

Developing Electronic Health Record-Based Electronic Phenotypes for Chronic Disease Indicators

Enhancing chronic disease surveillance with clinical data from electronic health records (EHRs) is of growing interest as public health agencies modernize their data systems. Although numerous clinical measures exist, they generally focus on measuring the operation and performance of health centers or quality of care provided. To advance chronic disease surveillance efforts, it is important to leverage EHR data to define relevant chronic disease measures, identify/construct electronic phenotypes (e-phenotypes), and address key gaps in public health knowledge and practice.

Defining chronic disease measures from the EHR—through the construction of e-phenotypes—enables the description of the population-level prevalence of health conditions and related factors. EHR data can address some limitations of established public health surveillance systems by providing more timely and geographically granular data, including more detailed clinical information about disease occurrence, diagnosis, treatment, and management (e.g., whether hypertension is controlled). Deciding whether to use existing chronic disease surveillance measures or develop new EHR-based e-phenotypes for chronic disease measures involves several considerations.

Glossary of Terms

Algorithm: a set of analytic instructions or decision rules that include/exclude specific clinical data (e.g., diagnosis codes), health encounters, and/or include a specific number of data sources.

Chronic Disease Indicator: the chronic disease condition or risk factor you want to describe.

Chronic Disease Measure: the measure of disease burden for a specific chronic disease (e.g., hypertension prevalence).

Electronic Phenotype: a single or set of definitions (e.g., case definition) for a clinical condition or characteristic that can be ascertained in an EHR system or clinical data repository by applying a defined algorithm through a computerized query.

Population: the group to which you want to extrapolate or generalize the disease burden estimates (e.g., adults over 50 years).

Repository: a centralized space for accessing data, algorithms, or measures.

Representativeness: the degree to which estimates from a sample reflect the intended population.

Validity: the accuracy of an electronic phenotype in capturing the intended chronic disease measure.

OBJECTIVE

This document describes a process that state, tribal, local, and territorial health departments can adopt to identify or develop chronic disease electronic phenotypes for surveillance of important chronic disease measures. These measures can be used to assess and monitor population health status as well as potentially evaluate public health programs.

Chronic Disease Measure and Electronic Phenotype Development Process

The first question that needs to be asked when considering a new chronic disease measure and related e-phenotype is, “What are our surveillance goals?”

The answer will help to identify conditions or concepts to measure. Second, it is important to define the population that you are trying to describe with a chronic disease measure. Once the goals, indicators, and population are clearly defined, public health departments can follow a systematic process to identify existing measures and e-phenotypes or develop new e-phenotypes as needed.

The figure on the following page describes a stepwise process that health departments can use to identify and/or develop e-phenotypes from EHR-based data that can advance population health surveillance.

The process of identifying and/or developing e-phenotypes for chronic disease measures involves six key steps. Although the figure shows it as a linear process, generally a more iterative approach is required. Each step is further described in the table that follows, and resources to address each step are provided.



Process for Identifying and Developing EHR-Based Electronic Phenotypes of Chronic Disease Measures



STEP 03 TIP:
Work with a healthcare professional throughout this process



Description of a Recommended Process for Identifying and Developing EHR-Based Electronic Phenotypes of Chronic Disease Measures (Hypertension prevalence as an example).

Step	Process	Purpose	Example	Resources
1A	Browse Existing Repositories	<p>It may not be necessary to create new electronic phenotypes (e-phenotypes).</p> <p>There may be existing e-phenotypes that can be used as is or as a starting point for further refinement based on your needs.</p> <p>Differentiating e-phenotypes for surveillance versus those for improving quality of clinical care or performance management is key.</p>	<p>We can search existing e-phenotype repositories such as:</p> <ul style="list-style-type: none"> • Observational Health Data Sciences and Informatics' (OHDSI) Phenotype Library on GitHub • ESP Public - Confluence • Chronic Conditions Data Warehouse • CMS' eCQM for relevant e-phenotypes 	<p>Hohman KH, Zambarano B, Klompas M, Wall HK, Kraus EM, Carton TW, et al. Development of a Hypertension Electronic Phenotype for Chronic Disease Surveillance in Electronic Health Records: Key Analytic Decisions and Their Effects. <i>Prev Chronic Dis</i> 2023;20:230026. DOI: http://dx.doi.org/10.5888/pcd20.230026</p>
1B	Search the Literature	<p>There may also be published literature discussing efforts and preliminary outcomes in creating an e-phenotype.</p> <p>Searching the literature can direct you to e-phenotypes that have been developed for clinical research.</p> <p>Although the e-phenotypes may be overly complicated to apply to epidemiological research or public health surveillance, they might provide a starting point for further development.</p>	<p>Search Google Scholar: (hypertension OR "high blood pressure") AND (EHR OR "electronic health record" OR "electronic medical record") AND (surveillance OR prevalence OR "public health") AND (indicator OR phenotype OR algorithm)</p>	<p>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. <i>J Public Health (Oxf)</i>. 2016 Sep;38(3):e392-e399. doi: 10.1093/pubmed/fdv155. Epub 2015 Nov 6. PMID: 26547088; PMCID: PMC5072168.</p>



Step	Process	Purpose	Example	Resources
02	<p>Understand the Components of the E-phenotype/Identify Appropriate Data Source</p>	<p>Commonly used phenotypes are created by bringing together different criteria—diagnosis codes, procedure codes, drug prescriptions and/or fills, vital signs, lab measurements, etc.</p> <p>Understanding the clinical data necessary to create the phenotype and determining what is available in the corresponding data source is critical to e-phenotype development.</p>		<p>OHDSI Observational Health Data Sciences and Informatics—Standardized Data: The OMOP Common Data Model</p>
03	<p>Translate E-phenotype Components into Codes</p>	<p>Once you know what components will be used, then you need to identify the actual codes or values that will constitute a “case.” These codes can include diagnosis codes (e.g., ICD), procedure codes (CPT, HCPCS), drug codes (NDC), or lab codes (LOINC).</p>		<p>Hohman KH, Zambarano B, Klompas M, Wall HK, Kraus EM, Carton TW, et al. Development of a Hypertension Electronic Phenotype for Chronic Disease Surveillance in Electronic Health Records: Key Analytic Decisions and Their Effects. <i>Prev Chronic Dis</i> 2023;20:230026. DOI: http://dx.doi.org/10.5888/pcd20.230026</p> <p>He S, Park S, Kuklina E, et al. Leveraging Electronic Health Records to Construct a Phenotype for Hypertension Surveillance in the United States. <i>Am J Hypertens</i>. 2023; 36(12):677–85. doi:10.1093/ajh/hpad081</p>



Step	Process	Purpose	Example	Resources
04	Create a New E-phenotype if Needed	Once you have worked through steps 2 and 3, you can create the algorithm to generate the desired electronic phenotype.	<p>Any of the following events within a 1-year period of time:</p> <ul style="list-style-type: none"> • A diagnosis of essential hypertension or secondary hypertension (ICD10: I10, I11, I12, I13, and I15) • A prescription for an antihypertensive medication* • High blood pressure (≥ 140 systolic or ≥ 90 diastolic) measured on two or more occasions within a year <p>Inclusion criteria:</p> <ul style="list-style-type: none"> • Patient had one or more valid blood pressure measures in the prior 2 years • Adults aged 18–85 years <p>Exclusion criteria:</p> <ul style="list-style-type: none"> • Diagnosis of end-stage-renal disease (I12.0, I13.11, I13.2, N18.6, 585.6) • Diagnosis of pregnancy within 1 year (ICD10: Z33, Z34, O09). 	<p>Hohman KH, Zambarano B, Klompas M, Wall HK, Kraus EM, Carton TW, et al. Development of a Hypertension Electronic Phenotype for Chronic Disease Surveillance in Electronic Health Records: Key Analytic Decisions and Their Effects. <i>Prev Chronic Dis</i> 2023;20:230026. DOI: http://dx.doi.org/10.5888/pcd20.230026</p> <p>He S, Park S, Kuklina E, et al. Leveraging Electronic Health Records to Construct a Phenotype for Hypertension Surveillance in the United States. <i>Am J Hypertens</i>. 2023;36(12):677-685. doi:10.1093/ajh/hpad081</p>



Step	Process	Purpose	Example	Resources
05	Extract EHR data and Generate Disease Burden Estimate/Assess Validity	<p>Once the electronic phenotype is developed, you can apply it to your population data (patient database) to generate disease burden estimates.</p> <p>Validation of the algorithm against expert chart review can determine accuracy.</p> <p>Validation against an external data source is important to determine the accuracy of the new phenotype.</p>		<p>Hohman KH, Zambarano B, Klompas M, Wall HK, Kraus EM, Carton TW, et al. Development of a Hypertension Electronic Phenotype for Chronic Disease Surveillance in Electronic Health Records: Key Analytic Decisions and Their Effects. <i>Prev Chronic Dis</i> 2023;20:230026. DOI: http://dx.doi.org/10.5888/pcd20.230026</p> <p>He S, Park S, Fujii Y, et al. State-Level Hypertension Prevalence and Control Among Adults in the U.S. <i>Am J Prev Med</i>. 2024;66(1):46-54. doi:10.1016/j.amepre.2023.09.010</p>
06	Consider Need for Bias Reduction	<p>The patient population in the EHR system may not represent the population to which you want to generalize.</p> <p>Weighting (e.g., post-stratification weighting or raking), imputation (for key variables based on missingness), or modeling (e.g., generalized linear mixed-models) methods may help to reduce bias.</p> <p>However, please note that, none of these methods is meant to replace data quality control methods prior to analysis.</p>		



Considerations and Conclusions

Identifying or developing EHR-based electronic phenotypes for chronic disease measures can enhance population-level chronic disease surveillance efforts. As with all data sources, there are limitations.

EHR data users need to consider that: 1) EHR data are collected primarily for billing and patient-level documentation, not for population-level disease surveillance; therefore, data quality and missingness can create analytic challenges; and 2) EHR data contain information about persons seeking medical care and may not be representative of the underlying population of interest.

Despite the limitations, as public health departments expand the types of data they use to modernize surveillance efforts, EHR data may be an important source to consider to enhance chronic disease surveillance and provide more timely, locally granular, and clinically detailed estimates of disease prevention and control.

MENDS, the Multi-State EHR-Based Network for Disease Surveillance, is an innovative data modernization pilot project that leverages timely electronic health record (EHR) data to enhance chronic disease surveillance. MENDS fosters information partnerships between healthcare organizations and health departments.

For more information, contact: mends@chronicdisease.org

The “Improving Chronic Disease Surveillance and Management Through the Use of Electronic Health Records/Health Information Systems” project is supported by the Centers for Disease Control and Prevention (CDC) of the U.S. Department of Health and Human Services (HHS) as part of a financial assistance award totaling \$1,800,000 with 100 percent funded by CDC/HHS. Disclaimer: The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement, by CDC/HHS, or the U.S. Government.

