

Original Paper

# Identification of Hypertension in Electronic Health Records Through Computable Phenotype Development and Validation for Use in Public Health Surveillance: Retrospective Study

Nimish Valvi<sup>1,2\*</sup>, MPH, DrPH; Timothy McFarlane<sup>3</sup>, MPH, PhD; Katie S Allen<sup>1,4</sup>, BS; P Joseph Gibson<sup>5</sup>, MPH, PhD; Brian Edward Dixon<sup>1,4\*</sup>, MPA, PhD

<sup>1</sup>Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN, United States

<sup>2</sup>Department of Nutrition and Health Science, College of Health, Ball State University, Muncie, IN, United States

<sup>3</sup>Indiana Family and Social Services Administration, Indianapolis, IN, United States

<sup>4</sup>Department of Health Policy & Management, Fairbanks School of Public Health, Indiana University, Indianapolis, IN, United States

<sup>5</sup>CDC Foundation, Atlanta, GA, United States

\*these authors contributed equally

**Corresponding Author:**

Brian Edward Dixon, MPA, PhD  
Department of Health Policy & Management  
Fairbanks School of Public Health  
Indiana University  
1050 Wishard Blvd  
RG 5000  
Indianapolis, IN, 46202  
United States  
Phone: 1 317 278 3072  
Email: [bedixon@regenstrief.org](mailto:bedixon@regenstrief.org)

## Abstract

**Background:** Electronic health record (EHR) systems are widely used in the United States to document care delivery and outcomes. Health information exchange (HIE) networks, which integrate EHR data from the various health care providers treating patients, are increasingly used to analyze population-level data. Existing methods for population health surveillance of essential hypertension by public health authorities may be complemented using EHR data from HIE networks to characterize disease burden at the community level.

**Objective:** We aimed to derive and validate computable phenotypes (CPs) to estimate hypertension prevalence for population-based surveillance using an HIE network.

**Methods:** Using existing data available from an HIE network, we developed 6 candidate CPs for essential (primary) hypertension in an adult population from a medium-sized Midwestern metropolitan area in the United States. A total of 2 independent clinician reviewers validated the phenotypes through a manual chart review of 150 randomly selected patient records. We assessed the precision of CPs by calculating sensitivity, specificity, positive predictive value (PPV),  $F_1$ -score, and validity of chart reviews using prevalence-adjusted bias-adjusted  $\kappa$ . We further used the most balanced CP to estimate the prevalence of hypertension in the population.

**Results:** Among a cohort of 548,232 adults, 6 CPs produced PPVs ranging from 71% (95% CI 64.3%-76.9%) to 95.7% (95% CI 84.9%-98.9%). The  $F_1$ -score ranged from 0.40 to 0.91. The prevalence-adjusted bias-adjusted  $\kappa$  revealed a high percentage agreement of 0.88 for hypertension. Similarly, interrater agreement for individual phenotype determination demonstrated substantial agreement (range 0.70-0.88) for all 6 phenotypes examined. A phenotype based solely on diagnostic codes possessed reasonable performance ( $F_1$ -score=0.63; PPV=95.1%) but was imbalanced with low sensitivity (47.6%). The most balanced phenotype ( $F_1$ -score=0.91; PPV=83.5%) included diagnosis, blood pressure measurements, and medications and identified 210,764 (38.4%) individuals with hypertension during the study period (2014-2015).

**Conclusions:** We identified several high-performing phenotypes to identify essential hypertension prevalence for local public health surveillance using EHR data. Given the increasing availability of EHR systems in the United States and other nations,

leveraging EHR data has the potential to enhance surveillance of chronic disease in health systems and communities. Yet given variability in performance, public health authorities will need to decide whether to seek optimal balance or declare a preference for algorithms that lean toward sensitivity or specificity to estimate population prevalence of disease.

(*JMIR Form Res* 2023;7:e46413) doi: [10.2196/46413](https://doi.org/10.2196/46413)

## KEYWORDS

computable phenotypes; electronic health records; health information exchange; hypertension; population surveillance; public health informatics

## Introduction

Hypertension is the most prevalent risk factor for mortality throughout the world, and it was reported as a primary or contributing cause of death for over 500,000 Americans in 2019 [1]. Moreover, hypertension is reported as a comorbid condition for nearly 70% of individuals who have their first myocardial infarction and almost 80% of those who have their first stroke [2]. Overall, approximately 1 out of 3 adults in the United States is diagnosed with hypertension, which translates to almost 75 million Americans [3]. Moreover, given recent changes to guidelines for the prevention, detection, evaluation, and management of hypertension, the number of Americans considered to have this condition is expected to increase in the future [4].

The surveillance of conditions such as hypertension is a cornerstone of public health practice [5], yet quantifying the prevalence of hypertension at granular person, place, and time levels remains a challenge for local health departments (LHDs). Existing methods for capturing prevalence data for chronic disease within a geographic area rely upon community-based surveys. The 2 surveys most widely used by LHDs are the Behavioral Risk Factor Surveillance System and the National Health and Nutrition Examination Survey [6,7]. Those surveys provide reliable hypertension estimates at state and national levels, respectively, but have important limitations in their timeliness, breadth, and cost [8]. Moreover, these surveys provide very imprecise disease estimates at the local level (eg, county and neighborhood), where most public health interventions occur.

In addition to the Behavioral Risk Factor Surveillance System and the National Health and Nutrition Examination Survey, some LHDs conduct their own community surveys of health and health behaviors. While community surveys may be representative of the local population and powered for granular estimation, they can be costly to perform and are often spaced years apart [9]. Furthermore, community surveys typically do not include any medical, dental, or physiological measurements, limiting their ability to reliably measure true disease states as well as adherence to clinical guidelines, such as proper management of diabetes, hypertension, and other chronic illnesses.

Given the limitations of existing methods, LHDs seek alternative methods for obtaining timely information on health behaviors and risk factors prevalent in their community. Since the passing of the Health Information Technology for Economic and Clinical Health Act of 2009, electronic health records (EHR) systems

have become more common in the United States [10], representing a potential source of chronic disease surveillance data among those seeking health care.

Over 70% of ambulatory providers in the United States have adopted EHR systems [10]. As health care systems increasingly capture data from routine health care visits in EHR systems, national initiatives, including the digital Learning Health System of the National Academy of Medicine, the Robert Wood Johnson Foundation's Data for Health, and the Multi-State EHR-Based Network for Disease Surveillance [11], aim to leverage such data to improve the delivery of health care and community health outcomes [12]. The hope is that by leveraging existing digital data sources, public health agencies may access more timely and complete information to better assess and improve health in their communities.

Previous studies leveraged EHR data from primary care information systems funded by public health agencies but did not capture EHR data from other systems [13]. To explore the use of data combined from multiple EHR systems covering a large population in a community, we sought to develop and validate hypertension computable phenotypes (CPs) using EHR data available through a community-based health information exchange (HIE) network. We further sought to assess the HIE network as a source for accurate estimates of hypertension prevalence for a geographically defined population. HIE networks are increasingly available in many jurisdictions and nations [14], potentially providing public health authorities with a mechanism to access multiple electronic records for the same patient from a wide array of hospitals, physician practices, and other health care organizations that account for a representative subset of the population.

To assess the measurement of hypertension prevalence using an HIE network for an LHD population, we empirically derived and examined 6 hypertension CPs using data extracted from multiple EHR systems deployed across 3 distinct, large integrated delivery networks operating in Marion County, Indiana. This paper reports the prevalence rates derived from 6 hypertension phenotypes as well as the performance of each CP compared to a human chart review.

## Methods

### Data Source

The Indiana Network for Patient Care (INPC), launched in the 1990s, is one of the largest interorganizational clinical data repository in the United States, with more than 10 billion clinical data elements [15]. The INPC serves as the primary platform for the Indiana Health Information Exchange, which connects

more than 100 hospitals, 14,000 practices, and nearly 40,000 providers. Each participating institution submits clinical data from its EHR system to a centralized repository (the INPC) managed by the Indiana Health Information Exchange. Incoming data are matched to existing patients using an enterprise master person index (eg, master patient index or client registry), and clinical concepts (eg, laboratory results or blood pressure readings) are mapped to standardized terminologies for storage in the common data repository structure [16].

The INPC captures information on 99% of the population of Marion County, Indiana, which is home to Indianapolis. According to the 2020 census [17], Marion County had a resident population of 971,102 with a racial composition of 62.4% White, 29.6% Black or African American, and 11.3% Hispanic; 51.6% female; and 13.1% adults aged 65 years or older.

For this study, a subset of 3 health systems was used, representing at least 80% (776,882/971,102) of the population of Marion County. The first is an essential health provider consisting of a large tertiary hospital and 10 federally qualified health centers geographically spread across the county. The second is the state's largest health system, with 3 tertiary hospitals in Marion County, including the region's largest children's hospital, a level 1 trauma center, and the state's largest cancer center. The final system is a regional health system that includes another large children's hospital, a tertiary hospital, and a large network of primary care and neighborhood emergency departments across the county. Each health system independently represents a large portion of Marion County residents based on available health market share data, and each included system-approved participation in the study.

### Study Population

To examine CPs for hypertension, we extracted a cohort of all adults (at least 18 years of age as of January 1, 2014) living in Marion County (all patient addresses in INPC are geocoded [18]) who sought care at 1 of the 3 large integrated delivery networks between January 1, 2014, and December 31, 2015.

### CPs for Essential (Primary) Hypertension

There is no standard CP for defining essential (primary) hypertension based on clinically derived data. To that effect, a central goal of this work was to propose and test EHR-based phenotypes of essential hypertension. These algorithms were developed using available definitions for essential hypertension from the US Centers for Disease Control and Prevention (CDC) as well as the American Heart Association (AHA) [19] along with input from epidemiologists and medical doctors. We examined a range of phenotype definitions (Table 1) because evidence of hypertension exists in several places across a patient's EHR, with data contributed from multiple health systems. All included data elements were required to be within the 2014–2015 period specified for this analysis. While many clinicians document hypertension through routine billing using International Classification of Diseases (ICD) codes, previous studies found this kind of documentation to be incomplete. In addition to clinical diagnoses, we incorporated physiologic blood pressure readings for some phenotypes, given that the CDC and the US Preventative Services Taskforce [19,20] recommend screening for hypertension using office-based measurements. When using blood pressure readings, hypertension was characterized by systolic blood pressure measurements of at least 140 mm Hg or diastolic blood pressure measurements of at least 90 mm Hg based on guidelines published before 2017 by the American College of Cardiology and the AHA [3]. However, blood pressure recordings were not available for all individuals in the population. Furthermore, a meta-analysis revealed major accuracy limitations, including misdiagnosis for office-based blood pressure measurements [21]. In addition, previous studies reported that medication data can be useful when identifying individuals with chronic illness, as they are typically on maintenance medication [22–24]. We therefore examined 1 phenotype that included pharmacy data downloaded from a national repository [25] of claims data. The list of hypertension medications used was originally created by pharmacist health services researchers at the Regenstrief Institute (in Multimedia Appendix 1).

**Table 1.** Computable phenotype definitions used to examine hypertension prevalence among adults in Marion County, Indiana (2014-2015).

Computable phenotype	Definition	Data examined
1	Individual has at least 1 hypertension ICD-9-CM <sup>a</sup> or ICD-10-CM <sup>b</sup> diagnostic code documented for at least inpatient or outpatient encounter ( $\geq 1$ hypertension diagnostic code)	Primary diagnosis as well as secondary diagnoses or comorbidities associated with a clinical encounter documented in the patient's EHR <sup>c</sup>
2	Individual has at least 1 BP <sup>d</sup> measurement in which the systolic BP was at least 140 mm Hg or diastolic BP was at least 90 mm Hg ( $\geq 1$ BP reading)	Point-of-care- and laboratory-based physiologic measurements recorded in the patient's EHR
3	Individual has at least 2 BP measurements in which the systolic BP was at least 140 mm Hg or diastolic BP was at least 90 mm Hg ( $\geq 2$ BP readings)	Point-of-care- and laboratory-based physiologic measurements recorded in the patient's EHR
4	Individual has at least 1 BP measurement in which the systolic BP was at least 140 mm Hg or diastolic was at least 90 mm Hg ( $\geq 1$ BP reading), and individual has at least 1 hypertension ICD-9-CM or ICD-10-CM diagnostic code documented for at least inpatient or outpatient encounter ( $\geq 1$ hypertension diagnostic code)	Primary diagnosis as well as secondary diagnoses or comorbidities, along with point-of-care-based physiologic measurements recorded in the patient's EHR
5	Individual has at least 1 BP measurement in which the systolic BP was at least 140 mm Hg or diastolic BP was at least 90 mm Hg ( $\geq 1$ BP reading), and individual has at least 2 hypertension ICD-9-CM or ICD-10-CM diagnostic codes documented for 2 different inpatient or outpatient encounters ( $\geq 2$ hypertension diagnostic codes)	Primary diagnosis as well as secondary diagnoses or comorbidities, along with point-of-care- and laboratory-based physiologic measurements recorded in the patient's EHR
6	Individual has at least 1 BP measurement in which the systolic BP was at least 140 mm Hg or diastolic BP was at least 90 mm Hg ( $\geq 1$ BP reading), individual has at least 1 hypertension ICD-9-CM or ICD-10-CM diagnostic code documented for at least inpatient or outpatient encounter ( $\geq 1$ hypertension diagnostic code), or individual has filled at least 1 prescription for any medication associated with hypertension ( $\geq 1$ hypertension medication)	Primary diagnosis as well as secondary diagnoses or comorbidities, along with point-of-care- and laboratory-based physiologic measurements recorded in the patient's EHR and pharmacy claims records for filled prescriptions downloaded from a national repository

<sup>a</sup>ICD-9-CM: International Classification of Disease-9 Clinical Modification.

<sup>b</sup>ICD-10-CM: International Classification of Disease-10 Clinical Modification.

<sup>c</sup>EHR: electronic health record.

<sup>d</sup>BP: blood pressure.

## Validation of CPs and Chart Review

To validate the phenotypes, we used a chart review using a sample of 299 individuals randomly selected from the population. A retired outpatient cardiovascular nurse with more than 40 years of service reviewed the merged, longitudinal EHR for patients in the sample using the INPC. In the chart review, based on clinical judgment, it was noted whether the patient had hypertension during the study period based on the recorded data in the chart. Chart reviewers had access to the full clinical record contained within the HIE, which included clinical text as well as structured data. In addition, the reviewer documented any evidence supporting the clinical assessment of hypertension. The phenotype definitions were applied to the 299 individuals in the random sample to examine their performance. Phenotype performance was quantified using sensitivity, specificity, and positive predictive value (PPV), including the 95% CIs for each measure, and the  $F_1$ -score.

These phenotypes were then applied to the entire population extracted from the INPC to estimate hypertension prevalence for Marion County, Indiana. Prevalence rates were calculated by age, sex, race, and ethnicity.

A blinded chart review was conducted on patient records at the Regenstrief Institute. The chart review was conducted independently, and the 2 reviewers were trained before the review. A total of 299 charts were randomly selected from the

INPC, of which 233 were reviewed by reviewer 1. To calculate interrater reliability, we compared the hypertension classifications from 150 charts reviewed by a second reviewer, another outpatient nurse with 20 years of experience, with those of the first reviewer.

The standard measure of agreement is Cohen  $\kappa$  coefficient ( $\kappa$ ). It adjusts the observed agreement by the agreement expected by chance. However, if no further adjustments are made,  $\kappa$  can be deceptive because it is sensitive to both the bias in reporting "yes" to the classification of hypertension, if any exists, and the prevalence of "yes" relative to "no" in the sample. We therefore calculated the prevalence-adjusted bias-adjusted  $\kappa$  (PABAK) [26-28] for the case definition and the phenotypes. PABAK values were interpreted according to the guidelines for  $\kappa$  provided by Landis and Koch [29]: 0.81-1.00=almost perfect agreement; 0.61-0.80=substantial agreement; 0.41-0.60=moderate agreement; 0.21-0.40=fair agreement; and 0.01-0.20=slight agreement.

All analyses and phenotype coding were conducted using SAS (version 9.4; SAS Institute Inc).

## Ethical Considerations

Study approval was obtained from the Indiana University Institutional Review Board (exempt protocol 1701925087). Informed consent was waived due to the retrospective use of preexisting, deidentified data from medical records.

Data managers at the Regenstrief Institute extracted EHR data from the 3 health systems using the INPC clinical data repository. Data from a third health system were extracted by analysts at the health system, then linked and merged with the INPC data as this health system does not contribute ambulatory clinic data to the INPC repository. A total of 6 CPs were derived from the data (Table 1), and prevalence was calculated by dividing the number of individuals meeting each essential hypertension case definition by the total cohort population.

## Results

### Population Characteristics

Data were extracted from the INPC from January 1, 2014, to December 31, 2015, and included 548,232 adults aged 18 years or older from the 3 health systems within Marion County. The majority of people were White (308,213/548,232, 56.2%), female (335,548/548,232, 61.2%), and adults aged 65 years or older (93,513/548,232, 17.1%; Table 2).

**Table 2.** Study characteristics and demographics using electronic health records from the Indiana Network for Patient Care, 2014-2015 (N=548,232).

Characteristic	Overall (N=548,232)	Hypertension	
		No (n=337,468)	Yes (n=210,764)
<b>Age (years), n (%)</b>			
18-39	214,655 (39.1)	161,878 (75.4)	52,777 (24.6)
40-64	240,064 (43.8)	138,648 (57.8)	101,416 (42.2)
≥65	93,513 (17.1)	36,942 (39.5)	56,571 (60.5)
<b>Sex, n (%)</b>			
Female	335,548 (61.2)	214,241 (63.8)	121,307 (36.2)
Male	212,684 (38.8)	123,227 (57.9)	89,457 (42.1)
<b>Race, n (%)</b>			
White	308,213 (56.2)	187,381 (60.8)	120,832 (39.2)
Black	148,117 (27)	78,057 (52.7)	70,060 (47.3)
Other	91,902 (16.8)	72,030 (78.4)	19,872 (21.6)

### CPs using INPC

A total of 299 records were randomly drawn from the INPC. The chart review was completed for 233 records by reviewer 1 (Table 3). According to chart reviewer 1 expert opinion, 167 individuals were identified with hypertension, a prevalence of 71.7%. Of the 167 individuals with hypertension, the phenotype algorithms identified 82 (CP1), 107 (CP2), 55 (CP3), 47 (CP4), 142 (CP5), and 200 (CP6), respectively. The final CP (CP6)

was determined to have a sensitivity of 99.9% (95% CI 97.8%-100.0%) with a PPV of 83.5% (95% CI 79.9%-86.6%). However, CP6 had a comparatively lower specificity of 50% (95% CI 37.4%-62.6%) compared to CP1, with a specificity of 93.9% (95% CI 85.2%-98.3%). For CP6, the  $F_1$ -score was closest to 1.0 at 0.91, suggesting that this algorithm has the most balanced performance across error types. Using the  $F_1$ -score as an evaluation metric, the next best-performing algorithm is CP5 (0.71), followed by CP1 (0.63).

**Table 3.** Performance of computable phenotypes (CPs) for hypertension validated using randomly selected electronic health records from the Indiana Network for Patient Care (n=233). The prevalence estimate on manual chart review (reviewer 1) was determined to be 71.7% (n=167).

Phenotypes	Total confirmed hypertension cases, n	Prevalence, n (%)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive predictive value, % (95% CI)	$F_1$ -score
≥1 Clinical diagnosis (CP1)	78	82 (35.2)	46.7 (38.9-54.6)	93.9 (85.2-98.3)	95.1 (88.1-98.0)	0.63
≥1 Vitals indicated (CP2)	76	107 (45.9)	45.5 (37.8-53.4)	53.0 (40.3-65.4)	71.0 (64.3-76.9)	0.55
≥2 Vitals indicated (CP3)	44	55 (23.6)	26.3 (19.8-33.7)	83.3 (72.1-91.4)	80.0 (68.8-87.9)	0.40
≥1 Clinical diagnosis and ≥1 vitals indicated (CP4)	45	47 (20.2)	26.9 (20.4-34.3)	97.0 (89.5-99.6)	95.7 (84.9-98.9)	0.42
≥1 Clinical diagnosis or ≥1 vitals indicated (CP5)	109	142 (60.9)	65.3 (57.5-72.5)	50.0 (37.4-62.6)	76.7 (71.7-81.2)	0.71
≥1 Clinical diagnosis, ≥1 vitals indicated, or ≥1 medications indicated (CP6)	167	200 (85.8)	99.9 (97.8-100.0)	50.0 (37.4-62.6)	83.5 (79.9-86.6)	0.91

## Interrater Reliability

We assessed interrater reliability between the 2 raters using PABAK. A total of 150 charts were reviewed by both reviewers out of the 233 charts that reviewer 1 examined. The interrater

agreement (PABAK=0.91) was an almost perfect agreement for the 2 reviewers. The estimated PABAK (Table 4) for the phenotypes ranged from substantial (0.70) to almost perfect agreement (0.88).

**Table 4.** Hypertension phenotypes agreement between 2 reviewers and prevalence-adjusted bias-adjusted  $\kappa$  (PABAK; n=150).

Phenotype	HTN <sup>a</sup> , n	HTN prevalence estimate, % (95% CI)	PABAK <sup>b</sup>
HTN case determination	101	67.3 (59.2-74.8)	0.91
≥1 Clinical diagnosis (CP1 <sup>c</sup> )	35	23.3 (16.8-30.9)	0.70
≥1 Vitals indicated overall (CP2)	55	36.7 (28.9-44.9)	0.73
≥2 Vitals indicated (CP3)	29	19.3 (13.3-26.6)	0.81
≥1 Clinical diagnosis and ≥1 vitals indicated (CP4)	20	13.3 (8.3-19.8)	0.76
≥1 Clinical diagnosis or ≥1 vitals indicated (CP5)	71	47.3 (39.1-55.6)	0.71
≥1 Clinical diagnosis, ≥1 vitals indicated, or ≥1 medications indicated (CP6)	117	78.0 (70.5-84.3)	0.88

<sup>a</sup>HTN: hypertension.

<sup>b</sup>PABAK =  $2 \times ((\text{positive agreement} + \text{negative agreement})/N) - 1$ .

<sup>c</sup>CP: computable phenotype.

In Table 2, the overall hypertension prevalence using CP6 during the study period was estimated at (210,764/548,232, 38.4%). The prevalence of hypertension was highest among adults aged 65 years or older (56,571/95,513, 60.5%), female individuals (121,307/335,548, 36.2%), and Black individuals (70,060/148,117, 47.3%) during the study period.

## Discussion

### Principal Findings

We developed and evaluated 6 CPs using a combination of ICD-9 and 10 codes, medications, and vitals for essential hypertension using EHR data extracted from multiple health systems. Despite variation in CP performance, most CPs possessed PPVs above 75%, making these CPs reasonable for use in public health surveillance. Unlike in controlled trials or other research contexts in which investigators seek to maximize specificity (eg, the ability to identify individuals without disease), public health surveillance seeks to balance specificity with sensitivity (eg, the ability to identify all cases in the population). In this study, CP1, having good specificity, might serve as an initial screening for selecting individuals who might benefit from more targeted interventions.

### Comparison With Previous Work

Our findings align with previous work examining the performance of approaches that rely primarily on diagnostic codes compared to other CP approaches. Relying solely on diagnostic codes (CP1) resulted in low sensitivity (46.7%) yet high specificity (93.9%) and PPV (95.1%) to identify individuals with hypertension. This performance is quite similar to the use of ICD codes for identifying cases of sexually transmitted infections, which also possess high specificity (99.9%) and reasonable PPV (≥85%) yet low sensitivity (<15%) [30]. The imbalance is also reflected in its  $F_1$ -score of 0.63, which is adequate but may not be sufficient for public health surveillance,

where most epidemiologists prefer to err on the side of sensitivity in case detection.

The broadest phenotype (CP6) that allowed the identification of individuals on antihypertensive therapies along with clinical diagnoses and blood pressure readings possessed high sensitivity (99.9%) but lower specificity (50%) and PPV (83.5%). However, the  $F_1$ -score (0.91) suggests CP6 as the best algorithm for overall balance. It is worth noting that one potential reason for the low specificity could be due to the treatment of hypertension, resulting in a reduced blood pressure measurement during the window of observation. This would indicate individuals with hypertension who meet the CP definition based on medication but may not have evidence of hypertension as a reason for a visit in the clinical text or abnormal blood pressure readings during the observation window. This and other indications for hypertensive classification could be explored further in future studies.

Multiple strategies to identify the prevalence of chronic diseases have been used previously, either from a single hospital or within a network-based EHR system [31-35]. Our hypertension phenotype based on diagnostic codes alone (CP1) was much lower in sensitivity (46.7%) compared to previous studies from the United States (83%) [36], Canada (85%) [34], and Switzerland (83%) [35]. A likely explanation for this is that the previous studies had participants mainly from primary-care centers, while this study used data from both outpatient and hospitalization encounters. Furthermore, many small primary care practices do not contribute data to the INPC, limiting data capture in some parts of the metropolitan area. When additional sources of evidence of hypertension were included in the phenotype, the sensitivity increased. However, after including additional sources, the specificity was reduced to 50%. The increase in sensitivity after including medications can partly be explained by a study [37] that found that individuals with an appropriate diagnosis of hypertension were more likely to be treated (92.6%). Population prevalence (38.4%) calculated using

CP6 was much higher compared to a diverse patient population in a California [37] health system (28.7%), which used a similar definition of hypertension [38]. However, this population prevalence was comparable to a study from New York (38.4% vs 39.2%) [36], with a large distributed EHR network having a similar time period.

Increasing the availability of standardized data within EHR-based surveillance systems, including the adoption and use of the US Core Data for Interoperability (USCDI), can further improve the accuracy and completeness of CPs to estimate the prevalence of hypertension. EHR-based diagnoses capture an individual's state or disease status; however, they are based on the interpretation of the clinical staff treating the patient. Using secondary data for estimating prevalence, public health professionals often do not have the power to enforce standardization of data capture from the originating health systems [39]. Further studies should explore the comparison of prevalence from EHR-based surveillance systems with that of standardized, representative [40] data captured through other methods. With the Office of the National Coordinator for Health Information Technology requiring implementation of USCDI, all of the CPs used in this study could be implemented through the Fast Healthcare Interoperability Resources interfaces throughout the United States in all certified EHR systems updated after December 31, 2022.

The data for this study were extracted from 3 different health systems that participate in the INPC, managed by the Indiana HIE. Leveraging HIE infrastructures to aggregate data across health care facilities for public health purposes is increasingly viewed as critical in the wake of the COVID-19 pandemic [41]. As discussed during the 2022 American College of Medical Informatics Symposium, many HIE networks are working to become health data utilities that serve as information infrastructures in support of public health [42]. Unfortunately, few states have robust HIE infrastructures such as that of the INPC. Yet, there is hope that policies such as TEFCA (Trusted Exchange Framework and Common Agreement) [43] and the emphasis by the Office of the National Coordinator for Health Information Technology can spur to creation and expansion of HIE infrastructures across the nation [44]. This will remain a key area for research and development as public health seeks to modernize its data infrastructure and HIE networks expand their support for population health. The work presented here is just the beginning of efforts to better support public health surveillance through the EHR and HIE systems.

### Limitations

We acknowledge the limitations of using EHR data, some of which are due to the administrative nature that can produce misclassification (ie, coding errors or missed fields). Additionally, in-care adult populations are more likely to be female, older, non-Hispanic, and insured compared to the not-in-care adult population [45]. In comparison to the

demographic distribution in the general population, there was a higher representation of women and older adults in our EHR cohort. This could either underestimate hypertension prevalence for younger age groups and for male individuals or overestimate for female individuals and older adults in our population. Data obtained from EHR-HIE systems can have variable quality since they are obtained from diverse health delivery systems due to differences in documentation processes across providers and clinicians. For instance, some practices may record information only from standardized fields, while others may capture values using free-text fields. All 3 health systems included in this study use a commercial EHR system. Once clinical encounter data are submitted from the EHR to the HIE, they undergo quality control steps to normalize data to the extent possible, including standardization of terminology. Additionally, the phenotypes did not leverage unstructured data, which may have resulted in missing data, depending on how the institutions store the elements required for the phenotype. We did anticipate some data quality issues such as missingness or data inaccuracies, but this limitation is applicable to a minority of cases [46,47]. We further note that the  $F_1$ -measure, which is widely used in information retrieval, is calculated based on precision and recall, and it does not account for accuracy (eg, true negatives), which means it may not be an ideal measure for performance in cases involving clinical diagnosis [48]. Lastly, we used both ICD-9 Clinical Modification and ICD-10 Clinical Modification codes to identify this study population, yet we could not account for performance differences between the coding systems.

Despite limitations, using EHR-based prevalence estimates for population health has several benefits. They provide larger sample sizes while affording granular person, place, and time that are unavailable from existing population-based self-reported surveys. The data are more timely and more affordable than locally commissioned surveys. Furthermore, the collection and tabulation of estimates can be done more frequently and require less human effort than traditional approaches. This would enable resource-limited LHDs to routinely assess chronic disease burden and trend data over time.

### Conclusions

We constructed 6 CPs to estimate the prevalence of hypertension in support of public health surveillance for chronic disease. With the help of manual chart reviews, we were able to capture the variation between phenotypes. In the future, we plan to use these phenotypes to compare prevalence with estimates from population-based health surveys. EHR-based estimates for chronic illnesses are helpful for public health surveillance and regional quality improvement efforts at much lower costs compared to traditional population survey approaches. As the adoption and use of HIE systems and standards such as the USCDI increase, the quality of population health metrics should improve over time.

### Acknowledgments

The authors acknowledge support from the Regenstrief Data Services team, especially Jane Wang, Amy Hancock, and Anna Roberts, who facilitated access to Indiana Network for Patient Care data for this study.

This work was supported by a contract, “Evaluation of DOH Capacity for Using EHR Data for Cardiovascular Disease Surveillance,” from the US Centers for Disease Control and Prevention (CDC) to the Task Force for Global Health. NV and BED are further supported by a contract (NU38OT000286) from the CDC to the National Association of Chronic Disease Directors for improving chronic disease surveillance and management through the use of EHRs and health information systems. NV and BED receive additional support from the CDC through grants and contracts for the study of COVID-19 and other infectious diseases. The funders did not have a role in the writing or editing of this manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the CDC.

### Data Availability

The data sets generated and analyzed during this study are not publicly available due to privacy and governance related issues but are available from the corresponding author on reasonable request and with appropriate governance.

### Authors' Contributions

The project was conceptualized by BED and PJG, with methodological input from TM. Data management and analysis were conducted by TM, NV, and KSA. NV drafted the manuscript. All authors contributed critical reviews and important edits to the manuscript. All authors approved the final version. All authors assume public responsibility for the accuracy and integrity of the work.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

List of hypertension medications using National Drug Codes.

[\[DOCX File , 592 KB-Multimedia Appendix 1\]](#)

### References

1. National Center for Health Statistics. About multiple cause of death, 1999-2019. CDC WONDER. Atlanta, GA. Centers for Disease Control and Prevention; 2019. URL: [https://web.archive.org/web/20211129042726if\\_/https://wonder.cdc.gov/controller/saved/D77/D241F661](https://web.archive.org/web/20211129042726if_/https://wonder.cdc.gov/controller/saved/D77/D241F661) [accessed 2023-11-10]
2. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. *Circulation*. 2015;131(4):e29-e322 [FREE Full text] [doi: [10.1161/CIR.000000000000152](https://doi.org/10.1161/CIR.000000000000152)] [Medline: [25520374](https://pubmed.ncbi.nlm.nih.gov/25520374/)]
3. Merai R, Siegel C, Rakotz M, Basch P, Wright J, Wong B, et al. CDC grand rounds: a public health approach to detect and control hypertension. *MMWR Morb Mortal Wkly Rep*. 2016;65(45):1261-1264 [FREE Full text] [doi: [10.15585/mmwr.mm6545a3](https://doi.org/10.15585/mmwr.mm6545a3)] [Medline: [27855138](https://pubmed.ncbi.nlm.nih.gov/27855138/)]
4. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Himmelfarb CD, et al. 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. *J Am Coll Cardiol*. 2018;71(19):e127-e248 [FREE Full text] [doi: [10.1016/j.jacc.2017.11.006](https://doi.org/10.1016/j.jacc.2017.11.006)] [Medline: [29146535](https://pubmed.ncbi.nlm.nih.gov/29146535/)]
5. Lee LM, Thacker SB, Centers for Disease Control and Prevention (CDC). The cornerstone of public health practice: public health surveillance, 1961--2011. *MMWR Suppl*. 2011;60(4):15-21 [Medline: [21976162](https://pubmed.ncbi.nlm.nih.gov/21976162/)]
6. Chowdhury PP, Mawokomatanda T, Xu F, Gamble S, Flegel D, Pierannunzi C, et al. Surveillance for certain health behaviors, chronic diseases, and conditions, access to health care, and use of preventive health services among states and selected local areas?- Behavioral risk factor surveillance system, United States, 2012. *MMWR Surveill Summ*. 2016;65(4):1-142 [FREE Full text] [doi: [10.15585/mmwr.ss6504a1](https://doi.org/10.15585/mmwr.ss6504a1)] [Medline: [27124212](https://pubmed.ncbi.nlm.nih.gov/27124212/)]
7. Hales CM, Carroll MD, Simon PA, Kuo T, Ogden CL. Hypertension prevalence, awareness, treatment, and control among adults aged ≥18 years—Los Angeles county, 1999-2006 and 2007-2014. *MMWR Morb Mortal Wkly Rep*. 2017;66(32):846-849 [FREE Full text] [doi: [10.15585/mmwr.mm6632a3](https://doi.org/10.15585/mmwr.mm6632a3)] [Medline: [28817553](https://pubmed.ncbi.nlm.nih.gov/28817553/)]
8. Pierannunzi C, Hu SS, Balluz L. A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004-2011. *BMC Med Res Methodol*. 2013;13:49 [FREE Full text] [doi: [10.1186/1471-2288-13-49](https://doi.org/10.1186/1471-2288-13-49)] [Medline: [23522349](https://pubmed.ncbi.nlm.nih.gov/23522349/)]
9. Stone K, Sierocki A, Shah V, Ylitalo KR, Horney JA. Conducting community health needs assessments in the local public health department: a comparison of random digit dialing and the community assessment for public health emergency response. *J Public Health Manag Pract*. 2018;24(2):155-163 [doi: [10.1097/PHH.0000000000000522](https://doi.org/10.1097/PHH.0000000000000522)] [Medline: [28141672](https://pubmed.ncbi.nlm.nih.gov/28141672/)]
10. Jamoom E, Yang N, Hing E. Adoption of certified electronic health record systems and electronic information sharing in physician offices: United States, 2013 and 2014. *NCHS Data Brief*. 2016(236):1-8 [FREE Full text] [Medline: [26828707](https://pubmed.ncbi.nlm.nih.gov/26828707/)]



11. Kraus EM, Brand B, Hohman KH, Baker EL. New directions in public health surveillance: using electronic health records to monitor chronic disease. *J Public Health Manag Pract*. 2022;28(2):203-206 [FREE Full text] [doi: [10.1097/PHH.0000000000001501](https://doi.org/10.1097/PHH.0000000000001501)] [Medline: [35100219](https://pubmed.ncbi.nlm.nih.gov/35100219/)]
12. Institute of Medicine (US). In: Grossmann C, Powers B, McGinnis JM, editors. *Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary*. Washington, DC. National Academies Press; 2011.
13. Perlman SE, McVeigh KH, Thorpe LE, Jacobson L, Greene CM, Gwynn RC. Innovations in population health surveillance: using electronic health records for chronic disease surveillance. *Am J Public Health*. 2017;107(6):853-857 [doi: [10.2105/AJPH.2017.303813](https://doi.org/10.2105/AJPH.2017.303813)] [Medline: [28426302](https://pubmed.ncbi.nlm.nih.gov/28426302/)]
14. Dixon BE. Introduction to health information exchange. In: *Health Information Exchange: Navigating and Managing a Network of Health Information Systems*. 2nd Edition. San Diego, CA, USA. Academic Press; 2022:3-20
15. Overhage JM, Kansky JP. The Indiana health information exchange. In: Dixon BE, editor. *Health Information Exchange: Navigating and Managing a Network of Health Information Systems*. 2nd Edition. San Diego, CA, USA. Academic Press; 2022:471-486
16. Zafar A, Dixon BE. Pulling back the covers: technical lessons of a real-world health information exchange. *Stud Health Technol Inform*. 2007;129(Pt 1):488-492 [Medline: [17911765](https://pubmed.ncbi.nlm.nih.gov/17911765/)]
17. Population estimates, American community survey, and current population survey. United States Census Bureau. 2022. URL: <https://www.census.gov/quickfacts/marioncountyindiana> [accessed 2023-11-10]
18. Comer KF, Grannis S, Dixon BE, Bodenhamer DJ, Wiehe SE. Incorporating geospatial capacity within clinical data systems to address social determinants of health. *Public Health Rep*. 2011;126(Suppl 3):54-61 [FREE Full text] [doi: [10.1177/00333549111260S310](https://doi.org/10.1177/00333549111260S310)] [Medline: [21836738](https://pubmed.ncbi.nlm.nih.gov/21836738/)]
19. Go AS, Bauman MA, King SMC, Fonarow GC, Lawrence W, Williams KA, et al. An effective approach to high blood pressure control: a science advisory from the American Heart Association, the American College of Cardiology, and the Centers for Disease Control and Prevention. *Hypertension*. 2014;63(4):878-885 [FREE Full text] [doi: [10.1161/HYP.0000000000000003](https://doi.org/10.1161/HYP.0000000000000003)] [Medline: [24243703](https://pubmed.ncbi.nlm.nih.gov/24243703/)]
20. Anderson AE, Kerr WT, Thames A, Li T, Xiao J, Cohen MS. Electronic health record phenotyping improves detection and screening of type 2 diabetes in the general United States population: a cross-sectional, unselected, retrospective study. *J Biomed Inform*. 2016;60:162-168 [FREE Full text] [doi: [10.1016/j.jbi.2015.12.006](https://doi.org/10.1016/j.jbi.2015.12.006)] [Medline: [26707455](https://pubmed.ncbi.nlm.nih.gov/26707455/)]
21. Guirguis-Blake JM, Evans CV, Webber EM, Coppola EL, Perdue LA, Weyrich MS. Screening for hypertension in adults: updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA*. 2021;325(16):1657-1669 [FREE Full text] [doi: [10.1001/jama.2020.21669](https://doi.org/10.1001/jama.2020.21669)] [Medline: [33904862](https://pubmed.ncbi.nlm.nih.gov/33904862/)]
22. Dixon BE, Zou JF, Comer KF, Rosenman M, Craig JL, Gibson P. Using electronic health record data to improve community health assessment. *Front Public Health Serv Sys Res*. 2016;5(5):50-56 [FREE Full text] [doi: [10.13023/FPHSSR.0505.08](https://doi.org/10.13023/FPHSSR.0505.08)]
23. Nguyen KA, Haggstrom DA, Ofner S, Perkins SM, French DD, Myers LJ, et al. Medication use among veterans across health care systems. *Appl Clin Inform*. 2017;8(1):235-249 [FREE Full text] [doi: [10.4338/ACI-2016-10-RA-0184](https://doi.org/10.4338/ACI-2016-10-RA-0184)] [Medline: [28271121](https://pubmed.ncbi.nlm.nih.gov/28271121/)]
24. Zhu VJ, Tu W, Rosenman MB, Overhage JM. Facilitating clinical research through the health information exchange: lipid control as an example. *AMIA Annu Symp Proc*. 2010;2010:947-951 [FREE Full text] [Medline: [21347118](https://pubmed.ncbi.nlm.nih.gov/21347118/)]
25. National drug code directory. U.S. Food and Drug Administration. URL: <https://www.accessdata.fda.gov/scripts/cder/ndc/index.cfm> [accessed 2023-11-10]
26. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993;46(5):423-429 [doi: [10.1016/0895-4356\(93\)90018-v](https://doi.org/10.1016/0895-4356(93)90018-v)] [Medline: [8501467](https://pubmed.ncbi.nlm.nih.gov/8501467/)]
27. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6(4):284-290 [doi: [10.1037/1040-3590.6.4.284](https://doi.org/10.1037/1040-3590.6.4.284)]
28. McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, For GG, et al. Evidence-Based Medicine Teaching Tips Working Group. Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). *CMAJ*. 2004;171(11):1369-1373 [FREE Full text] [doi: [10.1503/cmaj.1031981](https://doi.org/10.1503/cmaj.1031981)] [Medline: [15557592](https://pubmed.ncbi.nlm.nih.gov/15557592/)]
29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174 [Medline: [843571](https://pubmed.ncbi.nlm.nih.gov/843571/)]
30. Ho YA, Rahurkar S, Tao G, Patel CG, Arno JN, Wang J, et al. Validation of international classification of diseases, tenth revision, clinical modification codes for identifying cases of chlamydia and gonorrhoea. *Sex Transm Dis*. 2021;48(5):335-340 [FREE Full text] [doi: [10.1097/OLQ.0000000000001257](https://doi.org/10.1097/OLQ.0000000000001257)] [Medline: [32740450](https://pubmed.ncbi.nlm.nih.gov/32740450/)]
31. Aguilar-Palacio I, Carrera-Lasfuentes P, Poblador-Plou B, Prados-Torres A, Rabanaque-Hernández MJ, por el Grupo de Investigación en Servicios Sanitarios de Aragón (GRISSA). Morbidity and drug consumption. comparison of results between the National Health Survey and electronic medical records. *Gac Sanit*. 2014;28(1):41-47 [FREE Full text] [doi: [10.1016/j.gaceta.2013.04.004](https://doi.org/10.1016/j.gaceta.2013.04.004)] [Medline: [23680136](https://pubmed.ncbi.nlm.nih.gov/23680136/)]
32. Barber J, Muller S, Whitehurst T, Hay E. Measuring morbidity: self-report or health care records? *Fam Pract*. 2010;27(1):25-30 [FREE Full text] [doi: [10.1093/fampra/cmp098](https://doi.org/10.1093/fampra/cmp098)] [Medline: [20019091](https://pubmed.ncbi.nlm.nih.gov/20019091/)]

33. Catalán-Ramos A, Verdú JM, Grau M, Iglesias-Rodal M, del Val García JL, Consola A, et al. Population prevalence and control of cardiovascular risk factors: what electronic medical records tell us. *Aten Primaria*. 2014;46(1):15-24 [FREE Full text] [doi: [10.1016/j.aprim.2013.06.004](https://doi.org/10.1016/j.aprim.2013.06.004)] [Medline: [24325864](https://pubmed.ncbi.nlm.nih.gov/24325864/)]
34. Williamson T, Green ME, Birtwhistle R, Khan S, Garies S, Wong ST, et al. Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records. *Ann Fam Med*. 2014;12(4):367-372 [FREE Full text] [doi: [10.1370/afm.1644](https://doi.org/10.1370/afm.1644)] [Medline: [25024246](https://pubmed.ncbi.nlm.nih.gov/25024246/)]
35. Zellweger U, Bopp M, Holzer BM, Djalali S, Kaplan V. Prevalence of chronic medical conditions in Switzerland: exploring estimates validity by comparing complementary data sources. *BMC Public Health*. 2014;14:1157 [FREE Full text] [doi: [10.1186/1471-2458-14-1157](https://doi.org/10.1186/1471-2458-14-1157)] [Medline: [25377723](https://pubmed.ncbi.nlm.nih.gov/25377723/)]
36. Thorpe LE, McVeigh KH, Perlman S, Chan PY, Bartley K, Schreibstein L, et al. Monitoring prevalence, treatment, and control of metabolic conditions in New York city adults using 2013 primary care electronic health records: a surveillance validation study. *EGEMS (Wash DC)*. 2016;4(1):1266 [FREE Full text] [doi: [10.13063/2327-9214.1266](https://doi.org/10.13063/2327-9214.1266)] [Medline: [28154836](https://pubmed.ncbi.nlm.nih.gov/28154836/)]
37. 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 [FREE Full text] [doi: [10.1038/ajh.2011.179](https://doi.org/10.1038/ajh.2011.179)] [Medline: [22031453](https://pubmed.ncbi.nlm.nih.gov/22031453/)]
38. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, et al. 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](https://doi.org/10.1001/jama.289.19.2560)] [Medline: [12748199](https://pubmed.ncbi.nlm.nih.gov/12748199/)]
39. Richesson R, Wiley LK, Gold S, Rasmussen L. Electronic health records–based phenotyping: introduction. In: *Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials*. Bethesda, Maryland. NIH Pragmatic Trials Collaboratory; 2021.
40. Ostchega Y, Fryar CD, Nwankwo T, Nguyen DT. Hypertension prevalence among adults aged 18 and over: United States, 2017-2018. *NCHS Data Brief*. 2020(364):1-8 [FREE Full text] [Medline: [32487290](https://pubmed.ncbi.nlm.nih.gov/32487290/)]
41. Dixon BE, Grannis SJ, McAndrews C, Broyles AA, Mikels-Carrasco W, Wiensch A, et al. Leveraging data visualization and a statewide health information exchange to support COVID-19 surveillance and response: application of public health informatics. *J Am Med Inform Assoc*. 2021;28(7):1363-1373 [FREE Full text] [doi: [10.1093/jamia/ocab004](https://doi.org/10.1093/jamia/ocab004)] [Medline: [33480419](https://pubmed.ncbi.nlm.nih.gov/33480419/)]
42. Indiana Rural Health Association. The importance of health data utilities in supporting public health. 2022 Presented at: 2022 Spring Rural Summit; April 12, 2022; Indianapolis, Indiana URL: <https://www.indianaruralhealth.org/events/2022/04/12/in-person-events/2022-spring-rural-summit/>
43. User's guide to the trusted exchange framework and common agreement—TEFCA. The Sequoia Project. 2022. URL: <https://rce.sequoiaproject.org/wp-content/uploads/2022/01/Common-Agreement-Users-Guide.pdf> [accessed 2023-11-10]
44. Adler-Milstein J, Worzala C, Dixon BE. Future directions for health information exchange. In: Dixon BE, editor. *Health Information Exchange: Navigating and Managing a Network of Health Information Systems*. 2nd Edition. San Diego, CA, USA. Academic Press; 2022.
45. Romo ML, Chan PY, Lurie-Moroni E, Perlman SE, Newton-Dame R, Thorpe LE, et al. Characterizing adults receiving primary medical care in New York city: implications for using electronic health records for chronic disease surveillance. *Prev Chronic Dis*. 2016;13:E56 [FREE Full text] [doi: [10.5888/pcd13.150500](https://doi.org/10.5888/pcd13.150500)] [Medline: [27126554](https://pubmed.ncbi.nlm.nih.gov/27126554/)]
46. Berry DJ, Kessler M, Morrey BF. Maintaining a hip registry for 25 years. Mayo Clinic experience. *Clin Orthop Relat Res*. 1997(344):61-68 [FREE Full text] [doi: [10.1097/00003086-199711000-00007](https://doi.org/10.1097/00003086-199711000-00007)] [Medline: [9372758](https://pubmed.ncbi.nlm.nih.gov/9372758/)]
47. Wisniewski MF, Kieszkowski P, Zagorski BM, Trick WE, Sommers M, Weinstein RA. Development of a clinical data warehouse for hospital infection control. *J Am Med Inform Assoc*. 2003;10(5):454-462 [FREE Full text] [doi: [10.1197/jamia.M1299](https://doi.org/10.1197/jamia.M1299)] [Medline: [12807807](https://pubmed.ncbi.nlm.nih.gov/12807807/)]
48. Hand DJ, Christen P, Kirielle N. F\*: an interpretable transformation of the F-measure. *Mach Learn*. 2021;110(3):451-456 [FREE Full text] [doi: [10.1007/s10994-021-05964-1](https://doi.org/10.1007/s10994-021-05964-1)] [Medline: [33746357](https://pubmed.ncbi.nlm.nih.gov/33746357/)]

## Abbreviations

- AHA:** American Heart Association
- CDC:** US Centers for Disease Control and Prevention
- CP:** computable phenotype
- EHR:** electronic health record
- HIE:** health information exchange
- ICD:** International Classification of Diseases
- INPC:** Indiana Network for Patient Care
- LHD:** local health department
- PABAK:** prevalence-adjusted bias-adjusted  $\kappa$
- PPV:** positive predictive value
- TEFCA:** Trusted Exchange Framework and Common Agreement

**USCDI: US Core Data for Interoperability**

*Edited by A Mavragani; submitted 13.02.23; peer-reviewed by A Higaki, W Song; comments to author 17.05.23; revised version received 21.07.23; accepted 07.11.23; published 27.12.23*

*Please cite as:*

*Valvi N, McFarlane T, Allen KS, Gibson PJ, Dixon BE*

*Identification of Hypertension in Electronic Health Records Through Computable Phenotype Development and Validation for Use in Public Health Surveillance: Retrospective Study*

*JMIR Form Res 2023;7:e46413*

*URL: <https://formative.jmir.org/2023/1/e46413>*

*doi: [10.2196/46413](https://doi.org/10.2196/46413)*

*PMID: [38150296](https://pubmed.ncbi.nlm.nih.gov/38150296/)*

©Nimish Valvi, Timothy McFarlane, Katie S Allen, P Joseph Gibson, Brian Edward Dixon. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 27.12.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.