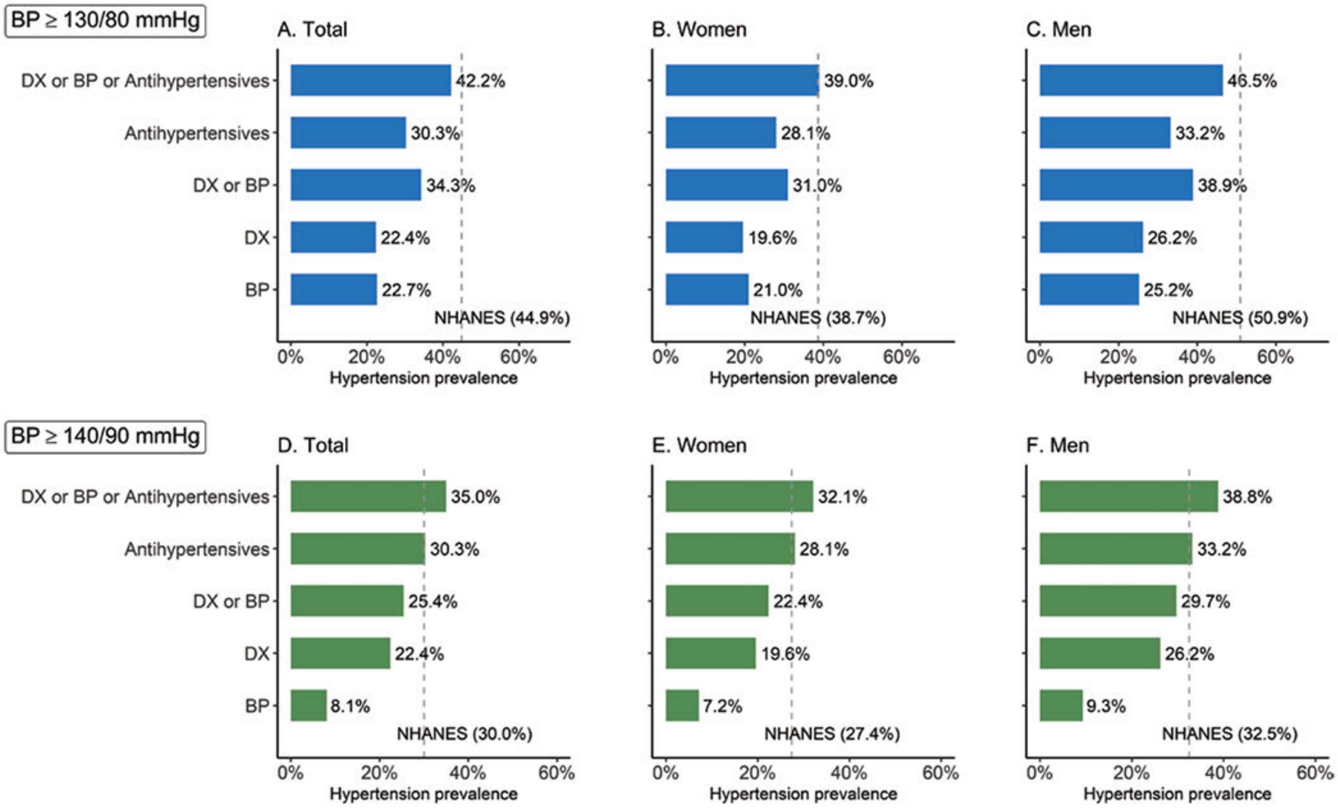


Figure 3.

Age-standardized hypertension prevalence by sex in AEMR-US based on selected criteria, compared with NHANES. BP cutoffs are consistent between NHANES and AEMR-US in each panel. Main measurement period, 2019. Abbreviations: AEMR-US, IQVIA Ambulatory Electronic Medical Record-US dataset (2019); BP, blood pressure, ≥2 BP measurements; DX, ≥1 diagnosis code; Antihypertensives, ≥1 antihypertensive medication; NHANES, National Health and Nutrition Examination Survey (2017–18); Hypertension in NHANES: BP ≥140/90 or 130/80, or self-reported antihypertensive medication use.

Table 1.

Characteristics of the study population based on candidate e-phenotypes for hypertension in AEMR-US

Characteristics	All eligible patients	Diagnosis codes only*		BP measurements only†		Antihypertensives only‡
		1DX	2DX	≥ 130/80 mm Hg	≥ 140/90 mm Hg	
Sample size (%)	11,031,368 (100.0%)	3,248,085 (29.4%)	2,088,583 (18.9%)	2,852,563 (25.9%)	1,103,481 (10.0%)	4,257,341 (38.6%)
Age, years, mean (SD)	52.3 (16.7)	61.8 (11.8)	62.4 (11.5)	57.4 (14.4)	60.1 (13.1)	60.9 (12.5)
Women, %	56.7	49.9	50.3	53.1	51.7	51.8
Race/ethnicity, %§						
Black	8.0	11.1	11.5	9.9	11.9	10.2
White	66.8	67.7	68.7	68.4	66.5	68.0
Asian	2.2	1.7	1.7	1.6	1.5	1.6
Hispanic	0.4	0.2	0.1	0.2	0.2	0.2
Other	4.6	5.0	5.6	6.2	6.6	5.1
Unknown	18.1	14.4	12.4	13.7	13.2	15.0
Prescribed ≥1 antihypertensive medication, %	38.6	89.1	92.8	61.9	73.5	100.0
BP measurements//						
No BP in the past 1 year, %	10.1	1.0	0.4	0.0	0.0	4.2
Number of BP measurements ≥130/80 mm Hg, median (IQR)	1 (0, 2)	2 (1,3)	2 (1,4)	3 (2, 4)	4 (3, 6)	1 (1, 3)
Number of BP measurements ≥140/90 mm Hg, median (IQR)	0 (0, 1)	1 (0, 2)	1 (0,2)	1 (1, 2)	3 (2, 4)	0 (0, 1)
Latest BP <130/80 mm Hg, %	40.0	30.8	31.9	17.0	11.9	32.9
Latest BP <140/90 mm Hg, %	70.3	66.5	68.6	60.8	33.5	66.5
Current smoker, %	10.3	13.5	14.5	14.7	15.9	12.8
BMI, kg/m^2, mean (SD)	29.9 (7.1)	32.0 (7.2)	32.2 (7.3)	31.9 (7.4)	32.2 (7.6)	31.8 (7.3)
Diabetes, %#	15.0	32.0	36.6	24.7	28.5	28.2

Abbreviations: AEMR-US, IQVIA Ambulatory Electronic Medical Record-US dataset (2019); BMI, body mass index; BP, blood pressure; DX, diagnosis code; SD, standard deviation.

* Only using diagnosis codes to define hypertension; 1DX (or 2) refers to ≥1 (or 2) diagnosis codes.

† Only using BP measurements to define hypertension. BP measurements were two different-day (systolic and diastolic) BP measurements.

‡ Only using antihypertensive medications to define hypertension.

[§]IQVIA combines race and ethnicity in one field, which may cause loss of true racial ethnic identity data at the patient level.

[//]Among eligible patients with ≥1 BP measurements in the two-year period (2018–2019), the number of BP measurements range from 0 to 352.

[#]Diabetes information was for adults ≥ 20 years based on a published algorithm.

Table 2.

Patients defined as having suspected hypertension using one, two, or three criteria in AEMR-US

Identification of suspected hypertension [*],[†]	**BP ≥130/80 mm Hg**		**BP ≥140/90 mm Hg**	
	Main period[§] ***N*** **= 5,569,867 (%)**	**Extended period**[§] ***N*** **= 6,667,488 (%)**	**Main period** ***N*** **= 4,851,279 (%)**	**Extended period** ***N*** **= 5,543,716 (%)**
Single criterion				
DX alone	4.1	2.3	6.2	4.4
BP alone	17.2	22.9	4.9	7.3
Antihypertensives alone	18.3	12.9	25.4	21.3
Two-criteria overlap				
DX and BP	2.3	2.5	1.1	1.4
DX and antihypertensives	26.4	18.1	45.6	39.3
BP and antihypertensives	6.2	8.2	2.7	4.0
Three-criteria overlap				
DX and BP and antihypertensives	25.5	33.1	14.0	22.3

Abbreviations: AEMR-US, IQVIA Ambulatory Electronic Medical Record-US data; BP, blood pressure; DX, diagnosis code.

[*] The denominator (100%) for each column is indicated in the secondary column heading.

[†] DX, one or more diagnosis codes; BP, two or more high BP measurements; antihypertensives, one or more prescription of antihypertensive medication.

[§] Main period, 2019; extended period, 2018–19, with 2018 serving as the lookback period in the numerator for case identification.

Table 3.

Controlled hypertension in 2019 in AEMR-US and NHANES 2017–18, by age and sex

Patients with controlled hypertension*, %	AEMR-US†		NHANES‡	
	Crude prevalence, %	Age-standardized prevalence§, %	Crude prevalence, %	Age-standardized prevalence, %
BP <130/80 mm Hg				
Total	32.8	31.5	19.4	14.5
By sex				
Female	34.8	36.1	22.6	17.3
Male	30.6	26.2	16.8	13.0
By age group, year				
18–44	31.3	–	5.9	–
45–64	30.2	–	20.6	–
65–79	35.7	–	30.8	–
BP <140/90 mm Hg				
Total	67.2	66.8	45.4	40.0
By sex				
Female	68.9	70.8	46.2	43.3
Male	65.2	62.1	44.6	38.7
By age group, year				
18–44	66.5	–	33.2	–
45–64	66.7	–	46.1	–
65–79	67.7	–	49.4	–

Abbreviations: AEMR-US, IQVIA Ambulatory Electronic Medical Record-US data; BP, blood pressure; NHANES, National Health and Nutrition Examination Survey (2017–18); "–", not applicable.

*Hypertension control is among patients who were identified as suspected hypertension cases, with at least one BP measurement during the main measurement period: if the most recent BP measurements were less than the cutoff points (<130/80, or <140/90 mm Hg, respectively), it is considered controlled; otherwise uncontrolled.

†The denominator for AEMR-US are patients identified as having hypertension using the e-phenotype of 1 diagnosis code, or 2 high BP, or antihypertensive medications.

‡The denominator for NHANES are participants who had high BP or reported taking antihypertensives.

§Standardized based on 2010 US census's age categories, including age groups 18–44, 45–64, and 65–79 years.