

---

**Review**

# Electronic Medical Record–Based Case Phenotyping for the Charlson Conditions: Scoping Review

---

Seungwon Lee<sup>1,2,3,4</sup>, MPH; Chelsea Doktorchik<sup>1,2</sup>, MSc; Elliot Asher Martin<sup>1,3</sup>, PhD; Adam Giles D'Souza<sup>1,3</sup>, PhD; Cathy Eastwood<sup>1,2</sup>, RN, PhD; Abdel Aziz Shaheen<sup>1,2,5</sup>, MBChB, MSc; Christopher Naugler<sup>2,6</sup>, MD; Joon Lee<sup>1,2,4,7</sup>, PhD; Hude Quan<sup>1,2</sup>, PhD, MD

<sup>1</sup>Centre for Health Informatics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

<sup>2</sup>Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

<sup>3</sup>Alberta Health Services, Calgary, AB, Canada

<sup>4</sup>Data Intelligence for Health Lab, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

<sup>5</sup>Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

<sup>6</sup>Department of Pathology and Laboratory Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

<sup>7</sup>Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

**Corresponding Author:**

Hude Quan, PhD, MD

Department of Community Health Sciences

Cumming School of Medicine

University of Calgary

Teaching, Research, & Wellness Building

3280 Hospital Dr NW

Calgary, AB, T2N 4Z6

Canada

Phone: 1 403 210 9317

Email: [hquan@ucalgary.ca](mailto:hquan@ucalgary.ca)

---

**Abstract**

**Background:** Electronic medical records (EMRs) contain large amounts of rich clinical information. Developing EMR-based case definitions, also known as EMR phenotyping, is an active area of research that has implications for epidemiology, clinical care, and health services research.

**Objective:** This review aims to describe and assess the present landscape of EMR-based case phenotyping for the Charlson conditions.

**Methods:** A scoping review of EMR-based algorithms for defining the Charlson comorbidity index conditions was completed. This study covered articles published between January 2000 and April 2020, both inclusive. Embase (Excerpta Medica database) and MEDLINE (Medical Literature Analysis and Retrieval System Online) were searched using keywords developed in the following 3 domains: terms related to EMR, terms related to case finding, and disease-specific terms. The manuscript follows the Preferred Reporting Items for Systematic reviews and Meta-analyses extension for Scoping Reviews (PRISMA) guidelines.

**Results:** A total of 274 articles representing 299 algorithms were assessed and summarized. Most studies were undertaken in the United States (181/299, 60.5%), followed by the United Kingdom (42/299, 14.0%) and Canada (15/299, 5.0%). These algorithms were mostly developed either in primary care (103/299, 34.4%) or inpatient (168/299, 56.2%) settings. Diabetes, congestive heart failure, myocardial infarction, and rheumatology had the highest number of developed algorithms. Data-driven and clinical rule–based approaches have been identified. EMR-based phenotype and algorithm development reflect the data access allowed by respective health systems, and algorithms vary in their performance.

**Conclusions:** Recognizing similarities and differences in health systems, data collection strategies, extraction, data release protocols, and existing clinical pathways is critical to algorithm development strategies. Several strategies to assist with phenotype-based case definitions have been proposed.

(*JMIR Med Inform* 2021;9(2):e23934) doi: [10.2196/23934](https://doi.org/10.2196/23934)

**KEYWORDS**

electronic medical records; Charlson comorbidity; EMR phenotyping; health services research

## Introduction

### Background

Recent advances in computational power, increased adoption of electronic medical records (EMRs), and the subsequent rise of big data analytics in health care have opened the door to precision medicine [1]. EMRs are systemized collections of patient health information and documentation, collected in real time and stored in a digital format. EMRs were originally designed to facilitate communication in support of clinical decision-making for individual patients and to improve the quality of care. Canada and other countries have heavily promoted EMR adoption [2,3]. Globally, EMR data have been used widely for secondary purposes, such as research.

Developing case definitions, a process known as phenotyping, has become an active area of research associated with EMRs. Establishing EMR data-based phenotyping is essential for setting up the operational framework toward pursuing precision medicine, which aims to tailor medical decisions and treatments to each patient in a timely manner. EMR phenotyping allows identification and surveillance of health conditions in a timely manner and can be integrated into existing clinical flows and infrastructure. Phenotyping comorbidities using EMR data have important implications on disease management. Comorbidity is a medical condition existing simultaneously with but independently from another condition in a patient. These diseases may be related to each other by some shared association [4]. The Charlson comorbidity index [4-6] is a measure that predicts 1-year mortality based on the presence or absence of specific chronic conditions. Typically, each condition is identified through the presence of specific International Classification of Diseases (ICD) codes and assigned a score depending on the risk of death. Scores are summed for each patient to provide a total score to predict mortality [7,8]. The Charlson [5] comorbidity algorithm is the most widely used comorbidity index at present and has demonstrated the importance of classifying conditions using health data [6,7], including risk adjustment analysis, developing patient safety indicators, and identifying specific disease cohorts for research and public health applications.

### Objectives

Few reviews [9-12] have been published on developing EMR case definitions or phenotyping algorithms for selected chronic conditions, but none specifically cover all of the Charlson comorbidities. Furthermore, these articles narrowed their scope to specific perspectives [10] or specific settings (eg, inpatient or primary care only) [9,11]. These reviews report few studies utilizing natural language processing (NLP) or machine learning (ML), which emphasizes the importance of data science techniques (eg, deep learning) in the present health research. The primary objective of this study is to provide an overview of EMR-based phenotyping algorithms for the Charlson conditions. The secondary objective is to provide

recommendations for health systems considering the adoption of EMR-based case phenotyping.

## Methods

### Article Screening

The methodology follows the guidelines recommended by the Preferred Reporting Items for Systematic Reviews and Meta-analysis Extension Protocols for Scoping Reviews (PRISMA-ScR) [13]. The Excerpta Medica database (Embase), and Medical Literature Analysis and Retrieval System Online (MEDLINE) databases were searched from January 2000 to April 2020 to identify peer-reviewed papers. The search strategy covered the following 3 domains: (1) terms related to EMRs, (2) terms related to case finding, and (3) disease-specific terms. We initially used validated clinical text descriptions from ICD-10 to derive search terms for selected conditions (Multimedia Appendix 1). Boolean algorithms were developed for each specific condition using the domain keywords (Multimedia Appendix 2). The cancer categories of metastatic cancer and malignant cancer were excluded, as there is already an existing review on this topic [11].

Manual screening was performed according to the following established study guidelines. Peer-reviewed journal papers were included if they were published between January 2000 and April 2020, written in English, involved human subjects and EMR, and were retrieved by the Boolean search algorithm for at least one Charlson condition. This review study focused only on case phenotyping using EMR data, and therefore, papers were excluded if they only involved administrative databases. Administrative data studies that linked EMR data were included. The presence of the Charlson conditions in each study, if reported, was defined by the presence of ICD-9 or ICD-10 codes stated in the manuscript. The full PRISMA flow diagram was created (Multimedia Appendix 3). The final search results were exported to a reference software (EndNote, Clarivate Inc) [14], and duplicates were removed.

### Characterizing the Identified Literature

A data extraction form was developed. The extracted data components included article characteristics (year and country), health care type (eg, inpatient, outpatient, and emergency), specific name of the data source, whether diagnostic codes (eg, ICD) were used, types of EMR data (eg, structured, unstructured, or imaging), techniques (eg, epidemiology/biostatistics, ML, or NLP), and whether a validation methodology was employed. The extracted data types (categorical) were recoded as binary variables to indicate whether they were employed in the algorithm. The frequencies of the algorithms, EMR settings, and countries were calculated. The identified algorithms were substratified into the following 7 types in this review based on the types of data used: (1) diagnostic codes only; (2) codes and structured data (demographics, labs, and medications); (3) diagnostic codes and free-text data; (4) diagnostic codes, structured, and free-text data; (5) structured data only; (6)

free-text data only; and (7) free-text and structured data. The detailed operational definitions of case definitions used in the identified studies were also extracted. The extracted data were summarized using frequencies and graphs where applicable. STATA 14 software (StataCorp LLC) [15] was used for statistical analysis. We further summarized the used data elements, disease context, data linkage, and validation of phenotyping algorithms using the extracted tables.

## Results

### Article Screening

After 1097 duplicates were removed, a total of 3691 abstracts were identified from the electronic databases. A total of 3402 abstracts were excluded based on the title and abstract screening, resulting in 289 full-text articles for full article review. Of these, 39 articles were excluded because they did not include any Charlson conditions, and 22 articles could not be retrieved, leading to the exclusion of 61 articles. The remaining 228 articles were considered eligible for this review and analyzed. References of eligible full articles were screened, and additional articles were identified for inclusion ( $n=46$ ), leading to a total of 274 articles for qualitative synthesis. Articles covering multiple disease phenotypes were counted once per phenotype, leading to a total of 299 disease phenotyping algorithms. The PRISMA diagram depicting this process is shown in [Multimedia Appendix 3](#).

### Characteristics of the Identified Literature

The frequencies of the algorithms, EMR settings, and countries are shown in [Table 1](#). The complete data extraction table is presented in [Multimedia Appendix 4](#) [16-285]. A total of 274 articles representing 299 algorithms from 22 countries were identified in this review. The majority of this work was undertaken in the United States (181/299, 60.5%), followed by the United Kingdom (42/299, 14.0%) and Canada (15/299,

5.0%). Algorithm development has steadily increased over the years, with the majority of work published after 2016. The distributions of these algorithms by the year of publication and by country are shown in [Figure 1](#). The breakdown of the disease areas of these algorithms is shown in [Figure 2](#).

[Table 2](#) provides a summary of the algorithm types used for each Charlson condition. The most common algorithm types were diagnostic codes and structured data (167/299, 55.9%), followed by diagnostic codes, structured and free-text data (51/299, 17.1%), and diagnostic codes only (40/299, 13.4%). Variations in the data sources used were observed based on disease context and data availability.

These algorithms were mostly developed either in primary care (103/299, 34.4%) or inpatient (168/299, 56.2%) settings. A total of 23 algorithms (23/299, 7.7%) used data sources from inpatient and outpatient EMR. This trend was consistent across the conditions assessed in this review. The United States had the highest algorithm count across most of the assessed conditions, followed by the United Kingdom, Canada, and other nations. Detailed information about the distribution of algorithms by disease, EMR setting, and country is shown in [Table 1](#).

We abstracted study objectives and classified different purposes for which algorithms were developed for, as well as the setting of each study ([Multimedia Appendix 4](#)). Phenotyping algorithm development was not always the primary objective for the identified studies; sometimes, it was part of a larger process. The most commonly occurring objectives of the algorithms were (1) phenotyping algorithm development (193/299, 64.5%), (2) epidemiological analysis (70/299, 23.4%), and (3) predictive modeling (19/299, 6.4%). Other objectives included designing clinical decision support and implementation tools, genome analysis, and registry development. These objectives reflect the health system delivery and clinical practice contexts in which the studies were situated.

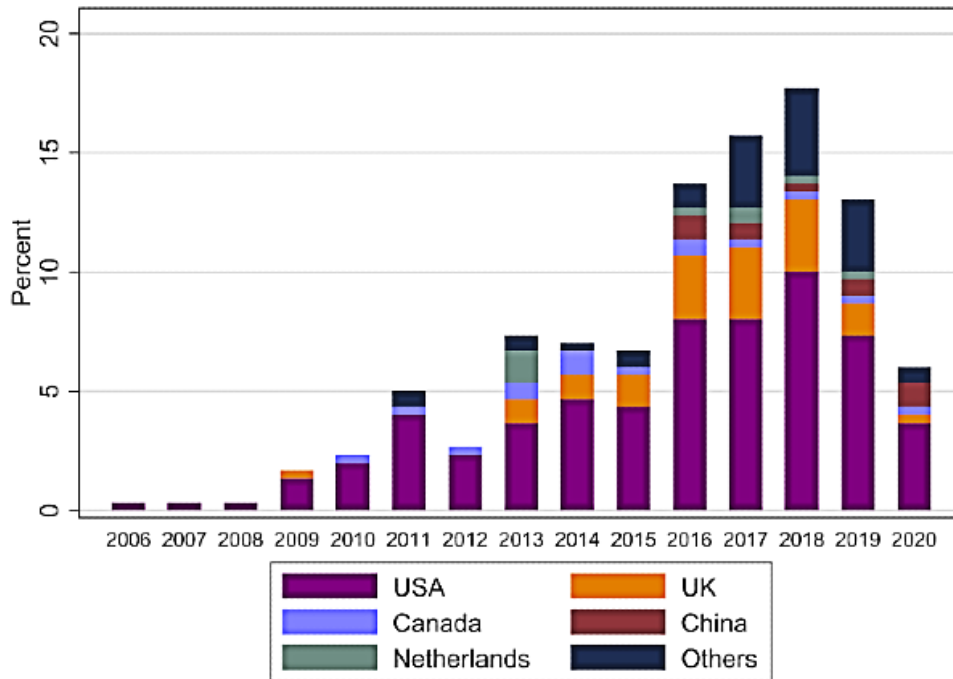
**Table 1.** Descriptive summary of the 299 Charlson algorithms.

Disease	Algorithm count	EMR <sup>a</sup> settings				Country
		Inpatient	Inpatient and outpatient	Outpatient	Other	
Myocardial infarction [16-38]	23	16	3	4	0	<ul style="list-style-type: none"> <li>• 16 United States</li> <li>• 2 United Kingdom</li> <li>• 5 Others</li> </ul>
Congestive heart failure [19,39-75]	38	22	1	14	1	<ul style="list-style-type: none"> <li>• 27 United States</li> <li>• 3 Sweden</li> <li>• 8 Others</li> </ul>
Peripheral vascular disease [19,76-89]	15	6	1	7	1	<ul style="list-style-type: none"> <li>• 9 United States</li> <li>• 5 United Kingdom</li> <li>• 1 Norway</li> </ul>
Cerebrovascular disease [19,47,57,78,83,90-107]	23	14	0	9	0	<ul style="list-style-type: none"> <li>• 10 United States</li> <li>• 6 United Kingdom</li> <li>• 7 Others</li> </ul>
Hemiplegia and paraplegia	0	0	0	0	0	0
Dementia [19,84,108-130]	25	13	1	10	1	<ul style="list-style-type: none"> <li>• 8 United States</li> <li>• 2 United Kingdom</li> <li>• 1 Netherlands</li> <li>• 1 Canada</li> <li>• 13 Others</li> </ul>
Chronic pulmonary disease [129,131-160]	31	14	1	16	0	<ul style="list-style-type: none"> <li>• 16 United States</li> <li>• 6 United Kingdom</li> <li>• 4 Canada</li> <li>• 5 Others</li> </ul>
Rheumatologic disease [161-185]	25	15	1	9	0	<ul style="list-style-type: none"> <li>• 15 United States</li> <li>• 7 United Kingdom</li> <li>• 3 Others</li> </ul>
Peptic ulcer disease [186-189]	4	3	0	1	0	<ul style="list-style-type: none"> <li>• 3 United States</li> <li>• 1 Singapore</li> </ul>
Diabetes [19,28,34,47,48,84,128,129,140,150,166,190-234]	56	30	6	20	0	<ul style="list-style-type: none"> <li>• 31 United States</li> <li>• 8 Canada</li> <li>• 4 United Kingdom</li> <li>• 13 Others</li> </ul>
Diabetes, with complications [57,235-242]	9	6	1	2	0	<ul style="list-style-type: none"> <li>• 5 United States</li> <li>• 2 United Kingdom</li> <li>• 1 Israel</li> <li>• 1 China</li> </ul>
Renal disease [47,57,243-262]	22	9	3	9	1	<ul style="list-style-type: none"> <li>• 16 United States</li> <li>• 2 United Kingdom</li> <li>• 2 Spain</li> <li>• 2 Others</li> </ul>
Mild liver disease [189,263-276]	15	11	3	0	1	<ul style="list-style-type: none"> <li>• 11 United States</li> <li>• 2 China</li> <li>• 1 Australia</li> <li>• 1 United Kingdom</li> </ul>
Moderate/severe liver disease [244,275-280]	7	5	0	2	0	<ul style="list-style-type: none"> <li>• 4 United States</li> <li>• 1 United Kingdom</li> <li>• 1 Netherlands</li> <li>• 1 China</li> </ul>

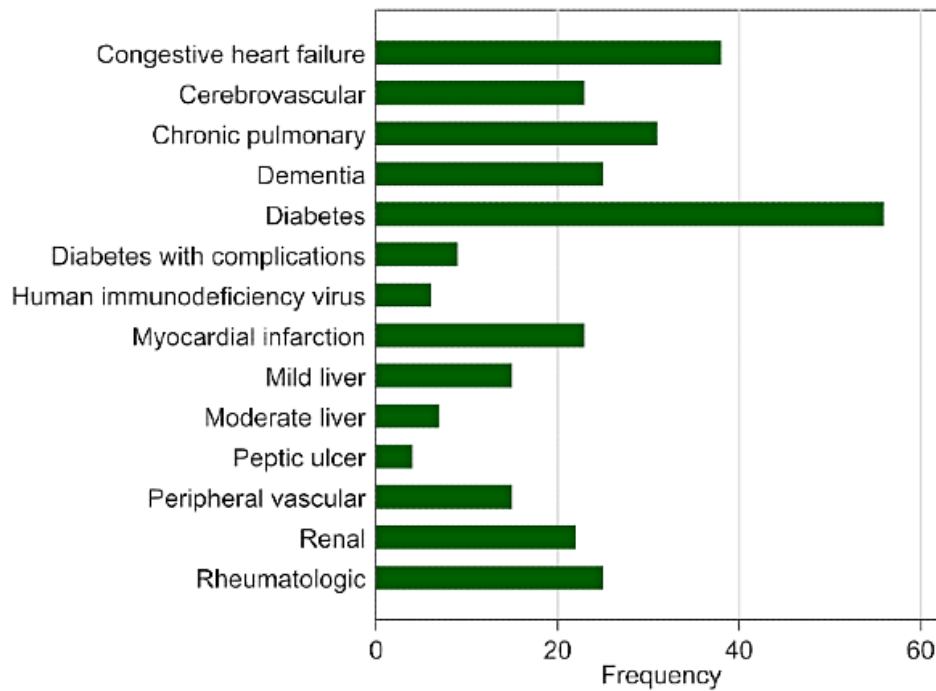
Disease	Algorithm count	EMR <sup>a</sup> settings				Country
		Inpatient	Inpatient and outpatient	Outpatient	Other	
HIV [137,281-285]	6	4	2	0	0	6 United States

<sup>a</sup>EMR: electronic medical record.

**Figure 1.** Distribution of published articles by country between January 2000 and April 2020.



**Figure 2.** Distribution of electronic medical record data-based algorithms by Charlson disease area.



**Table 2.** The Charlson algorithm types identified in this scoping review.

Charlson condition	Algorithm type						
	Codes only	Codes and structured data	Codes and free-text data	Codes, structured, and free-text data	Structured data only	Free-text data only	Free-text and structured data
Myocardial infarction (n=23, 7.7%)	7	10	0	4	0	2	0
Congestive heart failure (n=38, 12.7%)	7	19	2	9	0	0	1
Peripheral vascular disease (n=15, 5.0%)	1	6	1	2	2	2	1
Cerebrovascular disease (n=23, 7.7%)	6	14	0	1	1	1	0
Dementia (n=25, 8.4%)	4	8	4	4	1	1	3
Chronic pulmonary disease (n=31, 10.4%)	3	17	3	3	0	4	1
Rheumatologic disease (n=25, 8.4%)	2	9	2	12	0	0	0
Peptic ulcer disease (n=4, 1.3%)	0	1	1	2	0	0	0
Diabetes (n=56, 18.7%)	6	41	1	6	1	0	1
Diabetes with complications (n=9, 3.0%)	1	8	0	0	0	0	0
Renal disease (n=22, 7.4%)	2	15	1	2	1	0	1
Mild liver disease (n=15, 5.0%)	1	11	0	2	0	0	1
Moderate/severe liver disease (n=7, 2.3%)	0	3	0	3	0	0	1
HIV (n=6, 2.0%)	0	5	0	1	0	0	0
Combined (n=299, 100.0%)	40	167	15	51	6	10	10

### Data Elements: Structured Versus Unstructured

With regard to the EMR algorithms identified in this study, structured data most commonly consisted of demographics, diagnoses, procedures, vital signs, laboratory results, and medications. Structured data elements were the most common type of data employed by clinical rule-based algorithms and included basic demographics (eg, sex and age), medications, laboratory data, and diagnostic codes. A total of 233 out of 299 (77.9%) algorithms employed key laboratory diagnostic tests based on the present clinical practice.

These structured EMR components are typically available across EMR systems. Algorithms based on diagnostic codes and structured data were used primarily (213/299, 71.2%) for chronic conditions such as diabetes, where laboratory tests and medication may be necessary and sufficient for clinical decision-making. The use of diagnostic codes depended on the EMR setting (ie, outpatient or inpatient) and the health services jurisdiction (eg, United Kingdom vs United States vs Canada) where the work took place ([Multimedia Appendix 4](#)). Types of diagnostic codes identified included ICD-9, ICD-10, Read, Oxford Medical Information System, and International Classification of Primary Care (ICPC). ICD codes were used predominantly within inpatient settings (148/168, 88.1%). These basic structured data-based definitions were enhanced by incorporating unstructured data such as free text and imaging for designing classification algorithms ([Table 2](#)) for complicated chronic conditions. In summary, the disease context determined the data elements that were used.

Unstructured free-text data (eg, discharge summaries, consult notes, and nursing notes) were incorporated in approximately 86 out of 299 (28.8%) case phenotyping algorithms. NLP

techniques were used to analyze such unstructured free-text data. Many studies used controlled medical terminologies, such as the Unified Medical Language System [286] and the Systematized Nomenclature of Medicine Clinical Terms [287], in the processing of clinical notes. Both terminologies can be used by medical researchers. Many studies also employed custom vocabularies developed in consultation with clinicians or had clinicians manually annotate the free-text data to obtain the reference standard. Variations in the processing of the unstructured data were also noted. NLP processing programs such as clinical Text Analysis and Knowledge Extraction System [288], MedTagger [289], or in-house programs were employed using one of the terminologies mentioned above. This data processing converted unstructured free-text data into structured data. The converted data are often combined with existing structured data for phenotyping and disease prediction using a wide range of techniques in epidemiology, statistics, and ML. Cox regression modeling was used for survival analysis, along with incidence and prevalence in epidemiological studies. Supervised learning classification algorithms such as Naive Bayes, support vector machines, logistic regression, and neural networks are commonly used in the ML studies. The manually annotated notes or reference standard obtained from the chart review provided labels for supervised ML.

### Disease Context

Case phenotyping algorithms exhibited 2 distinct types of approaches: clinician-derived rule-based (ie, expert-driven) and data-driven approaches. Clinician-derived rule-based approaches for defining cases were based on clinical criteria dictated by guidelines or clinical practice. These rule-based methods are generally easy to interpret and are accepted as clinically relevant. However, criteria were inconsistent within and across multiple

diseases even for the clinical rule-based case phenotyping, implying that the interpretation of algorithm results may depend on choices made during the algorithm development process [290]. Despite these variations, common structured data elements were identified in each disease discipline within each context of patient care. In contrast, data-driven approaches to defining cases use information extracted from available data to determine the disease status of the patient, often with improved performance (eg, sensitivity, positive predictive value [PPV], and F1 score) compared with baseline rule-based algorithms. For example, feeding all available free-text and laboratory data for congestive heart failure (CHF) into a prediction model can classify the CHF status [73]. One study employed principal component analysis [34]. However, the association between the predictor variables and outcomes is often difficult to ascertain, and the model may be difficult to interpret.

The algorithms used various EMR data elements depending on the clinical disease context. For each disease area, unique diagnostic methods or clinical data elements were observed. Diabetes was the most commonly identified disease in our literature search (56/299, 18.7%) and will be used as an example. Case phenotyping for diabetes had fewer data element variations compared with other diseases, and algorithms involved hemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>), glucose levels, and fasting glucose as key laboratory tests and antidiabetic medications. Most diabetes algorithms did not define the severity of the disease but classified the conditions in terms of the presence or absence of type 1 or type 2 diabetes. Diabetes phenotyping studies designed patient cohort selection taking this into consideration. Developing phenotypes for identifying severe complications of diabetes did require additional data (ie, clinical narratives) and advanced methodological approaches (eg, NLP and ML), as structured data alone would not readily identify these unless diagnostic codes were included for such complications. EMR phenotypes for disease severity were sometimes developed, in the case of chronic conditions that have a widely accepted clinical severity definition. Using chronic kidney disease as an example, severity was defined according to the Kidney Disease Improving Global Outcomes [291] and the National Kidney Foundation [292] guidelines based on estimated glomerular filtration rate.

### Data Linkage

A subset of phenotyping algorithms (30/299, 10.0%) linked EMR data to disease registries or genomics data. A total of 24 out of 299 (8.0%) algorithms linked clinical and health administrative databases. All data linkage occurred in studies that used diagnostic codes. The most commonly occurring diagnostic codes were ICD-9 and ICD-10, with some regional or national diagnostic codes (eg, Read codes among UK studies). The EMR administrative data linkage context appeared mostly within primary care data-based algorithms (14/24). The UK Clinical Practice Research Datalink was linked with Hospital Episode Statistics and other administrative data to primary care EMR. The most commonly linked inpatient care data came from the Electronic Medical Records and Genomics (eMERGE) consortium [293], which provided additional validation between clinical documentation and scientific (ie, genomic) observation. These data linkage studies were employed for epidemiological

analyses (improved accuracy of incidence and prevalence estimates) of diseases at the population level [83,96,212].

### Validity of Phenotyping Algorithms

Studies varied in their reporting metrics for the validity of case definition algorithms. Commonly reported metrics were sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and F1 score. A total of 185 algorithms (185/299, 62.1%) employed chart review as the reference standard to calculate some of the aforementioned validation metrics. Of these 185 algorithms, 9 employed ML, 39 employed NLP, and 17 employed both ML and NLP. Of the 114 algorithms that did not conduct a chart review, 17 incorporated ML, 14 incorporated NLP, and 7 employed both ML and NLP techniques. Including free-text data as a data source in phenotyping algorithms tended to yield higher performance, with an average sensitivity of 0.906 (SD 0.110) and PPV of 0.913 (SD 0.120) when compared with studies that did not use free-text or ML (average sensitivity of 0.825 (SD 0.214) and average PPV of 0.853 (SD 0.174)). Incorporation of ML as part of the data-driven phenotyping also led to similar performance in sensitivity but weaker PPV, with an average sensitivity of 0.832 (SD 0.095) and average PPV of 0.633 (SD 0.358). In total, 59 out of 166 (35.5%) inpatient algorithms employed NLP, whereas 10 out of 93 (10.8%) primary care algorithms employed NLP. Among the works that used NLP, terminology standards were based on either Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) or Unified Medical Language System (UMLS), although many developed their own in-house keywords. Coding standards within inpatients were based on either ICD-9 or ICD-10 depending on the timing of the study and the jurisdictions where each study took place. Similarly, primary care code standards also varied. For example, mostly Read or ICPC codes were used in the United Kingdom, whereas ICD codes were used in North America (United States and Canada). The additional data provide a specific list of ML techniques that were used in each study, if employed ([Multimedia Appendix 4](#)).

## Discussion

### EMR Phenotyping and Precision Medicine

Achieving precision medicine requires the right information to be delivered to the right personnel at the right time. Developing EMR data-based phenotypes and integrating them into existing health information systems is a pivotal step for building a learning health system. EMR phenotypes allow rapid detection of diseases and accelerate the delivery of information to clinicians who may need it to make informed clinical decisions, policymakers who may use them to obtain population information for making public health decisions, and health services organizations that may need such information for planning clinical operations or developing risk adjustment models for patient safety programs. The purposes of the case definitions identified in this review were largely achieving one of the stated objectives above.

EMR-based phenotype and algorithm development reflected the structure and data available within respective health systems. Diagnostic codes, such as ICD and present procedural

terminology codes, are often used for billing purposes within inpatient and outpatient (ie, primary care) settings in certain countries (eg, the United States). These codes were also built into EMR systems (eg, problem lists). Consequently, these diagnostic codes were used extensively in algorithm development with the assumption that billing and problem list practices accurately reflect the provided care. In jurisdictional settings where ICD-based billing was not recorded directly in the EMR system during patient care (eg, inpatient care in Alberta), such assumptions could not be made and influenced the algorithm development process. Recognizing similarities and differences in data collection strategies, extraction, data release protocols, and existing clinical pathways is critical and will inform algorithm development strategies. ML and NLP techniques are increasingly being adopted in phenotyping algorithms. This is a testament to the fact that detailed records, available from free-text data, can assist with building high-performance classification algorithms.

### Data Extraction, Validity, and Quality

Developing data-driven case finding algorithms is not feasible without electronic data [294]. However, EMR data are not always easy to work with [295], as they are primarily intended to support clinical practice rather than research. EMR settings influence data collection and extraction strategies. Inpatient facilities often set up electronic data warehouses where EMR data are collected into centralized repositories, including free-text data. Primary care settings, in contrast, have variations in their systems, and studies based on primary care data often only use more common data elements such as laboratory data and demographics for multisite studies. Free-text data are less available when compared with inpatient facilities. Primary care clinics, including specialist clinics, are privately operated in many jurisdictions, whereas inpatient care may be publicly or privately operated. These different entities may not always be required to share health data or may have different data management protocols. These considerations influenced the algorithm development process, and a stark contrast in the used data elements can be observed between algorithms developed in outpatient and inpatient settings. To mitigate some of these issues, researchers conducted data linkage between data sources to expand the scope of the available data.

In addition, significant changes in the terminology and coding standards and practices in EMRs have occurred and are actively occurring. This often makes it difficult or impossible to compare or share algorithms developed for different EMR systems using different coding standards (eg, ICD-9, ICD-10, Read, SNOMED RT, SNOMED CT, and MEDCIN for diagnostic codes). Furthermore, many investigators noted that their studies were based on data from a single center, as they did not have access to external EMR data outside of their own institution. Thus, the potential lack of generalizability was a limitation for some studies. However, algorithms developed using commonly available data elements were often externally validated in multiple studies. In particular, simpler algorithms involving diagnostic codes or laboratory data appeared to be externally validated more commonly. This trend was observed in diabetes and rheumatic conditions and occurred mostly in the United States.

Variation in reported metrics (eg, sensitivity, specificity, positive predictive value, negative predictive value, area under the receiver-operator characteristic curve, and F1 score) was observed in the identified literature. Standardized metrics used in health care should be reported, including sensitivity, specificity, positive predictive value, and negative predictive value. As there is a trade-off between sensitivity and positive predictive value and both are important, it is also useful to report the F1 score, which is the harmonic mean of these 2 quantities. In addition, as class imbalance is frequently a problem in the context of disease classification, with positive instances far less common than negative instances, studies are encouraged to report metrics that account for this, such as area under the precision-recall curve [296]. At present, there are no universally accepted EMR data quality assessment metrics available, although there are various proposed data quality assessment frameworks [297]. Data quality must be assessed based on the suitability of the data to achieve a specific research objective or downstream task. We discuss this later in the recommendations.

### Limitations

This study is not without limitations. First, it is possible that our search did not encompass all qualifying articles in the field. However, our search strategy was refined and improved by systematic review search experts and librarians, and we believe our search successfully captured a broad spectrum of articles on the Charlson conditions. Second, manual screening was carried out by one individual. The objectivity of the review may have been increased by including a second reviewer. Finally, our review did not discuss methods employed for assessing EMR data quality, which depends on the context and clinical application, and is a difficult concept to measure in general. To date, there is no universally accepted data quality metric developed for EMR data, and few of the papers in this review discuss whether or how data quality was assessed in their study. Further research is required to establish the scope of practice for EMR data quality assessment.

### Recommendations on the Basis of Findings

Our review identified that case phenotyping algorithms depends on the health delivery system and disease context. We present a few observed strategies to assist with refining phenotype case definitions using the following key strategies: (1) understanding the health system structure and setting (eg, outpatient vs inpatient, coding practice) will provide a general sense of the type of EMR data that may be available; (2) considering data linkage can increase the scope of data available for algorithm development, it is important to recognize that data may not be standardized or comparable between different data sources. Additional data processing such as data recoding or data imputation may be needed; (3) identifying the relevant clinical and/or health services pathway and involving respective specialty physicians and other stakeholders as part of the algorithm development process can assist with knowledge translation; (4) employing a common data model (eg, observational medical outcomes partnership [298]) and using commonly available data elements to the possible extent can encourage widespread deployment and external validation. A



common data model may differ between disease disciplines and health system areas; and (5) considering how to customize the algorithm to the needs of the end user. The needs are largely divided into clinical decision support through risk adjustment analysis, population-scale disease identification for public health initiatives, or developing methodologies to improve algorithm performance.

Health care is a unique environment, and a one-size-fits-all approach may not be appropriate. This review identified variations in EMR phenotyping, which were heavily influenced by the health care delivery setting and the disease context. To optimize performance, researchers should develop tailored algorithms that focus on the specific population of interest and the particular structure of the health system (eg, developing a primary care diabetes definition), while accounting for data issues such as variations in coding systems, clinical practice guidelines, and data quality. Once a locally developed algorithm is in place, health systems may consider implementing their case finding algorithms on standardized data models. This review identified several studies that either validated previously validated case definitions in a new setting or were refined to appropriately identify disease patients within a new setting. Having locally developed algorithms converted to standard data models will facilitate external validation and implementation, which can otherwise be a critical roadblock to the adoption of these algorithms, allowing for improved algorithm interoperability between health care systems.

The interoperability of algorithms across systems facilitates implementation within existing real-time clinical decision support systems. Easy access to developed code is also critical in validating and replicating published algorithms, after their

computability has been confirmed. Analytical code and resources could be shared publicly (eg, on GitHub) to allow access for validation and implementation. The eMERGE consortium [293], CALIBER [299], and Canadian Primary Care Sentinel Surveillance Network [300], for example, have made their algorithms publicly available and have been widely adopted.

## Conclusions

We assessed EMR-based phenotyping of the Charlson conditions in health care settings. The phenotyping algorithms were locally developed and tailored to the needs and objectives of the individual studies. The health system structure and disease context determined data availability and type. The disease context dictated the common data types used for algorithm development. NLP with free-text data was employed for complex diseases that were difficult to identify with algorithms using readily available structured data. Supervised ML was employed in phenotyping algorithms, where applicable, which worked with reference standards obtained from medical chart review. Studies are encouraged to report standard health system metrics and metrics that account for class imbalance. Locally developed algorithms were validated or refined for adoption in the new setting. Locally developed disease- and setting-specific algorithms could be translated into a common data model for easier interoperability of algorithms across systems. Integrating EMR phenotyping algorithms within a health system could lead to the development of a clinical decision support system that makes use of refined existing risk adjustment scoring for risk stratification in clinical point-of-care and inform the public health and health system decision-making process, thus, leading to learning health systems.

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Developed search terms (Medical Subject Headings) for scoping literature review.  
[\[DOCX File , 17 KB-Multimedia Appendix 1\]](#)

---

## Multimedia Appendix 2

Embase and Medical Literature Analysis and Retrieval System Online search results of Charlson terms.  
[\[DOCX File , 22 KB-Multimedia Appendix 2\]](#)

---

## Multimedia Appendix 3

Preferred Reporting Items for Systematic reviews and Meta-analyses flow diagram.  
[\[PNG File , 171 KB-Multimedia Appendix 3\]](#)

---

## Multimedia Appendix 4

Summary spreadsheet of identified articles between January 2000 and April 2020.  
[\[XLSX File \(Microsoft Excel File\), 208 KB-Multimedia Appendix 4\]](#)

---

## References

1. Jameson JL, Longo DL. Precision medicine--personalized, problematic, and promising. *N Engl J Med* 2015 Jun 04;372(23):2229-2234. [doi: [10.1056/NEJMs1503104](https://doi.org/10.1056/NEJMs1503104)] [Medline: [26014593](https://pubmed.ncbi.nlm.nih.gov/26014593/)]

2. Adler-Milstein J, Jha AK. HITECH Act Drove Large Gains In Hospital Electronic Health Record Adoption. *Health Aff (Millwood)* 2017 Aug 01;36(8):1416-1422. [doi: [10.1377/hlthaff.2016.1651](https://doi.org/10.1377/hlthaff.2016.1651)] [Medline: [28784734](https://pubmed.ncbi.nlm.nih.gov/28784734/)]
3. Gagnon M, Payne-Gagnon J, Breton E, Fortin J, Khoury L, Dolovich L, et al. Adoption of Electronic Personal Health Records in Canada: Perceptions of Stakeholders. *Int J Health Policy Manag* 2016 Jul 01;5(7):425-433 [FREE Full text] [doi: [10.15171/ijhpm.2016.36](https://doi.org/10.15171/ijhpm.2016.36)] [Medline: [27694670](https://pubmed.ncbi.nlm.nih.gov/27694670/)]
4. Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M. Defining comorbidity: implications for understanding health and health services. *Ann Fam Med* 2009;7(4):357-363 [FREE Full text] [doi: [10.1370/afm.983](https://doi.org/10.1370/afm.983)] [Medline: [19597174](https://pubmed.ncbi.nlm.nih.gov/19597174/)]
5. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40(5):373-383. [Medline: [3558716](https://pubmed.ncbi.nlm.nih.gov/3558716/)]
6. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol* 2011 Mar 15;173(6):676-682. [doi: [10.1093/aje/kwq433](https://doi.org/10.1093/aje/kwq433)] [Medline: [21330339](https://pubmed.ncbi.nlm.nih.gov/21330339/)]
7. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005 Nov;43(11):1130-1139. [Medline: [16224307](https://pubmed.ncbi.nlm.nih.gov/16224307/)]
8. Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol* 1993 Oct;46(10):1075-9; discussion 1081. [Medline: [8410092](https://pubmed.ncbi.nlm.nih.gov/8410092/)]
9. McBrien KA, Souri S, Symonds NE, Rouhi A, Lethebe BC, Williamson TS, et al. Identification of validated case definitions for medical conditions used in primary care electronic medical record databases: a systematic review. *J Am Med Inform Assoc* 2018 Nov 01;25(11):1567-1578. [doi: [10.1093/jamia/ocy094](https://doi.org/10.1093/jamia/ocy094)] [Medline: [30137498](https://pubmed.ncbi.nlm.nih.gov/30137498/)]
10. Nissen F, Quint JK, Wilkinson S, Mullerova H, Smeeth L, Douglas IJ. Validation of asthma recording in electronic health records: a systematic review. *Clin Epidemiol* 2017;9:643-656 [FREE Full text] [doi: [10.2147/CLEP.S143718](https://doi.org/10.2147/CLEP.S143718)] [Medline: [29238227](https://pubmed.ncbi.nlm.nih.gov/29238227/)]
11. Wang P, Garza M, Zozus M. Cancer Phenotype Development: A Literature Review. *Stud Health Technol Inform* 2019;257:468-472. [Medline: [30741241](https://pubmed.ncbi.nlm.nih.gov/30741241/)]
12. Xu J, Rasmussen LV, Shaw PL, Jiang G, Kiefer RC, Mo H, et al. Review and evaluation of electronic health records-driven phenotype algorithm authoring tools for clinical and translational research. *J Am Med Inform Assoc* 2015 Dec;22(6):1251-1260 [FREE Full text] [doi: [10.1093/jamia/ocv070](https://doi.org/10.1093/jamia/ocv070)] [Medline: [26224336](https://pubmed.ncbi.nlm.nih.gov/26224336/)]
13. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med* 2018 Oct 02;169(7):467-473. [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
14. EndNote Version X8. <https://endnote.com>. Philadelphia, PA: Clarivate; 2013.
15. STATA SSR14. College Station, TX: StataCorp LLC; 2015. URL: <https://www.stata.com>
16. Ammann EM, Schweizer ML, Robinson JG, Eschol JO, Kafa R, Girotra S, et al. Chart validation of inpatient ICD-9-CM administrative diagnosis codes for acute myocardial infarction (AMI) among intravenous immune globulin (IGIV) users in the Sentinel Distributed Database. *Pharmacoepidemiol Drug Saf* 2018 Apr;27(4):398-404 [FREE Full text] [doi: [10.1002/pds.4398](https://doi.org/10.1002/pds.4398)] [Medline: [29446185](https://pubmed.ncbi.nlm.nih.gov/29446185/)]
17. Ando T, Ooba N, Mochizuki M, Koide D, Kimura K, Lee SL, et al. Positive predictive value of ICD-10 codes for acute myocardial infarction in Japan: a validation study at a single center. *BMC Health Serv Res* 2018 Dec 26;18(1):895 [FREE Full text] [doi: [10.1186/s12913-018-3727-0](https://doi.org/10.1186/s12913-018-3727-0)] [Medline: [30477501](https://pubmed.ncbi.nlm.nih.gov/30477501/)]
18. Backenroth D, Chase H, Friedman C, Wei Y. Using Rich Data on Comorbidities in Case-Control Study Design with Electronic Health Record Data Improves Control of Confounding in the Detection of Adverse Drug Reactions. *PLoS One* 2016;11(10):e0164304 [FREE Full text] [doi: [10.1371/journal.pone.0164304](https://doi.org/10.1371/journal.pone.0164304)] [Medline: [27716785](https://pubmed.ncbi.nlm.nih.gov/27716785/)]
19. Bent-Ennakhil N, Cécile Périer M, Sobocki P, Gothefors D, Johansson G, Milea D, et al. Incidence of cardiovascular diseases and type-2-diabetes mellitus in patients with psychiatric disorders. *Nord J Psychiatry* 2018 Oct;72(7):455-461. [doi: [10.1080/08039488.2018.1463392](https://doi.org/10.1080/08039488.2018.1463392)] [Medline: [30513230](https://pubmed.ncbi.nlm.nih.gov/30513230/)]
20. Bjerking LH, Hansen KW, Madsen M, Jensen JS, Madsen JK, Sørensen R, et al. Use of diagnostic coronary angiography in women and men presenting with acute myocardial infarction: a matched cohort study. *BMC Cardiovasc Disord* 2016 Jun 01;16:120 [FREE Full text] [doi: [10.1186/s12872-016-0248-9](https://doi.org/10.1186/s12872-016-0248-9)] [Medline: [27250115](https://pubmed.ncbi.nlm.nih.gov/27250115/)]
21. Coloma PM, Valkhoff VE, Mazzaglia G, Nielsson MS, Pedersen L, Molokhia M, EU-ADR Consortium. Identification of acute myocardial infarction from electronic healthcare records using different disease coding systems: a validation study in three European countries. *BMJ Open* 2013 Jun 20;3(6) [FREE Full text] [doi: [10.1136/bmjopen-2013-002862](https://doi.org/10.1136/bmjopen-2013-002862)] [Medline: [23794587](https://pubmed.ncbi.nlm.nih.gov/23794587/)]
22. Cross DS, McCarty CA, Steinhubl SR, Carey DJ, Erlich PM. Development of a multi-institutional cohort to facilitate cardiovascular disease biomarker validation using existing biorepository samples linked to electronic health records. *Clin Cardiol* 2013 Aug;36(8):486-491 [FREE Full text] [doi: [10.1002/clc.22146](https://doi.org/10.1002/clc.22146)] [Medline: [23740530](https://pubmed.ncbi.nlm.nih.gov/23740530/)]
23. Findlay I, Morris T, Zhang R, McCowan C, Shield S, Forbes B, et al. Linking hospital patient records for suspected or established acute coronary syndrome in a complex secondary care system: a proof-of-concept e-registry in National Health Service Scotland. *Eur Heart J Qual Care Clin Outcomes* 2018 Jul 01;4(3):155-167. [doi: [10.1093/ehjqcco/qcy007](https://doi.org/10.1093/ehjqcco/qcy007)] [Medline: [29462281](https://pubmed.ncbi.nlm.nih.gov/29462281/)]

24. FitzHenry F, Murff HJ, Matheny ME, Gentry N, Fielstein EM, Brown SH, et al. Exploring the frontier of electronic health record surveillance: the case of postoperative complications. *Med Care* 2013 Jul;51(6):509-516 [[FREE Full text](#)] [doi: [10.1097/MLR.0b013e31828d1210](https://doi.org/10.1097/MLR.0b013e31828d1210)] [Medline: [23673394](#)]
25. Floyd JS, Blondon M, Moore KP, Boyko EJ, Smith NL. Validation of methods for assessing cardiovascular disease using electronic health data in a cohort of Veterans with diabetes. *Pharmacoepidemiol Drug Saf* 2016 May;25(4):467-471 [[FREE Full text](#)] [doi: [10.1002/pds.3921](https://doi.org/10.1002/pds.3921)] [Medline: [26555025](#)]
26. Goldstein BA, Assimes T, Winkelmayr WC, Hastie T. Detecting clinically meaningful biomarkers with repeated measurements: An illustration with electronic health records. *Biometrics* 2015 Jul;71(2):478-486 [[FREE Full text](#)] [doi: [10.1111/biom.12283](https://doi.org/10.1111/biom.12283)] [Medline: [25652566](#)]
27. Herrett E, Bhaskaran K, Timmis A, Denaxas S, Hemingway H, Smeeth L. Association between clinical presentations before myocardial infarction and coronary mortality: a prospective population-based study using linked electronic records. *Eur Heart J* 2014 Oct 14;35(35):2363-2371 [[FREE Full text](#)] [doi: [10.1093/eurheartj/ehu286](https://doi.org/10.1093/eurheartj/ehu286)] [Medline: [25038774](#)]
28. Hivert M, Grant RW, Shrader P, Meigs JB. Identifying primary care patients at risk for future diabetes and cardiovascular disease using electronic health records. *BMC Health Serv Res* 2009 Sep 22;9:170 [[FREE Full text](#)] [doi: [10.1186/1472-6963-9-170](https://doi.org/10.1186/1472-6963-9-170)] [Medline: [19772639](#)]
29. Mahler SA, Lenoir KM, Wells BJ, Burke GL, Duncan PW, Case LD, et al. Safely Identifying Emergency Department Patients With Acute Chest Pain for Early Discharge. *Circulation* 2018 Nov 27;138(22):2456-2468 [[FREE Full text](#)] [doi: [10.1161/CIRCULATIONAHA.118.036528](https://doi.org/10.1161/CIRCULATIONAHA.118.036528)] [Medline: [30571347](#)]
30. Manemann SM, Gerber Y, Chamberlain AM, Dunlay SM, Bell MR, Jaffe AS, et al. Acute coronary syndromes in the community. *Mayo Clin Proc* 2015 May;90(5):597-605. [doi: [10.1016/j.mayocp.2015.02.015](https://doi.org/10.1016/j.mayocp.2015.02.015)] [Medline: [25794453](#)]
31. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA* 2011 Aug 24;306(8):848-855. [doi: [10.1001/jama.2011.1204](https://doi.org/10.1001/jama.2011.1204)] [Medline: [21862746](#)]
32. Persell SD, Dunne AP, Lloyd-Jones DM, Baker DW. Electronic health record-based cardiac risk assessment and identification of unmet preventive needs. *Med Care* 2009 May;47(4):418-424. [doi: [10.1097/MLR.0b013e31818dce21](https://doi.org/10.1097/MLR.0b013e31818dce21)] [Medline: [19238100](#)]
33. Reynolds K, Go AS, Leong TK, Boudreau DM, Cassidy-Bushrow AE, Fortmann SP, et al. Trends in Incidence of Hospitalized Acute Myocardial Infarction in the Cardiovascular Research Network (CVRN). *Am J Med* 2017 Mar;130(3):317-327 [[FREE Full text](#)] [doi: [10.1016/j.amjmed.2016.09.014](https://doi.org/10.1016/j.amjmed.2016.09.014)] [Medline: [27751900](#)]
34. Song W, Huang H, Zhang CZ, Bates DW, Wright A. Using whole genome scores to compare three clinical phenotyping methods in complex diseases. *Sci Rep* 2018 Jul 27;8(1):11360 [[FREE Full text](#)] [doi: [10.1038/s41598-018-29634-w](https://doi.org/10.1038/s41598-018-29634-w)] [Medline: [30054501](#)]
35. Tien M, Kashyap R, Wilson GA, Hernandez-Torres V, Jacob AK, Schroeder DR, et al. Retrospective Derivation and Validation of an Automated Electronic Search Algorithm to Identify Post Operative Cardiovascular and Thromboembolic Complications. *Appl Clin Inform* 2015;6(3):565-576 [[FREE Full text](#)] [doi: [10.4338/ACI-2015-03-RA-0026](https://doi.org/10.4338/ACI-2015-03-RA-0026)] [Medline: [26448798](#)]
36. Torabi A, Cleland JGF, Sherwi N, Atkin P, Panahi H, Kilpatrick E, et al. Influence of case definition on incidence and outcome of acute coronary syndromes. *Open Heart* 2016;3(2):e000487 [[FREE Full text](#)] [doi: [10.1136/openhrt-2016-000487](https://doi.org/10.1136/openhrt-2016-000487)] [Medline: [28123755](#)]
37. Wang N, Li T, Du Q. Risk factors of upper gastrointestinal hemorrhage with acute coronary syndrome. *Am J Emerg Med* 2019 Apr;37(4):615-619. [doi: [10.1016/j.ajem.2018.06.055](https://doi.org/10.1016/j.ajem.2018.06.055)] [Medline: [30381146](#)]
38. Zheng J, Yarzebski J, Ramesh BP, Goldberg RJ, Yu H. Automatically Detecting Acute Myocardial Infarction Events from EHR Text: A Preliminary Study. *AMIA Annu Symp Proc* 2014;2014:1286-1293 [[FREE Full text](#)] [Medline: [25954440](#)]
39. Bielinski SJ, Pathak J, Carrell DS, Takahashi PY, Olson JE, Larson NB, et al. A Robust e-Epidemiology Tool in Phenotyping Heart Failure with Differentiation for Preserved and Reduced Ejection Fraction: the Electronic Medical Records and Genomics (eMERGE) Network. *J Cardiovasc Transl Res* 2015 Dec;8(8):475-483 [[FREE Full text](#)] [doi: [10.1007/s12265-015-9644-2](https://doi.org/10.1007/s12265-015-9644-2)] [Medline: [26195183](#)]
40. Blecker S, Katz SD, Horwitz LI, Kuperman G, Park H, Gold A, et al. Comparison of Approaches for Heart Failure Case Identification From Electronic Health Record Data. *JAMA Cardiol* 2016 Dec 01;1(9):1014-1020 [[FREE Full text](#)] [doi: [10.1001/jamacardio.2016.3236](https://doi.org/10.1001/jamacardio.2016.3236)] [Medline: [27706470](#)]
41. Bosch L, Assmann P, de Grauw WJC, Schalk BWM, Biermans MCJ. Heart failure in primary care: prevalence related to age and comorbidity. *Prim Health Care Res Dev* 2019 Jul 29;20:e79 [[FREE Full text](#)] [doi: [10.1017/S1463423618000889](https://doi.org/10.1017/S1463423618000889)] [Medline: [31868152](#)]
42. Bosco-Lévy P, Duret S, Picard F, Dos Santos P, Puymirat E, Gilleron V, et al. Diagnostic accuracy of the International Classification of Diseases, Tenth Revision, codes of heart failure in an administrative database. *Pharmacoepidemiol Drug Saf* 2019 Feb;28(2):194-200. [doi: [10.1002/pds.4690](https://doi.org/10.1002/pds.4690)] [Medline: [30395375](#)]
43. Byrd RJ, Steinhubl SR, Sun J, Ebadollahi S, Stewart WF. Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records. *Int J Med Inform* 2014 Dec;83(12):983-992 [[FREE Full text](#)] [doi: [10.1016/j.ijmedinf.2012.12.005](https://doi.org/10.1016/j.ijmedinf.2012.12.005)] [Medline: [23317809](#)]

44. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. *J Am Med Inform Assoc* 2017 Dec 01;24(2):361-370 [FREE Full text] [doi: [10.1093/jamia/ocw112](https://doi.org/10.1093/jamia/ocw112)] [Medline: [27521897](https://pubmed.ncbi.nlm.nih.gov/27521897/)]
45. Dai W, Brisimi TS, Adams WG, Mela T, Saligrama V, Paschalidis IC. Prediction of hospitalization due to heart diseases by supervised learning methods. *Int J Med Inform* 2015 Mar;84(3):189-197 [FREE Full text] [doi: [10.1016/j.ijmedinf.2014.10.002](https://doi.org/10.1016/j.ijmedinf.2014.10.002)] [Medline: [25497295](https://pubmed.ncbi.nlm.nih.gov/25497295/)]
46. Evans RS, Benuzillo J, Horne BD, Lloyd JF, Bradshaw A, Budge D, et al. Automated identification and predictive tools to help identify high-risk heart failure patients: pilot evaluation. *J Am Med Inform Assoc* 2016 Sep;23(5):872-878. [doi: [10.1093/jamia/ocv197](https://doi.org/10.1093/jamia/ocv197)] [Medline: [26911827](https://pubmed.ncbi.nlm.nih.gov/26911827/)]
47. Frigaard M, Rubinsky A, Lowell L, Malkina A, Karliner L, Kohn M, et al. Validating laboratory defined chronic kidney disease in the electronic health record for patients in primary care. *BMC Nephrol* 2019 Jan 03;20(1):3 [FREE Full text] [doi: [10.1186/s12882-018-1156-2](https://doi.org/10.1186/s12882-018-1156-2)] [Medline: [30606109](https://pubmed.ncbi.nlm.nih.gov/30606109/)]
48. Gini R, Schuemie MJ, Mazzaglia G, Lapi F, Francesconi P, Pasqua A, et al. Automatic identification of type 2 diabetes, hypertension, ischaemic heart disease, heart failure and their levels of severity from Italian General Practitioners' electronic medical records: a validation study. *BMJ Open* 2016 Dec 09;6(12):e012413 [FREE Full text] [doi: [10.1136/bmjopen-2016-012413](https://doi.org/10.1136/bmjopen-2016-012413)] [Medline: [27940627](https://pubmed.ncbi.nlm.nih.gov/27940627/)]
49. Huusko J, Purmonen T, Toppila I, Lassenius M, Ukkonen H. Real-world clinical diagnostics of heart failure patients with reduced or preserved ejection fraction. *ESC Heart Fail* 2020 Jul;7(3):1039-1048 [FREE Full text] [doi: [10.1002/ehf2.12665](https://doi.org/10.1002/ehf2.12665)] [Medline: [32187879](https://pubmed.ncbi.nlm.nih.gov/32187879/)]
50. Jonnalagadda SR, Adupa AK, Garg RP, Corona-Cox J, Shah SJ. Text Mining of the Electronic Health Record: An Information Extraction Approach for Automated Identification and Subphenotyping of HFpEF Patients for Clinical Trials. *J Cardiovasc Transl Res* 2017 Jul;10(3):313-321. [doi: [10.1007/s12265-017-9752-2](https://doi.org/10.1007/s12265-017-9752-2)] [Medline: [28585184](https://pubmed.ncbi.nlm.nih.gov/28585184/)]
51. Kaspar M, Fette G, Güder G, Seidlmayer L, Ertl M, Dietrich G, et al. Underestimated prevalence of heart failure in hospital inpatients: a comparison of ICD codes and discharge letter information. *Clin Res Cardiol* 2018 Oct;107(9):778-787 [FREE Full text] [doi: [10.1007/s00392-018-1245-z](https://doi.org/10.1007/s00392-018-1245-z)] [Medline: [29667017](https://pubmed.ncbi.nlm.nih.gov/29667017/)]
52. Koudstaal S, Pujades-Rodriguez M, Denaxas S, Gho JM, Shah AD, Yu N, et al. Prognostic burden of heart failure recorded in primary care, acute hospital admissions, or both: a population-based linked electronic health record cohort study in 2.1 million people. *Eur J Heart Fail* 2017 Sep;19(9):1119-1127. [doi: [10.1002/ehf.709](https://doi.org/10.1002/ehf.709)] [Medline: [28008698](https://pubmed.ncbi.nlm.nih.gov/28008698/)]
53. Kurgansky KE, Schubert P, Parker R, Djousse L, Riebman JB, Gagnon DR, et al. Association of pulse rate with outcomes in heart failure with reduced ejection fraction: a retrospective cohort study. *BMC Cardiovasc Disord* 2020 Feb 26;20(1):92 [FREE Full text] [doi: [10.1186/s12872-020-01384-6](https://doi.org/10.1186/s12872-020-01384-6)] [Medline: [32101141](https://pubmed.ncbi.nlm.nih.gov/32101141/)]
54. Lindmark K, Boman K, Olofsson M, Törnblom M, Levine A, Castelo-Branco A, et al. Epidemiology of heart failure and trends in diagnostic work-up: a retrospective, population-based cohort study in Sweden. *Clin Epidemiol* 2019;11:231-244. [doi: [10.2147/CLEP.S170873](https://doi.org/10.2147/CLEP.S170873)] [Medline: [30962724](https://pubmed.ncbi.nlm.nih.gov/30962724/)]
55. Magnusson P, Palm A, Branden E, Mörner S. Misclassification of hypertrophic cardiomyopathy: validation of diagnostic codes. *Clin Epidemiol* 2017;9:403-410 [FREE Full text] [doi: [10.2147/CLEP.S139300](https://doi.org/10.2147/CLEP.S139300)] [Medline: [28848363](https://pubmed.ncbi.nlm.nih.gov/28848363/)]
56. Panahiazar M, Taslimitehrani V, Pereira N, Pathak J. Using EHRs and Machine Learning for Heart Failure Survival Analysis. *Stud Health Technol Inform* 2015;216:40-44 [FREE Full text] [Medline: [26262006](https://pubmed.ncbi.nlm.nih.gov/26262006/)]
57. Navaneethan SD, Jolly SE, Schold JD, Arrigain S, Saupé W, Sharp J, et al. Development and validation of an electronic health record-based chronic kidney disease registry. *Clin J Am Soc Nephrol* 2011 Jan;6(1):40-49. [doi: [10.2215/CJN.04230510](https://doi.org/10.2215/CJN.04230510)] [Medline: [21051745](https://pubmed.ncbi.nlm.nih.gov/21051745/)]
58. Ng K, Steinhubl SR, deFilippi C, Dey S, Stewart WF. Early Detection of Heart Failure Using Electronic Health Records: Practical Implications for Time Before Diagnosis, Data Diversity, Data Quantity, and Data Density. *Circ Cardiovasc Qual Outcomes* 2016 Nov;9(6):649-658. [doi: [10.1161/CIRCOUTCOMES.116.002797](https://doi.org/10.1161/CIRCOUTCOMES.116.002797)] [Medline: [28263940](https://pubmed.ncbi.nlm.nih.gov/28263940/)]
59. Pakhomov S, Weston SA, Jacobsen SJ, Chute CG, Meverden R, Roger VL. Electronic medical records for clinical research: application to the identification of heart failure. *Am J Manag Care* 2007 Jun;13(6 Part 1):281-288 [FREE Full text] [Medline: [17567225](https://pubmed.ncbi.nlm.nih.gov/17567225/)]
60. Patel YR, Robbins JM, Kurgansky KE, Imran T, Orkaby AR, McLean RR, et al. Development and validation of a heart failure with preserved ejection fraction cohort using electronic medical records. *BMC Cardiovasc Disord* 2018 Jun 28;18(1):128. [doi: [10.1186/s12872-018-0866-5](https://doi.org/10.1186/s12872-018-0866-5)] [Medline: [29954337](https://pubmed.ncbi.nlm.nih.gov/29954337/)]
61. Pike MM, Decker PA, Larson NB, St Sauver JL, Takahashi PY, Roger VL, et al. Improvement in Cardiovascular Risk Prediction with Electronic Health Records. *J Cardiovasc Transl Res* 2016 Jun;9(3):214-222 [FREE Full text] [doi: [10.1007/s12265-016-9687-z](https://doi.org/10.1007/s12265-016-9687-z)] [Medline: [26960568](https://pubmed.ncbi.nlm.nih.gov/26960568/)]
62. Rasmy L, Wu Y, Wang N, Geng X, Zheng WJ, Wang F, et al. A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set. *J Biomed Inform* 2018 Aug;84:11-16 [FREE Full text] [doi: [10.1016/j.jbi.2018.06.011](https://doi.org/10.1016/j.jbi.2018.06.011)] [Medline: [29908902](https://pubmed.ncbi.nlm.nih.gov/29908902/)]
63. Shameer K, Johnson K, Yahi A. Predictive Modeling of Hospital Readmission Rates Using Electronic Medical Record-Wide Machine Learning: A Case-Study Using Mount Sinai Heart Failure Cohort. *Pacific Symposium on Biocomputing; Pacific Symposium on Biocomputing* 2016:22. [doi: [10.1142/9789813207813\\_0027](https://doi.org/10.1142/9789813207813_0027)] [Medline: [27896982](https://pubmed.ncbi.nlm.nih.gov/27896982/)]

64. Stålhammar J, Stern L, Linder R, Sherman S, Parikh R, Ariely R, et al. The burden of preserved ejection fraction heart failure in a real-world Swedish patient population. *J Med Econ* 2014 Jan;17(1):43-51. [doi: [10.3111/13696998.2013.848808](https://doi.org/10.3111/13696998.2013.848808)] [Medline: [24066839](https://pubmed.ncbi.nlm.nih.gov/24066839/)]
65. Sun J, Hu J, Luo D, Markatou M, Wang F, Edabollahi S, et al. Combining knowledge and data driven insights for identifying risk factors using electronic health records. *AMIA Annu Symp Proc* 2012;2012:901-910 [FREE Full text] [Medline: [23304365](https://pubmed.ncbi.nlm.nih.gov/23304365/)]
66. Taslimitehrani V, Dong G, Pereira NL, Panahiazar M, Pathak J. Developing EHR-driven heart failure risk prediction models using CPXR(Log) with the probabilistic loss function. *J Biomed Inform* 2016 Apr;60:260-269. [doi: [10.1016/j.jbi.2016.01.009](https://doi.org/10.1016/j.jbi.2016.01.009)] [Medline: [26844760](https://pubmed.ncbi.nlm.nih.gov/26844760/)]
67. Thomas IC, Nishimura M, Ma J, Dickson SD, Alshawabkeh L, Adler E, et al. Clinical Characteristics and Outcomes of Patients With Heart Failure and Methamphetamine Abuse. *J Card Fail* 2020 Mar;26(3):202-209. [doi: [10.1016/j.cardfail.2019.10.002](https://doi.org/10.1016/j.cardfail.2019.10.002)] [Medline: [31618697](https://pubmed.ncbi.nlm.nih.gov/31618697/)]
68. Tison GH, Chamberlain AM, Pletcher MJ, Dunlay SM, Weston SA, Killian JM, et al. Identifying heart failure using EMR-based algorithms. *Int J Med Inform* 2018 Dec;120:1-7 [FREE Full text] [doi: [10.1016/j.ijmedinf.2018.09.016](https://doi.org/10.1016/j.ijmedinf.2018.09.016)] [Medline: [30409334](https://pubmed.ncbi.nlm.nih.gov/30409334/)]
69. Vijayakrishnan R, Steinhubl SR, Ng K, Sun J, Byrd RJ, Daar Z, et al. Prevalence of heart failure signs and symptoms in a large primary care population identified through the use of text and data mining of the electronic health record. *J Card Fail* 2014 Jul;20(7):459-464 [FREE Full text] [doi: [10.1016/j.cardfail.2014.03.008](https://doi.org/10.1016/j.cardfail.2014.03.008)] [Medline: [24709663](https://pubmed.ncbi.nlm.nih.gov/24709663/)]
70. Wang Y, Luo J, Hao S, Xu H, Shin AY, Jin B, et al. NLP based congestive heart failure case finding: A prospective analysis on statewide electronic medical records. *Int J Med Inform* 2015 Dec;84(12):1039-1047. [doi: [10.1016/j.ijmedinf.2015.06.007](https://doi.org/10.1016/j.ijmedinf.2015.06.007)] [Medline: [26254876](https://pubmed.ncbi.nlm.nih.gov/26254876/)]
71. Wang Y, Ng K, Byrd R. Early detection of heart failure with varying prediction windows by structured and unstructured data in electronic health records. Conference proceedings : ;Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference 2015:2530-2533. [doi: [10.1109/embc.2015.7318907](https://doi.org/10.1109/embc.2015.7318907)] [Medline: [26736807](https://pubmed.ncbi.nlm.nih.gov/26736807/)]
72. Wu J, Roy J, Stewart WF. Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. *Med Care* 2010 Jun;48(6 Suppl):S106-S113. [doi: [10.1097/MLR.0b013e3181de9e17](https://doi.org/10.1097/MLR.0b013e3181de9e17)] [Medline: [20473190](https://pubmed.ncbi.nlm.nih.gov/20473190/)]
73. Xu Y, Lee S, Martin E, D'souza AG, Doktorchik CTA, Jiang J, et al. Enhancing ICD-Code-Based Case Definition for Heart Failure Using Electronic Medical Record Data. *J Card Fail* 2020 Jul;26(7):610-617. [doi: [10.1016/j.cardfail.2020.04.003](https://doi.org/10.1016/j.cardfail.2020.04.003)] [Medline: [32304875](https://pubmed.ncbi.nlm.nih.gov/32304875/)]
74. Yang X, Gong Y, Waheed N, March K, Bian J, Hogan WR, et al. Identifying Cancer Patients at Risk for Heart Failure Using Machine Learning Methods. *AMIA Annu Symp Proc* 2019;2019:933-941 [FREE Full text] [Medline: [32308890](https://pubmed.ncbi.nlm.nih.gov/32308890/)]
75. Zhang R, Ma S, Shanahan L, Munroe J, Horn S, Speedie S. Discovering and identifying New York heart association classification from electronic health records. *BMC Med Inform Decis Mak* 2018 Jul 23;18(Suppl 2):48. [doi: [10.1186/s12911-018-0625-7](https://doi.org/10.1186/s12911-018-0625-7)] [Medline: [30066653](https://pubmed.ncbi.nlm.nih.gov/30066653/)]
76. Afzal N, Mallipeddi VP, Sohn S, Liu H, Chaudhry R, Scott CG, et al. Natural language processing of clinical notes for identification of critical limb ischemia. *Int J Med Inform* 2018 Mar;111:83-89 [FREE Full text] [doi: [10.1016/j.ijmedinf.2017.12.024](https://doi.org/10.1016/j.ijmedinf.2017.12.024)] [Medline: [29425639](https://pubmed.ncbi.nlm.nih.gov/29425639/)]
77. Afzal N, Sohn S, Abram S, Scott CG, Chaudhry R, Liu H, et al. Mining peripheral arterial disease cases from narrative clinical notes using natural language processing. *J Vasc Surg* 2017 Jun;65(6):1753-1761 [FREE Full text] [doi: [10.1016/j.jvs.2016.11.031](https://doi.org/10.1016/j.jvs.2016.11.031)] [Medline: [28189359](https://pubmed.ncbi.nlm.nih.gov/28189359/)]
78. Archangelidi O, Pujades-Rodriguez M, Timmis A, Jouven X, Denaxas S, Hemingway H. Clinically recorded heart rate and incidence of 12 coronary, cardiac, cerebrovascular and peripheral arterial diseases in 233,970 men and women: A linked electronic health record study. *Eur J Prev Cardiol* 2018 Sep;25(14):1485-1495. [doi: [10.1177/2047487318785228](https://doi.org/10.1177/2047487318785228)] [Medline: [29966429](https://pubmed.ncbi.nlm.nih.gov/29966429/)]
79. Arruda-Olson AM, Afzal N, Priya Mallipeddi V, Said A, Moussa Pacha H, Moon S, et al. Leveraging the Electronic Health Record to Create an Automated Real-Time Prognostic Tool for Peripheral Arterial Disease. *J Am Heart Assoc* 2018 Dec 04;7(23):e009680 [FREE Full text] [doi: [10.1161/JAHA.118.009680](https://doi.org/10.1161/JAHA.118.009680)] [Medline: [30571601](https://pubmed.ncbi.nlm.nih.gov/30571601/)]
80. Caleyachetty R, Thomas GN, Toulis KA, Mohammed N, Gokhale KM, Balachandran K, et al. Metabolically Healthy Obese and Incident Cardiovascular Disease Events Among 3.5 Million Men and Women. *J Am Coll Cardiol* 2017 Oct 19;70(12):1429-1437 [FREE Full text] [doi: [10.1016/j.jacc.2017.07.763](https://doi.org/10.1016/j.jacc.2017.07.763)] [Medline: [28911506](https://pubmed.ncbi.nlm.nih.gov/28911506/)]
81. Daskivich T, Abedi G, Kaplan S. Electronic Health Record Problem Lists: Accurate Enough for Risk Adjustment? *Am J Manag Care* 2018;24(1):A. [Medline: [29350512](https://pubmed.ncbi.nlm.nih.gov/29350512/)]
82. Emdin CA, Anderson SG, Callender T, Conrad N, Salimi-Khorshidi G, Mohseni H, et al. Usual blood pressure, peripheral arterial disease, and vascular risk: cohort study of 4.2 million adults. *BMJ* 2015 Oct 29;351:h4865 [FREE Full text] [doi: [10.1136/bmj.h4865](https://doi.org/10.1136/bmj.h4865)] [Medline: [26419648](https://pubmed.ncbi.nlm.nih.gov/26419648/)]
83. George J, Rapsomaniki E, Pujades-Rodriguez M, Shah AD, Denaxas S, Herrett E, et al. How Does Cardiovascular Disease First Present in Women and Men? Incidence of 12 Cardiovascular Diseases in a Contemporary Cohort of 1,937,360 People.

- Circulation 2015 Oct 06;132(14):1320-1328 [[FREE Full text](#)] [doi: [10.1161/CIRCULATIONAHA.114.013797](https://doi.org/10.1161/CIRCULATIONAHA.114.013797)] [Medline: [26330414](#)]
84. Kho AN, Pacheco JA, Peissig PL, Rasmussen L, Newton KM, Weston N, et al. Electronic medical records for genetic research: results of the eMERGE consortium. *Sci Transl Med* 2011 Apr 20;3(79):79re1 [[FREE Full text](#)] [doi: [10.1126/scitranslmed.3001807](https://doi.org/10.1126/scitranslmed.3001807)] [Medline: [21508311](#)]
  85. Kullo IJ, Shameer K, Jouni H, Lesnick TG, Pathak J, Chute CG, et al. The ATXN2-SH2B3 locus is associated with peripheral arterial disease: an electronic medical record-based genome-wide association study. *Front Genet* 2014;5:166 [[FREE Full text](#)] [doi: [10.3389/fgene.2014.00166](https://doi.org/10.3389/fgene.2014.00166)] [Medline: [25009551](#)]
  86. Man A, Zhu Y, Zhang Y, Dubreuil M, Rho YH, Peloquin C, et al. The risk of cardiovascular disease in systemic sclerosis: a population-based cohort study. *Ann Rheum Dis* 2013 Jul;72(7):1188-1193 [[FREE Full text](#)] [doi: [10.1136/annrheumdis-2012-202007](https://doi.org/10.1136/annrheumdis-2012-202007)] [Medline: [22904260](#)]
  87. Ross EG, Jung K, Dudley JT, Li L, Leeper NJ, Shah NH. Predicting Future Cardiovascular Events in Patients With Peripheral Artery Disease Using Electronic Health Record Data. *Circ Cardiovasc Qual Outcomes* 2019 Mar;12(3):e004741 [[FREE Full text](#)] [doi: [10.1161/CIRCOUTCOMES.118.004741](https://doi.org/10.1161/CIRCOUTCOMES.118.004741)] [Medline: [30857412](#)]
  88. Savova GK, Fan J, Ye Z, Murphy SP, Zheng J, Chute CG, et al. Discovering peripheral arterial disease cases from radiology notes using natural language processing. *AMIA Annu Symp Proc* 2010 Nov 13;2010:722-726 [[FREE Full text](#)] [Medline: [21347073](#)]
  89. Wolfson J, Vock DM, Bandyopadhyay S, Kottke T, Vazquez-Benitez G, Johnson P, et al. Use and Customization of Risk Scores for Predicting Cardiovascular Events Using Electronic Health Record Data. *J Am Heart Assoc* 2017 May 24;6(4) [[FREE Full text](#)] [doi: [10.1161/JAHA.116.003670](https://doi.org/10.1161/JAHA.116.003670)] [Medline: [28438733](#)]
  90. Ammann EM, Leira EC, Winiecki SK, Nagaraja N, Dandapat S, Carnahan RM, et al. Chart validation of inpatient ICD-9-CM administrative diagnosis codes for ischemic stroke among IGIV users in the Sentinel Distributed Database. *Medicine (Baltimore)* 2017 Dec;96(52):e9440 [[FREE Full text](#)] [doi: [10.1097/MD.00000000000009440](https://doi.org/10.1097/MD.00000000000009440)] [Medline: [29384925](#)]
  91. Bell S, Daskalopoulou M, Rapsomaniki E, George J, Britton A, Bobak M, et al. Association between clinically recorded alcohol consumption and initial presentation of 12 cardiovascular diseases: population based cohort study using linked health records. *BMJ* 2017 Mar 22;356:j909. [doi: [10.1136/bmj.j909](https://doi.org/10.1136/bmj.j909)] [Medline: [28331015](#)]
  92. Castro VM, Dligach D, Finan S, Yu S, Can A, Abd-El-Barr M, et al. Large-scale identification of patients with cerebral aneurysms using natural language processing. *Neurology* 2017 Jan 10;88(2):164-168 [[FREE Full text](#)] [doi: [10.1212/WNL.0000000000003490](https://doi.org/10.1212/WNL.0000000000003490)] [Medline: [27927935](#)]
  93. Esteban S, Rodríguez Tablado M, Ricci RI, Terrasa S, Kopitowski K. A rule-based electronic phenotyping algorithm for detecting clinically relevant cardiovascular disease cases. *BMC Res Notes* 2017 Jul 14;10(1):281 [[FREE Full text](#)] [doi: [10.1186/s13104-017-2600-2](https://doi.org/10.1186/s13104-017-2600-2)] [Medline: [28705240](#)]
  94. Garg R, Oh E, Naidech A, Kording K, Prabhakaran S. Automating Ischemic Stroke Subtype Classification Using Machine Learning and Natural Language Processing. *J Stroke Cerebrovasc Dis* 2019 Jul;28(7):2045-2051. [doi: [10.1016/j.jstrokecerebrovasdis.2019.02.004](https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.02.004)] [Medline: [31103549](#)]
  95. Gon Y, Kabata D, Yamamoto K, Shintani A, Todo K, Mochizuki H, et al. Validation of an algorithm that determines stroke diagnostic code accuracy in a Japanese hospital-based cancer registry using electronic medical records. *BMC Med Inform Decis Mak* 2017 Dec 04;17(1):157 [[FREE Full text](#)] [doi: [10.1186/s12911-017-0554-x](https://doi.org/10.1186/s12911-017-0554-x)] [Medline: [29202795](#)]
  96. Gulliford MC, Charlton J, Ashworth M, Rudd AG, Toschke AM, eCRT Research Team. Selection of medical diagnostic codes for analysis of electronic patient records. Application to stroke in a primary care database. *PLoS One* 2009 Oct 24;4(9):e7168 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0007168](https://doi.org/10.1371/journal.pone.0007168)] [Medline: [19777060](#)]
  97. Imran TF, Posner D, Honerlaw J, Vassy JL, Song RJ, Ho Y, et al. A phenotyping algorithm to identify acute ischemic stroke accurately from a national biobank: the Million Veteran Program. *Clin Epidemiol* 2018;10:1509-1521 [[FREE Full text](#)] [doi: [10.2147/CLEP.S160764](https://doi.org/10.2147/CLEP.S160764)] [Medline: [30425582](#)]
  98. Kivimäki M, Batty GD, Singh-Manoux A, Britton A, Brunner EJ, Shipley MJ. Validity of Cardiovascular Disease Event Ascertainment Using Linkage to UK Hospital Records. *Epidemiology* 2017 Sep;28(5):735-739 [[FREE Full text](#)] [doi: [10.1097/EDE.0000000000000688](https://doi.org/10.1097/EDE.0000000000000688)] [Medline: [28570383](#)]
  99. Kogan E, Twyman K, Heap J, Milentijevic D, Lin JH, Alberts M. Assessing stroke severity using electronic health record data: a machine learning approach. *BMC Med Inform Decis Mak* 2020 Jan 08;20(1):8 [[FREE Full text](#)] [doi: [10.1186/s12911-019-1010-x](https://doi.org/10.1186/s12911-019-1010-x)] [Medline: [31914991](#)]
  100. Kreuger AL, Middelburg RA, Beckers EAM, de Vooght KMK, Zwaginga JJ, Kerkhoffs JH, et al. The identification of cases of major hemorrhage during hospitalization in patients with acute leukemia using routinely recorded healthcare data. *PLoS One* 2018;13(8):e0200655 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0200655](https://doi.org/10.1371/journal.pone.0200655)] [Medline: [30110326](#)]
  101. Ljubisavljevic S, Milosevic V, Stojanov A, Ljubisavljevic M, Dunjic O, Zivkovic M. Identification of clinical and paraclinical findings predictive for headache occurrence during spontaneous subarachnoid hemorrhage. *Clin Neurol Neurosurg* 2017 Jul;158:40-45. [doi: [10.1016/j.clineuro.2017.04.017](https://doi.org/10.1016/j.clineuro.2017.04.017)] [Medline: [28458057](#)]
  102. Ni Y, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, et al. Towards phenotyping stroke: Leveraging data from a large-scale epidemiological study to detect stroke diagnosis. *PLoS One* 2018;13(2):e0192586 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0192586](https://doi.org/10.1371/journal.pone.0192586)] [Medline: [29444182](#)]

103. Øie LR, Madsbu MA, Giannadakis C, Vorhaug A, Jensberg H, Salvesen , et al. Validation of intracranial hemorrhage in the Norwegian Patient Registry. *Brain Behav* 2018 Feb;8(2):e00900 [FREE Full text] [doi: [10.1002/brb3.900](https://doi.org/10.1002/brb3.900)] [Medline: [29484261](https://pubmed.ncbi.nlm.nih.gov/29484261/)]
104. Oostema JA, Konen J, Chassee T, Nasiri M, Reeves MJ. Clinical predictors of accurate prehospital stroke recognition. *Stroke* 2015 Jun;46(6):1513-1517. [doi: [10.1161/STROKEAHA.115.008650](https://doi.org/10.1161/STROKEAHA.115.008650)] [Medline: [25922507](https://pubmed.ncbi.nlm.nih.gov/25922507/)]
105. Pouwels KB, Voorham J, Hak E, Denig P. Identification of major cardiovascular events in patients with diabetes using primary care data. *BMC Health Serv Res* 2016 May 02;16:110 [FREE Full text] [doi: [10.1186/s12913-016-1361-2](https://doi.org/10.1186/s12913-016-1361-2)] [Medline: [27038959](https://pubmed.ncbi.nlm.nih.gov/27038959/)]
106. Weinstein R, Ess K, Sirdar B, Song S, Cutting S. Primary Intraventricular Hemorrhage: Clinical Characteristics and Outcomes. *J Stroke Cerebrovasc Dis* 2017 May;26(5):995-999. [doi: [10.1016/j.jstrokecerebrovasdis.2016.11.114](https://doi.org/10.1016/j.jstrokecerebrovasdis.2016.11.114)] [Medline: [28111008](https://pubmed.ncbi.nlm.nih.gov/28111008/)]
107. Wheeler E, Mair G, Sudlow C, Alex B, Grover C, Whiteley W. A validated natural language processing algorithm for brain imaging phenotypes from radiology reports in UK electronic health records. *BMC Med Inform Decis Mak* 2019 Sep 09;19(1):184. [doi: [10.1186/s12911-019-0908-7](https://doi.org/10.1186/s12911-019-0908-7)] [Medline: [31500613](https://pubmed.ncbi.nlm.nih.gov/31500613/)]
108. Amra S, O'Horo JC, Singh TD, Wilson GA, Kashyap R, Petersen R, et al. Derivation and validation of the automated search algorithms to identify cognitive impairment and dementia in electronic health records. *J Crit Care* 2017 Feb;37:202-205. [doi: [10.1016/j.jcrc.2016.09.026](https://doi.org/10.1016/j.jcrc.2016.09.026)] [Medline: [27969571](https://pubmed.ncbi.nlm.nih.gov/27969571/)]
109. Anzaldi LJ, Davison A, Boyd CM, Leff B, Kharrazi H. Comparing clinician descriptions of frailty and geriatric syndromes using electronic health records: a retrospective cohort study. *BMC Geriatr* 2017 Dec 25;17(1):248 [FREE Full text] [doi: [10.1186/s12877-017-0645-7](https://doi.org/10.1186/s12877-017-0645-7)] [Medline: [29070036](https://pubmed.ncbi.nlm.nih.gov/29070036/)]
110. Barnes DE, Zhou J, Walker RL, Larson EB, Lee SJ, Boscardin WJ, et al. Development and Validation of eRADAR: A Tool Using EHR Data to Detect Unrecognized Dementia. *J Am Geriatr Soc* 2020 Jan;68(1):103-111. [doi: [10.1111/jgs.16182](https://doi.org/10.1111/jgs.16182)] [Medline: [31612463](https://pubmed.ncbi.nlm.nih.gov/31612463/)]
111. Boustani M, Perkins AJ, Khandker RK, Duong S, Dexter PR, Lipton R, et al. Passive Digital Signature for Early Identification of Alzheimer's Disease and Related Dementia. *J Am Geriatr Soc* 2020 Mar;68(3):511-518. [doi: [10.1111/jgs.16218](https://doi.org/10.1111/jgs.16218)] [Medline: [31784987](https://pubmed.ncbi.nlm.nih.gov/31784987/)]
112. Corradi JP, Chhabra J, Mather JF, Waszynski CM, Dicks RS. Analysis of multi-dimensional contemporaneous EHR data to refine delirium assessments. *Comput Biol Med* 2016 Aug 01;75:267-274. [doi: [10.1016/j.combiomed.2016.06.013](https://doi.org/10.1016/j.combiomed.2016.06.013)] [Medline: [27340924](https://pubmed.ncbi.nlm.nih.gov/27340924/)]
113. Ernecoff NC, Wessell KL, Gabriel S, Carey TS, Hanson LC. A Novel Screening Method to Identify Late-Stage Dementia Patients for Palliative Care Research and Practice. *J Pain Symptom Manage* 2018 Apr;55(4):1152-1158.e1 [FREE Full text] [doi: [10.1016/j.jpainsymman.2017.12.480](https://doi.org/10.1016/j.jpainsymman.2017.12.480)] [Medline: [29288881](https://pubmed.ncbi.nlm.nih.gov/29288881/)]
114. Ford E, Rooney P, Oliver S, Hoile R, Hurley P, Banerjee S, et al. Identifying undetected dementia in UK primary care patients: a retrospective case-control study comparing machine-learning and standard epidemiological approaches. *BMC Med Inform Decis Mak* 2019 Dec 02;19(1):248 [FREE Full text] [doi: [10.1186/s12911-019-0991-9](https://doi.org/10.1186/s12911-019-0991-9)] [Medline: [31791325](https://pubmed.ncbi.nlm.nih.gov/31791325/)]
115. Halpern R, Seare J, Tong J, Hartry A, Olaoye A, Aigbogun MS. Using electronic health records to estimate the prevalence of agitation in Alzheimer disease/dementia. *Int J Geriatr Psychiatry* 2019 Mar;34(3):420-431 [FREE Full text] [doi: [10.1002/gps.5030](https://doi.org/10.1002/gps.5030)] [Medline: [30430642](https://pubmed.ncbi.nlm.nih.gov/30430642/)]
116. Jaakkimainen RL, Bronskill SE, Tierney MC, Herrmann N, Green D, Young J, et al. Identification of Physician-Diagnosed Alzheimer's Disease and Related Dementias in Population-Based Administrative Data: A Validation Study Using Family Physicians' Electronic Medical Records. *J Alzheimers Dis* 2016 Aug 10;54(1):337-349. [doi: [10.3233/JAD-160105](https://doi.org/10.3233/JAD-160105)] [Medline: [27567819](https://pubmed.ncbi.nlm.nih.gov/27567819/)]
117. Kharrazi H, Anzaldi LJ, Hernandez L, Davison A, Boyd CM, Leff B, et al. The Value of Unstructured Electronic Health Record Data in Geriatric Syndrome Case Identification. *J Am Geriatr Soc* 2018 Aug;66(8):1499-1507. [doi: [10.1111/jgs.15411](https://doi.org/10.1111/jgs.15411)] [Medline: [29972595](https://pubmed.ncbi.nlm.nih.gov/29972595/)]
118. Lewis G, Werbeloff N, Hayes JF, Howard R, Osborn DPJ. Diagnosed depression and sociodemographic factors as predictors of mortality in patients with dementia. *Br J Psychiatry* 2018 Aug;213(2):471-476 [FREE Full text] [doi: [10.1192/bjp.2018.86](https://doi.org/10.1192/bjp.2018.86)] [Medline: [29898791](https://pubmed.ncbi.nlm.nih.gov/29898791/)]
119. McCoy TH, Han L, Pellegrini AM, Tanzi RE, Berretta S, Perlis RH. Stratifying risk for dementia onset using large-scale electronic health record data: A retrospective cohort study. *Alzheimers Dement* 2020 Mar;16(3):531-540. [doi: [10.1016/j.jalz.2019.09.084](https://doi.org/10.1016/j.jalz.2019.09.084)] [Medline: [31859230](https://pubmed.ncbi.nlm.nih.gov/31859230/)]
120. Perera G, Pedersen L, Ansel D, Alexander M, Arrighi HM, Avillach P, et al. Dementia prevalence and incidence in a federation of European Electronic Health Record databases: The European Medical Informatics Framework resource. *Alzheimers Dement* 2018 Feb;14(2):130-139 [FREE Full text] [doi: [10.1016/j.jalz.2017.06.2270](https://doi.org/10.1016/j.jalz.2017.06.2270)] [Medline: [28734783](https://pubmed.ncbi.nlm.nih.gov/28734783/)]
121. Pham TM, Petersen I, Walters K, Raine R, Manthorpe J, Mukadam N, et al. Trends in dementia diagnosis rates in UK ethnic groups: analysis of UK primary care data. *Clin Epidemiol* 2018;10:949-960 [FREE Full text] [doi: [10.2147/CLEP.S152647](https://doi.org/10.2147/CLEP.S152647)] [Medline: [30123007](https://pubmed.ncbi.nlm.nih.gov/30123007/)]

122. Ponjoan A, Garre-Olmo J, Blanch J, Fages E, Alves-Cabratos L, Martí-Lluch R, et al. How well can electronic health records from primary care identify Alzheimer's disease cases? *Clin Epidemiol* 2019;11:509-518 [FREE Full text] [doi: [10.2147/CLEP.S206770](https://doi.org/10.2147/CLEP.S206770)] [Medline: [31456649](https://pubmed.ncbi.nlm.nih.gov/31456649/)]
123. Ponjoan A, Garre-Olmo J, Blanch J, Fages E, Alves-Cabratos L, Martí-Lluch R, et al. Epidemiology of dementia: prevalence and incidence estimates using validated electronic health records from primary care. *Clin Epidemiol* 2019;11:217-228 [FREE Full text] [doi: [10.2147/CLEP.S186590](https://doi.org/10.2147/CLEP.S186590)] [Medline: [30881138](https://pubmed.ncbi.nlm.nih.gov/30881138/)]
124. Pujades-Rodriguez M, Assi V, Gonzalez-Izquierdo A, Wilkinson T, Schnier C, Sudlow C, et al. The diagnosis, burden and prognosis of dementia: A record-linkage cohort study in England. *PLoS One* 2018;13(6):e0199026 [FREE Full text] [doi: [10.1371/journal.pone.0199026](https://doi.org/10.1371/journal.pone.0199026)] [Medline: [29944675](https://pubmed.ncbi.nlm.nih.gov/29944675/)]
125. Reuben DB, Hackbarth AS, Wenger NS, Tan ZS, Jennings LA. An Automated Approach to Identifying Patients with Dementia Using Electronic Medical Records. *J Am Geriatr Soc* 2017 Mar;65(3):658-659. [doi: [10.1111/jgs.14744](https://doi.org/10.1111/jgs.14744)] [Medline: [28152164](https://pubmed.ncbi.nlm.nih.gov/28152164/)]
126. Sommerlad A, Perera G, Mueller C, Singh-Manoux A, Lewis G, Stewart R, et al. Hospitalisation of people with dementia: evidence from English electronic health records from 2008 to 2016. *Eur J Epidemiol* 2019 Jul;34(6):567-577 [FREE Full text] [doi: [10.1007/s10654-019-00481-x](https://doi.org/10.1007/s10654-019-00481-x)] [Medline: [30649705](https://pubmed.ncbi.nlm.nih.gov/30649705/)]
127. van Bussel EF, Richard E, Arts DL, Nooyens ACJ, Coloma PM, de Waal MWM, et al. Dementia incidence trend over 1992-2014 in the Netherlands: Analysis of primary care data. *PLoS Med* 2017 Mar;14(3):e1002235 [FREE Full text] [doi: [10.1371/journal.pmed.1002235](https://doi.org/10.1371/journal.pmed.1002235)] [Medline: [28267788](https://pubmed.ncbi.nlm.nih.gov/28267788/)]
128. Wei W, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. *J Am Med Inform Assoc* 2016 Apr;23(e1):e20-e27 [FREE Full text] [doi: [10.1093/jamia/ocv130](https://doi.org/10.1093/jamia/ocv130)] [Medline: [26338219](https://pubmed.ncbi.nlm.nih.gov/26338219/)]
129. 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 Jul;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/)]
130. Wu C, Kuo C, Su C, Wang S, Dai H. Using text mining to extract depressive symptoms and to validate the diagnosis of major depressive disorder from electronic health records. *J Affect Disord* 2020 Jan 01;260:617-623. [doi: [10.1016/j.jad.2019.09.044](https://doi.org/10.1016/j.jad.2019.09.044)] [Medline: [31541973](https://pubmed.ncbi.nlm.nih.gov/31541973/)]
131. Afzal Z, Engelkes M, Verhamme KMC, Janssens HM, Sturkenboom MCJM, Kors JA, et al. Automatic generation of case-detection algorithms to identify children with asthma from large electronic health record databases. *Pharmacoepidemiol Drug Saf* 2013 Aug;22(8):826-833. [doi: [10.1002/pds.3438](https://doi.org/10.1002/pds.3438)] [Medline: [23592573](https://pubmed.ncbi.nlm.nih.gov/23592573/)]
132. Akgün KM, Sigel K, Cheung K, Kidwai-Khan F, Bryant AK, Brandt C, et al. Extracting lung function measurements to enhance phenotyping of chronic obstructive pulmonary disease (COPD) in an electronic health record using automated tools. *PLoS One* 2020;15(1):e0227730 [FREE Full text] [doi: [10.1371/journal.pone.0227730](https://doi.org/10.1371/journal.pone.0227730)] [Medline: [31945115](https://pubmed.ncbi.nlm.nih.gov/31945115/)]
133. Almoquera B, Vazquez L, Mentch F, Connolly J, Pacheco JA, Sundaresan AS, et al. Identification of Four Novel Loci in Asthma in European American and African American Populations. *Am J Respir Crit Care Med* 2017 Mar 15;195(4):456-463. [doi: [10.1164/rccm.201604-0861OC](https://doi.org/10.1164/rccm.201604-0861OC)] [Medline: [27611488](https://pubmed.ncbi.nlm.nih.gov/27611488/)]
134. Asche C, Said Q, Joish V, Hall CO, Brixner D. Assessment of COPD-related outcomes via a national electronic medical record database. *Int J Chron Obstruct Pulmon Dis* 2008;3(2):323-326 [FREE Full text] [doi: [10.2147/copd.s1857](https://doi.org/10.2147/copd.s1857)] [Medline: [18686742](https://pubmed.ncbi.nlm.nih.gov/18686742/)]
135. Borlée F, Yzermans CJ, Krop E, Aalders B, Rooijackers J, Zock J, et al. Spirometry, questionnaire and electronic medical record based COPD in a population survey: Comparing prevalence, level of agreement and associations with potential risk factors. *PLoS One* 2017;12(3):e0171494 [FREE Full text] [doi: [10.1371/journal.pone.0171494](https://doi.org/10.1371/journal.pone.0171494)] [Medline: [28273094](https://pubmed.ncbi.nlm.nih.gov/28273094/)]
136. Cave AJ, Davey C, Ahmadi E, Drummond N, Fuentes S, Kazemi-Bajestani SMR, et al. Development of a validated algorithm for the diagnosis of paediatric asthma in electronic medical records. *NPJ Prim Care Respir Med* 2016 Dec 24;26:16085 [FREE Full text] [doi: [10.1038/npjpcrm.2016.85](https://doi.org/10.1038/npjpcrm.2016.85)] [Medline: [27882997](https://pubmed.ncbi.nlm.nih.gov/27882997/)]
137. Crothers K, Rodriguez CV, Nance RM, Akgun K, Shahrir S, Kim J, et al. Accuracy of electronic health record data for the diagnosis of chronic obstructive pulmonary disease in persons living with HIV and uninfected persons. *Pharmacoepidemiol Drug Saf* 2019 Feb;28(2):140-147. [doi: [10.1002/pds.4567](https://doi.org/10.1002/pds.4567)] [Medline: [29923258](https://pubmed.ncbi.nlm.nih.gov/29923258/)]
138. DiSantostefano RL, Sampson T, Le HV, Hinds D, Davis KJ, Bakerly ND. Risk of pneumonia with inhaled corticosteroid versus long-acting bronchodilator regimens in chronic obstructive pulmonary disease: a new-user cohort study. *PLoS One* 2014;9(5):e97149 [FREE Full text] [doi: [10.1371/journal.pone.0097149](https://doi.org/10.1371/journal.pone.0097149)] [Medline: [24878543](https://pubmed.ncbi.nlm.nih.gov/24878543/)]
139. Hsu J, Pacheco JA, Stevens WW, Smith ME, Avila PC. Accuracy of phenotyping chronic rhinosinusitis in the electronic health record. *Am J Rhinol Allergy* 2014;28(2):140-144 [FREE Full text] [doi: [10.2500/ajra.2014.28.4012](https://doi.org/10.2500/ajra.2014.28.4012)] [Medline: [24717952](https://pubmed.ncbi.nlm.nih.gov/24717952/)]
140. Kadhim-Saleh A, Green M, Williamson T, Hunter D, Birtwhistle R. Validation of the diagnostic algorithms for 5 chronic conditions in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN): a Kingston Practice-based Research Network (PBRN) report. *J Am Board Fam Med* 2013;26(2):159-167 [FREE Full text] [doi: [10.3122/jabfm.2013.02.120183](https://doi.org/10.3122/jabfm.2013.02.120183)] [Medline: [23471929](https://pubmed.ncbi.nlm.nih.gov/23471929/)]



141. Kurmi OP, Vaucher J, Xiao D, Holmes MV, Guo Y, Davis KJ, et al. Validity of COPD diagnoses reported through nationwide health insurance systems in the People's Republic of China. *Int J Chron Obstruct Pulmon Dis* 2016;11:419-430 [[FREE Full text](#)] [doi: [10.2147/COPD.S100736](https://doi.org/10.2147/COPD.S100736)] [Medline: [27042034](#)]
142. Lee TM, Tu K, Wing LL, Gershon AS. Identifying individuals with physician-diagnosed chronic obstructive pulmonary disease in primary care electronic medical records: a retrospective chart abstraction study. *NPJ Prim Care Respir Med* 2017 May 15;27(1):34 [[FREE Full text](#)] [doi: [10.1038/s41533-017-0035-9](https://doi.org/10.1038/s41533-017-0035-9)] [Medline: [28507288](#)]
143. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med* 2019 Mar;25(3):433-438. [doi: [10.1038/s41591-018-0335-9](https://doi.org/10.1038/s41591-018-0335-9)] [Medline: [30742121](#)]
144. Nissen F, Morales DR, Mullerova H, Smeeth L, Douglas IJ, Quint JK. Validation of asthma recording in the Clinical Practice Research Datalink (CPRD). *BMJ Open* 2017 Aug 11;7(8):e017474 [[FREE Full text](#)] [doi: [10.1136/bmjopen-2017-017474](https://doi.org/10.1136/bmjopen-2017-017474)] [Medline: [28801439](#)]
145. Nissen F, Morales DR, Mullerova H, Smeeth L, Douglas IJ, Quint JK. Concomitant diagnosis of asthma and COPD: a quantitative study in UK primary care. *Br J Gen Pract* 2018 Dec;68(676):e775-e782. [doi: [10.3399/bjgp18X699389](https://doi.org/10.3399/bjgp18X699389)] [Medline: [30249612](#)]
146. Pacheco JA, Avila PC, Thompson JA, Law M, Quraishi JA, Greiman AK, et al. A highly specific algorithm for identifying asthma cases and controls for genome-wide association studies. *AMIA Annu Symp Proc* 2009 Dec 14;2009:497-501. [Medline: [20351906](#)]
147. Pennington AF, Strickland MJ, Freedle KA, Klein M, Drews-Botsch C, Hansen C, et al. Evaluating early-life asthma definitions as a marker for subsequent asthma in an electronic medical record setting. *Pediatr Allergy Immunol* 2016 Sep;27(6):591-596. [doi: [10.1111/pai.12586](https://doi.org/10.1111/pai.12586)] [Medline: [27116587](#)]
148. Rothnie KJ, Chandan JS, Goss HG, Müllerová H, Quint JK. Validity and interpretation of spirometric recordings to diagnose COPD in UK primary care. *Int J Chron Obstruct Pulmon Dis* 2017;12:1663-1668 [[FREE Full text](#)] [doi: [10.2147/COPD.S133891](https://doi.org/10.2147/COPD.S133891)] [Medline: [28652719](#)]
149. Rothnie KJ, Müllerová H, Hurst JR, Smeeth L, Davis K, Thomas SL, et al. Validation of the Recording of Acute Exacerbations of COPD in UK Primary Care Electronic Healthcare Records. *PLoS One* 2016;11(3):e0151357 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0151357](https://doi.org/10.1371/journal.pone.0151357)] [Medline: [26959820](#)]
150. Schulz S, Seddig T, Hanser S, Zaiss A, Daumke P. Checking coding completeness by mining discharge summaries. *Stud Health Technol Inform* 2011;169:594-598. [Medline: [21893818](#)]
151. Seol HY, Rolfes MC, Chung W, Sohn S, Ryu E, Park MA, et al. Expert artificial intelligence-based natural language processing characterises childhood asthma. *BMJ Open Res* 2020 Feb 04;7(1):e000524. [doi: [10.1136/bmjresp-2019-000524](https://doi.org/10.1136/bmjresp-2019-000524)]
152. Sohn S, Wang Y, Wi C, Krusemark EA, Ryu E, Ali MH, et al. Clinical documentation variations and NLP system portability: a case study in asthma birth cohorts across institutions. *J Am Med Inform Assoc* 2017 Nov 30. [doi: [10.1093/jamia/ocx138](https://doi.org/10.1093/jamia/ocx138)] [Medline: [29202185](#)]
153. Sohn S, Wi C, Wu ST, Liu H, Ryu E, Krusemark E, et al. Ascertainment of asthma prognosis using natural language processing from electronic medical records. *J Allergy Clin Immunol* 2018 Jun;141(6):2292-2294.e3 [[FREE Full text](#)] [doi: [10.1016/j.jaci.2017.12.1003](https://doi.org/10.1016/j.jaci.2017.12.1003)] [Medline: [29438770](#)]
154. Sperrin M, Webb DJ, Patel P, Davis KJ, Collier S, Pate A, et al. Chronic obstructive pulmonary disease exacerbation episodes derived from electronic health record data validated using clinical trial data. *Pharmacoepidemiol Drug Saf* 2019 Oct;28(10):1369-1376 [[FREE Full text](#)] [doi: [10.1002/pds.4883](https://doi.org/10.1002/pds.4883)] [Medline: [31385428](#)]
155. Sundaresan AS, Schneider G, Reynolds J, Kirchner HL. Identifying Asthma Exacerbation-Related Emergency Department Visit Using Electronic Medical Record and Claims Data. *Appl Clin Inform* 2018 Jul;9(3):528-540 [[FREE Full text](#)] [doi: [10.1055/s-0038-1666994](https://doi.org/10.1055/s-0038-1666994)] [Medline: [30040112](#)]
156. Vazquez Guillamet R, Ursu O, Iwamoto G, Moseley PL, Oprea T. Chronic obstructive pulmonary disease phenotypes using cluster analysis of electronic medical records. *Health Informatics J* 2018 Dec;24(4):394-409 [[FREE Full text](#)] [doi: [10.1177/1460458216675661](https://doi.org/10.1177/1460458216675661)] [Medline: [27856785](#)]
157. Wi C, Sohn S, Ali M, Krusemark E, Ryu E, Liu H, et al. Natural Language Processing for Asthma Ascertainment in Different Practice Settings. *J Allergy Clin Immunol Pract* 2018;6(1):126-131 [[FREE Full text](#)] [doi: [10.1016/j.jaip.2017.04.041](https://doi.org/10.1016/j.jaip.2017.04.041)] [Medline: [28634104](#)]
158. Wi C, Sohn S, Rolfes MC, Seabright A, Ryu E, Voge G, et al. Application of a Natural Language Processing Algorithm to Asthma Ascertainment. An Automated Chart Review. *Am J Respir Crit Care Med* 2017 Aug 15;196(4):430-437 [[FREE Full text](#)] [doi: [10.1164/rccm.201610-2006OC](https://doi.org/10.1164/rccm.201610-2006OC)] [Medline: [28375665](#)]
159. Wu ST, Sohn S, Ravikumar KE, Waghlikar K, Jonnalagadda SR, Liu H, et al. Automated chart review for asthma cohort identification using natural language processing: an exploratory study. *Ann Allergy Asthma Immunol* 2013 Dec;111(5):364-369 [[FREE Full text](#)] [doi: [10.1016/j.anai.2013.07.022](https://doi.org/10.1016/j.anai.2013.07.022)] [Medline: [24125142](#)]
160. Zeng QT, Goryachev S, Weiss S, Sordo M, Murphy SN, Lazarus R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. *BMC Med Inform Decis Mak* 2006;6:30 [[FREE Full text](#)] [doi: [10.1186/1472-6947-6-30](https://doi.org/10.1186/1472-6947-6-30)] [Medline: [16872495](#)]

161. Barnado A, Casey C, Carroll RJ, Wheless L, Denny JC, Crofford LJ. Developing Electronic Health Record Algorithms That Accurately Identify Patients With Systemic Lupus Erythematosus. *Arthritis Care Res (Hoboken)* 2017 May;69(5):687-693 [FREE Full text] [doi: [10.1002/acr.22989](https://doi.org/10.1002/acr.22989)] [Medline: [27390187](https://pubmed.ncbi.nlm.nih.gov/27390187/)]
162. Carroll RJ, Eyer AE, Denny JC. Naïve Electronic Health Record phenotype identification for Rheumatoid arthritis. *AMIA Annu Symp Proc* 2011;2011:189-196. [Medline: [22195070](https://pubmed.ncbi.nlm.nih.gov/22195070/)]
163. Carroll RJ, Thompson WK, Eyer AE, Mandelin AM, Cai T, Zink RM, et al. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. *J Am Med Inform Assoc* 2012 Jun;19(e1):e162-e169 [FREE Full text] [doi: [10.1136/amiajnl-2011-000583](https://doi.org/10.1136/amiajnl-2011-000583)] [Medline: [22374935](https://pubmed.ncbi.nlm.nih.gov/22374935/)]
164. Cote J, Berger A, Kirchner LH, Bili A. Low vitamin D level is not associated with increased incidence of rheumatoid arthritis. *Rheumatol Int* 2014 Oct;34(10):1475-1479. [doi: [10.1007/s00296-014-3019-x](https://doi.org/10.1007/s00296-014-3019-x)] [Medline: [24748471](https://pubmed.ncbi.nlm.nih.gov/24748471/)]
165. de Abreu MM, Maiorano AC, Tedeschi SK, Yoshida K, Lin T, Solomon DH. Outcomes of lupus and rheumatoid arthritis patients with primary dengue infection: A seven-year report from Brazil. *Semin Arthritis Rheum* 2018 Apr;47(5):749-755. [doi: [10.1016/j.semarthrit.2017.09.001](https://doi.org/10.1016/j.semarthrit.2017.09.001)] [Medline: [29061443](https://pubmed.ncbi.nlm.nih.gov/29061443/)]
166. Escudié J, Rance B, Malamut G, Khater S, Burgun A, Cellier C, et al. A novel data-driven workflow combining literature and electronic health records to estimate comorbidities burden for a specific disease: a case study on autoimmune comorbidities in patients with celiac disease. *BMC Med Inform Decis Mak* 2017 Oct 29;17(1):140 [FREE Full text] [doi: [10.1186/s12911-017-0537-y](https://doi.org/10.1186/s12911-017-0537-y)] [Medline: [28962565](https://pubmed.ncbi.nlm.nih.gov/28962565/)]
167. Ford E, Carroll J, Smith H, Davies K, Koeling R, Petersen I, et al. What evidence is there for a delay in diagnostic coding of RA in UK general practice records? An observational study of free text. *BMJ Open* 2016 Jun 28;6(6):e010393 [FREE Full text] [doi: [10.1136/bmjopen-2015-010393](https://doi.org/10.1136/bmjopen-2015-010393)] [Medline: [27354069](https://pubmed.ncbi.nlm.nih.gov/27354069/)]
168. Ford E, Nicholson A, Koeling R, Tate A, Carroll J, Axelrod L, et al. Optimising the use of electronic health records to estimate the incidence of rheumatoid arthritis in primary care: what information is hidden in free text? *BMC Med Res Methodol* 2013 Aug 21;13:105 [FREE Full text] [doi: [10.1186/1471-2288-13-105](https://doi.org/10.1186/1471-2288-13-105)] [Medline: [23964710](https://pubmed.ncbi.nlm.nih.gov/23964710/)]
169. Jamian L, Wheless L, Crofford LJ, Barnado A. Rule-based and machine learning algorithms identify patients with systemic sclerosis accurately in the electronic health record. *Arthritis Res Ther* 2019 Dec 30;21(1):305. [doi: [10.1186/s13075-019-2092-7](https://doi.org/10.1186/s13075-019-2092-7)] [Medline: [31888720](https://pubmed.ncbi.nlm.nih.gov/31888720/)]
170. Jorge A, Castro VM, Barnado A, Gainer V, Hong C, Cai T, et al. Identifying lupus patients in electronic health records: Development and validation of machine learning algorithms and application of rule-based algorithms. *Semin Arthritis Rheum* 2019 Aug;49(1):84-90 [FREE Full text] [doi: [10.1016/j.semarthrit.2019.01.002](https://doi.org/10.1016/j.semarthrit.2019.01.002)] [Medline: [30665626](https://pubmed.ncbi.nlm.nih.gov/30665626/)]
171. Kronzer VL, Wang L, Liu H, Davis JM, Sparks JA, Crowson CS. Investigating the impact of disease and health record duration on the eMERGE algorithm for rheumatoid arthritis. *J Am Med Inform Assoc* 2020 May 01;27(4):601-605. [doi: [10.1093/jamia/ocaa014](https://doi.org/10.1093/jamia/ocaa014)] [Medline: [32134444](https://pubmed.ncbi.nlm.nih.gov/32134444/)]
172. Liao KP, Cai T, Gainer V, Goryachev S, Zeng-treitler Q, Raychaudhuri S, et al. Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2010 Aug;62(8):1120-1127 [FREE Full text] [doi: [10.1002/acr.20184](https://doi.org/10.1002/acr.20184)] [Medline: [20235204](https://pubmed.ncbi.nlm.nih.gov/20235204/)]
173. Lin C, Karlson EW, Canhao H, Miller TA, Dligach D, Chen PJ, et al. Automatic prediction of rheumatoid arthritis disease activity from the electronic medical records. *PLoS One* 2013;8(8):e69932 [FREE Full text] [doi: [10.1371/journal.pone.0069932](https://doi.org/10.1371/journal.pone.0069932)] [Medline: [23976944](https://pubmed.ncbi.nlm.nih.gov/23976944/)]
174. Muller S, Hider SL, Raza K, Stack RJ, Hayward RA, Mallen CD. An algorithm to identify rheumatoid arthritis in primary care: a Clinical Practice Research Datalink study. *BMJ Open* 2015 Dec 23;5(12):e009309 [FREE Full text] [doi: [10.1136/bmjopen-2015-009309](https://doi.org/10.1136/bmjopen-2015-009309)] [Medline: [26700281](https://pubmed.ncbi.nlm.nih.gov/26700281/)]
175. Murray SG, Avati A, Schmajuk G, Yazdany J. Automated and flexible identification of complex disease: building a model for systemic lupus erythematosus using noisy labeling. *J Am Med Inform Assoc* 2019 Jan 01;26(1):61-65. [doi: [10.1093/jamia/ocy154](https://doi.org/10.1093/jamia/ocy154)] [Medline: [30476175](https://pubmed.ncbi.nlm.nih.gov/30476175/)]
176. Nicholson A, Ford E, Davies KA, Smith HE, Rait G, Tate AR, et al. Optimising use of electronic health records to describe the presentation of rheumatoid arthritis in primary care: a strategy for developing code lists. *PLoS One* 2013;8(2):e54878 [FREE Full text] [doi: [10.1371/journal.pone.0054878](https://doi.org/10.1371/journal.pone.0054878)] [Medline: [23451024](https://pubmed.ncbi.nlm.nih.gov/23451024/)]
177. Nielen MMJ, Ursum J, Schellevis FG, Korevaar JC. The validity of the diagnosis of inflammatory arthritis in a large population-based primary care database. *BMC Fam Pract* 2013 Jul 07;14:79 [FREE Full text] [doi: [10.1186/1471-2296-14-79](https://doi.org/10.1186/1471-2296-14-79)] [Medline: [24128086](https://pubmed.ncbi.nlm.nih.gov/24128086/)]
178. Nikiphorou E, de Lusignan S, Mallen CD, Khavandi K, Bedarida G, Buckley CD, et al. Cardiovascular risk factors and outcomes in early rheumatoid arthritis: a population-based study. *Heart* 2020 Mar 24 [FREE Full text] [doi: [10.1136/heartjnl-2019-316193](https://doi.org/10.1136/heartjnl-2019-316193)] [Medline: [32209618](https://pubmed.ncbi.nlm.nih.gov/32209618/)]
179. Partington RJ, Muller S, Helliwell T, Mallen CD, Abdul Sultan A. Incidence, prevalence and treatment burden of polymyalgia rheumatica in the UK over two decades: a population-based study. *Ann Rheum Dis* 2018 Dec;77(12):1750-1756. [doi: [10.1136/annrheumdis-2018-213883](https://doi.org/10.1136/annrheumdis-2018-213883)] [Medline: [30297332](https://pubmed.ncbi.nlm.nih.gov/30297332/)]
180. Redd D, Frech TM, Murtaugh MA, Rhiannon J, Zeng QT. Informatics can identify systemic sclerosis (SSc) patients at risk for scleroderma renal crisis. *Comput Biol Med* 2014 Oct;53:203-205 [FREE Full text] [doi: [10.1016/j.compbiomed.2014.07.022](https://doi.org/10.1016/j.compbiomed.2014.07.022)] [Medline: [25168254](https://pubmed.ncbi.nlm.nih.gov/25168254/)]

181. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet* 2010 Apr 09;86(4):560-572 [FREE Full text] [doi: [10.1016/j.ajhg.2010.03.003](https://doi.org/10.1016/j.ajhg.2010.03.003)] [Medline: [20362271](https://pubmed.ncbi.nlm.nih.gov/20362271/)]
182. Turner CA, Jacobs AD, Marques CK, Oates JC, Kamen DL, Anderson PE, et al. Word2Vec inversion and traditional text classifiers for phenotyping lupus. *BMC Med Inform Decis Mak* 2017 Aug 22;17(1):126 [FREE Full text] [doi: [10.1186/s12911-017-0518-1](https://doi.org/10.1186/s12911-017-0518-1)] [Medline: [28830409](https://pubmed.ncbi.nlm.nih.gov/28830409/)]
183. Verma A, Basile AO, Bradford Y, Kuivaniemi H, Tromp G, Carey D, et al. Phenome-Wide Association Study to Explore Relationships between Immune System Related Genetic Loci and Complex Traits and Diseases. *PLoS One* 2016;11(8):e0160573 [FREE Full text] [doi: [10.1371/journal.pone.0160573](https://doi.org/10.1371/journal.pone.0160573)] [Medline: [27508393](https://pubmed.ncbi.nlm.nih.gov/27508393/)]
184. Wang L, Rastegar-Mojarad M, Ji Z, Liu S, Liu K, Moon S, et al. Detecting Pharmacovigilance Signals Combining Electronic Medical Records With Spontaneous Reports: A Case Study of Conventional Disease-Modifying Antirheumatic Drugs for Rheumatoid Arthritis. *Front Pharmacol* 2018;9:875. [doi: [10.3389/fphar.2018.00875](https://doi.org/10.3389/fphar.2018.00875)] [Medline: [30131701](https://pubmed.ncbi.nlm.nih.gov/30131701/)]
185. Zhou S, Fernandez-Gutierrez F, Kennedy J, Cooksey R, Atkinson M, Denaxas S, UK Biobank Follow-up Outcomes Group, et al. Defining Disease Phenotypes in Primary Care Electronic Health Records by a Machine Learning Approach: A Case Study in Identifying Rheumatoid Arthritis. *PLoS One* 2016;11(5):e0154515 [FREE Full text] [doi: [10.1371/journal.pone.0154515](https://doi.org/10.1371/journal.pone.0154515)] [Medline: [27135409](https://pubmed.ncbi.nlm.nih.gov/27135409/)]
186. Gill JM, Mainous AG, Koopman RJ, Player MS, Everett CJ, Chen YX, et al. Impact of EHR-based clinical decision support on adherence to guidelines for patients on NSAIDs: a randomized controlled trial. *Ann Fam Med* 2011;9(1):22-30 [FREE Full text] [doi: [10.1370/afm.1172](https://doi.org/10.1370/afm.1172)] [Medline: [21242557](https://pubmed.ncbi.nlm.nih.gov/21242557/)]
187. Salmasian H, Freedberg DE, Abrams JA, Friedman C. An automated tool for detecting medication overuse based on the electronic health records. *Pharmacoepidemiol Drug Saf* 2013 Feb;22(2):183-189 [FREE Full text] [doi: [10.1002/pds.3387](https://doi.org/10.1002/pds.3387)] [Medline: [23233423](https://pubmed.ncbi.nlm.nih.gov/23233423/)]
188. Shelat VG, Ahmed S, Chia CLK, Cheah YL. Strict Selection Criteria During Surgical Training Ensures Good Outcomes in Laparoscopic Omental Patch Repair (LOPR) for Perforated Peptic Ulcer (PPU). *Int Surg* 2015 Mar;100(2):370-375. [doi: [10.9738/INTSURG-D-13-00241.1](https://doi.org/10.9738/INTSURG-D-13-00241.1)] [Medline: [25692444](https://pubmed.ncbi.nlm.nih.gov/25692444/)]
189. Singh B, Singh A, Ahmed A, Wilson GA, Pickering BW, Herasevich V, et al. Derivation and validation of automated electronic search strategies to extract Charlson comorbidities from electronic medical records. *Mayo Clin Proc* 2012 Sep;87(9):817-824 [FREE Full text] [doi: [10.1016/j.mayocp.2012.04.015](https://doi.org/10.1016/j.mayocp.2012.04.015)] [Medline: [22958988](https://pubmed.ncbi.nlm.nih.gov/22958988/)]
190. Ahmad FS, Chan C, Rosenman MB, Post WS, Fort DG, Greenland P, et al. Validity of Cardiovascular Data From Electronic Sources: The Multi-Ethnic Study of Atherosclerosis and HealthLNK. *Circulation* 2017 Oct 26;136(13):1207-1216. [doi: [10.1161/CIRCULATIONAHA.117.027436](https://doi.org/10.1161/CIRCULATIONAHA.117.027436)] [Medline: [28687707](https://pubmed.ncbi.nlm.nih.gov/28687707/)]
191. Chi GC, Li X, Tartof SY, Slezak JM, Koebnick C, Lawrence JM. Validity of ICD-10-CM codes for determination of diabetes type for persons with youth-onset type 1 and type 2 diabetes. *BMJ Open Diabetes Res Care* 2019;7(1):e000547 [FREE Full text] [doi: [10.1136/bmjdr-2018-000547](https://doi.org/10.1136/bmjdr-2018-000547)] [Medline: [30899525](https://pubmed.ncbi.nlm.nih.gov/30899525/)]
192. Coleman N, Halas G, Peeler W, Casaclang N, Williamson T, Katz A. From patient care to research: a validation study examining the factors contributing to data quality in a primary care electronic medical record database. *BMC Fam Pract* 2015 Feb 05;16:11 [FREE Full text] [doi: [10.1186/s12875-015-0223-z](https://doi.org/10.1186/s12875-015-0223-z)] [Medline: [25649201](https://pubmed.ncbi.nlm.nih.gov/25649201/)]
193. Crawford AG, Cote C, Couto J, Daskiran M, Gunnarsson C, Haas K, et al. Prevalence of obesity, type II diabetes mellitus, hyperlipidemia, and hypertension in the United States: findings from the GE Centricity Electronic Medical Record database. *Popul Health Manag* 2010 Jul;13(3):151-161. [doi: [10.1089/pop.2009.0039](https://doi.org/10.1089/pop.2009.0039)] [Medline: [20521902](https://pubmed.ncbi.nlm.nih.gov/20521902/)]
194. Esteban S, Rodríguez Tablado M, Peper F, Mahumud YS, Ricci RI, Kopitowski K, et al. Development and Validation of Various Phenotyping Algorithms for Diabetes Mellitus Using Data from Electronic Health Records. *Stud Health Technol Inform* 2017;245:366-369. [Medline: [29295117](https://pubmed.ncbi.nlm.nih.gov/29295117/)]
195. Gjelsvik B, Tran AT, Berg TJ, Bakke , Mdala I, Nøkleby K, et al. Exploring the relationship between coronary heart disease and type 2 diabetes: a cross-sectional study of secondary prevention among diabetes patients. *BJGP Open* 2019 May;3(1):bjgpopen18X101636. [doi: [10.3399/bjgpopen18X101636](https://doi.org/10.3399/bjgpopen18X101636)] [Medline: [31049416](https://pubmed.ncbi.nlm.nih.gov/31049416/)]
196. Harris SB, Glazier RH, Tompkins JW, Wilton AS, Chevendra V, Stewart MA, et al. Investigating concordance in diabetes diagnosis between primary care charts (electronic medical records) and health administrative data: a retrospective cohort study. *BMC Health Serv Res* 2010 Dec 23;10:347. [doi: [10.1186/1472-6963-10-347](https://doi.org/10.1186/1472-6963-10-347)] [Medline: [21182790](https://pubmed.ncbi.nlm.nih.gov/21182790/)]
197. Henderson J, Barnett S, Ghosh A, Pollack AJ, Hodgkins A, Win KT, et al. Validation of electronic medical data: Identifying diabetes prevalence in general practice. *Health Inf Manag* 2019 Jan;48(1):3-11. [doi: [10.1177/1833358318798123](https://doi.org/10.1177/1833358318798123)] [Medline: [30278786](https://pubmed.ncbi.nlm.nih.gov/30278786/)]
198. Ho ML, Lawrence N, van Walraven C, Manuel D, Keely E, Malcolm J, et al. The accuracy of using integrated electronic health care data to identify patients with undiagnosed diabetes mellitus. *J Eval Clin Pract* 2012 Jul;18(3):606-611. [doi: [10.1111/j.1365-2753.2011.01633.x](https://doi.org/10.1111/j.1365-2753.2011.01633.x)] [Medline: [21332609](https://pubmed.ncbi.nlm.nih.gov/21332609/)]
199. Kadhim-Saleh A, Green M, Williamson T, Hunter D, Birtwhistle R. Validation of the diagnostic algorithms for 5 chronic conditions in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN): a Kingston Practice-based Research Network (PBRN) report. *J Am Board Fam Med* 2013;26(2):159-167 [FREE Full text] [doi: [10.3122/jabfm.2013.02.120183](https://doi.org/10.3122/jabfm.2013.02.120183)] [Medline: [23471929](https://pubmed.ncbi.nlm.nih.gov/23471929/)]

200. Ke C, Stukel TA, Luk A, Shah BR, Jha P, Lau E, et al. Development and validation of algorithms to classify type 1 and 2 diabetes according to age at diagnosis using electronic health records. *BMC Med Res Methodol* 2020 Feb 24;20(1):35 [FREE Full text] [doi: [10.1186/s12874-020-00921-3](https://doi.org/10.1186/s12874-020-00921-3)] [Medline: [32093635](https://pubmed.ncbi.nlm.nih.gov/32093635/)]
201. Kho AN, Hayes MG, Rasmussen-Torvik L, Pacheco JA, Thompson WK, Armstrong LL, et al. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. *J Am Med Inform Assoc* 2012;19(2):212-218 [FREE Full text] [doi: [10.1136/amiajnl-2011-000439](https://doi.org/10.1136/amiajnl-2011-000439)] [Medline: [22101970](https://pubmed.ncbi.nlm.nih.gov/22101970/)]
202. Khokhar B, Quan H, Kaplan GG, Butalia S, Rabi D. Exploring novel diabetes surveillance methods: a comparison of administrative, laboratory and pharmacy data case definitions using THIN. *J Public Health (Oxf)* 2018 Sep 01;40(3):652-658. [doi: [10.1093/pubmed/fox096](https://doi.org/10.1093/pubmed/fox096)] [Medline: [28977382](https://pubmed.ncbi.nlm.nih.gov/28977382/)]
203. Klompas M, Eggleston E, McVetta J, Lazarus R, Li L, Platt R. Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. *Diabetes Care* 2013 Apr;36(4):914-921 [FREE Full text] [doi: [10.2337/dc12-0964](https://doi.org/10.2337/dc12-0964)] [Medline: [23193215](https://pubmed.ncbi.nlm.nih.gov/23193215/)]
204. Kosowan L, Wicklow B, Queenan J, Yeung R, Amed S, Singer A. Enhancing Health Surveillance: Validation of a Novel Electronic Medical Records-Based Definition of Cases of Pediatric Type 1 and Type 2 Diabetes Mellitus. *Can J Diabetes* 2019 Aug;43(6):392-398. [doi: [10.1016/j.jcjd.2019.02.005](https://doi.org/10.1016/j.jcjd.2019.02.005)] [Medline: [30956098](https://pubmed.ncbi.nlm.nih.gov/30956098/)]
205. Kudryakov R, Bowen J, Ewen E, West SL, Daoud Y, Fleming N, et al. Electronic health record use to classify patients with newly diagnosed versus preexisting type 2 diabetes: infrastructure for comparative effectiveness research and population health management. *Popul Health Manag* 2012 Mar;15(1):3-11. [doi: [10.1089/pop.2010.0084](https://doi.org/10.1089/pop.2010.0084)] [Medline: [21877923](https://pubmed.ncbi.nlm.nih.gov/21877923/)]
206. Lawrence JM, Black MH, Zhang JL, Slezak JM, Takhar HS, Koebnick C, et al. Validation of pediatric diabetes case identification approaches for diagnosed cases by using information in the electronic health records of a large integrated managed health care organization. *Am J Epidemiol* 2014 Jan 01;179(1):27-38. [doi: [10.1093/aje/kwt230](https://doi.org/10.1093/aje/kwt230)] [Medline: [24100956](https://pubmed.ncbi.nlm.nih.gov/24100956/)]
207. Lipscombe LL, Hwee J, Webster L, Shah BR, Booth GL, Tu K. Identifying diabetes cases from administrative data: a population-based validation study. *BMC Health Serv Res* 2018 May 02;18(1):316. [doi: [10.1186/s12913-018-3148-0](https://doi.org/10.1186/s12913-018-3148-0)] [Medline: [29720153](https://pubmed.ncbi.nlm.nih.gov/29720153/)]
208. Makam AN, Nguyen OK, Moore B, Ma Y, Amarasingham R. Identifying patients with diabetes and the earliest date of diagnosis in real time: an electronic health record case-finding algorithm. *BMC Med Inform Decis Mak* 2013 Aug 01;13:81 [FREE Full text] [doi: [10.1186/1472-6947-13-81](https://doi.org/10.1186/1472-6947-13-81)] [Medline: [23915139](https://pubmed.ncbi.nlm.nih.gov/23915139/)]
209. Moreno-Iribas C, Sayon-Orea C, Delfrade J, Ardanaz E, Gorricho J, Burgui R, et al. Validity of type 2 diabetes diagnosis in a population-based electronic health record database. *BMC Med Inform Decis Mak* 2017 Apr 08;17(1):34. [doi: [10.1186/s12911-017-0439-z](https://doi.org/10.1186/s12911-017-0439-z)] [Medline: [28390396](https://pubmed.ncbi.nlm.nih.gov/28390396/)]
210. Nichols GA, Desai J, Elston LJ, Lawrence JM, O'Connor PJ, Pathak RD, et al. Construction of a multisite DataLink using electronic health records for the identification, surveillance, prevention, and management of diabetes mellitus: the SUPREME-DM project. *Prev Chronic Dis* 2012;9:E110 [FREE Full text] [Medline: [22677160](https://pubmed.ncbi.nlm.nih.gov/22677160/)]
211. Nichols GA, Schroeder EB, Karter AJ, Gregg EW, Desai J, Lawrence JM, et al. Trends in diabetes incidence among 7 million insured adults, 2006-2011: the SUPREME-DM project. *Am J Epidemiol* 2015 Jan 1;181(1):32-39. [doi: [10.1093/aje/kwu255](https://doi.org/10.1093/aje/kwu255)] [Medline: [25515167](https://pubmed.ncbi.nlm.nih.gov/25515167/)]
212. Young JB, Gauthier-Loiselle M, Bailey RA, Manceur AM, Lefebvre P, Greenberg M, et al. Development of predictive risk models for major adverse cardiovascular events among patients with type 2 diabetes mellitus using health insurance claims data. *Cardiovasc Diabetol* 2018 Aug 24;17(1):118 [FREE Full text] [doi: [10.1186/s12933-018-0759-z](https://doi.org/10.1186/s12933-018-0759-z)] [Medline: [30143045](https://pubmed.ncbi.nlm.nih.gov/30143045/)]
213. Pacheco JA, Thompson W, Kho A. Automatically detecting problem list omissions of type 2 diabetes cases using electronic medical records. *AMIA Annu Symp Proc* 2011;2011:1062-1069 [FREE Full text] [Medline: [22195167](https://pubmed.ncbi.nlm.nih.gov/22195167/)]
214. Pantalone KM, Misra-Hebert AD, Hobbs TM, Wells BJ, Kong SX, Chagin K, et al. Effect of glycemic control on the Diabetes Complications Severity Index score and development of complications in people with newly diagnosed type 2 diabetes. *J Diabetes* 2018 Mar;10(3):192-199. [doi: [10.1111/1753-0407.12613](https://doi.org/10.1111/1753-0407.12613)] [Medline: [28976724](https://pubmed.ncbi.nlm.nih.gov/28976724/)]
215. Paul SK, Shaw JE, Montvida O, Klein K. Weight gain in insulin-treated patients by body mass index category at treatment initiation: new evidence from real-world data in patients with type 2 diabetes. *Diabetes Obes Metab* 2016 Dec;18(12):1244-1252. [doi: [10.1111/dom.12761](https://doi.org/10.1111/dom.12761)] [Medline: [27502528](https://pubmed.ncbi.nlm.nih.gov/27502528/)]
216. Richesson RL, Rusincovitch SA, Wixted D, Batch BC, Feinglos MN, Miranda ML, et al. A comparison of phenotype definitions for diabetes mellitus. *J Am Med Inform Assoc* 2013 Dec;20(e2):e319-e326 [FREE Full text] [doi: [10.1136/amiajnl-2013-001952](https://doi.org/10.1136/amiajnl-2013-001952)] [Medline: [24026307](https://pubmed.ncbi.nlm.nih.gov/24026307/)]
217. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet* 2010 Apr 09;86(4):560-572 [FREE Full text] [doi: [10.1016/j.ajhg.2010.03.003](https://doi.org/10.1016/j.ajhg.2010.03.003)] [Medline: [20362271](https://pubmed.ncbi.nlm.nih.gov/20362271/)]
218. Rodgers LR, Weedon MN, Henley WE, Hattersley AT, Shields BM. Cohort profile for the MASTERMIND study: using the Clinical Practice Research Datalink (CPRD) to investigate stratification of response to treatment in patients with type 2 diabetes. *BMJ Open* 2017 Oct 12;7(10):e017989 [FREE Full text] [doi: [10.1136/bmjopen-2017-017989](https://doi.org/10.1136/bmjopen-2017-017989)] [Medline: [29025846](https://pubmed.ncbi.nlm.nih.gov/29025846/)]

219. Schroeder EB, Donahoo WT, Goodrich GK, Raebel MA. Validation of an algorithm for identifying type 1 diabetes in adults based on electronic health record data. *Pharmacoepidemiol Drug Saf* 2018 Oct;27(10):1053-1059 [FREE Full text] [doi: [10.1002/pds.4377](https://doi.org/10.1002/pds.4377)] [Medline: [29292555](https://pubmed.ncbi.nlm.nih.gov/29292555/)]
220. Sharma M, Petersen I, Nazareth I, Coton SJ. An algorithm for identification and classification of individuals with type 1 and type 2 diabetes mellitus in a large primary care database. *Clin Epidemiol* 2016;8:373-380 [FREE Full text] [doi: [10.2147/CLEP.S113415](https://doi.org/10.2147/CLEP.S113415)] [Medline: [27785102](https://pubmed.ncbi.nlm.nih.gov/27785102/)]
221. Spratt SE, Pereira K, Granger BB, Batch BC, Phelan M, Pencina M, et al. Assessing electronic health record phenotypes against gold-standard diagnostic criteria for diabetes mellitus. *J Am Med Inform Assoc* 2017 May 01;24(e1):e121-e128 [FREE Full text] [doi: [10.1093/jamia/ocw123](https://doi.org/10.1093/jamia/ocw123)] [Medline: [27616701](https://pubmed.ncbi.nlm.nih.gov/27616701/)]
222. Teltsch DY, Fazeli Farsani S, Swain RS, Kaspers S, Huse S, Cristaldi C, et al. Development and validation of algorithms to identify newly diagnosed type 1 and type 2 diabetes in pediatric population using electronic medical records and claims data. *Pharmacoepidemiol Drug Saf* 2019 Feb;28(2):234-243. [doi: [10.1002/pds.4728](https://doi.org/10.1002/pds.4728)] [Medline: [30677205](https://pubmed.ncbi.nlm.nih.gov/30677205/)]
223. Tu K, Manuel D, Lam K, Kavanagh D, Mitiku TF, Guo H. Diabetics can be identified in an electronic medical record using laboratory tests and prescriptions. *J Clin Epidemiol* 2011 May;64(4):431-435. [doi: [10.1016/j.jclinepi.2010.04.007](https://doi.org/10.1016/j.jclinepi.2010.04.007)] [Medline: [20638237](https://pubmed.ncbi.nlm.nih.gov/20638237/)]
224. Upadhyaya SG, Murphree DH, Ngufor CG, Knight AM, Cronk DJ, Cima RR, et al. Automated Diabetes Case Identification Using Electronic Health Record Data at a Tertiary Care Facility. *Mayo Clin Proc Innov Qual Outcomes* 2017 Jul;1(1):100-110 [FREE Full text] [doi: [10.1016/j.mayocpiqo.2017.04.005](https://doi.org/10.1016/j.mayocpiqo.2017.04.005)] [Medline: [30225406](https://pubmed.ncbi.nlm.nih.gov/30225406/)]
225. Wei W, Leibson CL, Ransom JE, Kho AN, Caraballo PJ, Chai HS, et al. Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus. *J Am Med Inform Assoc* 2012;19(2):219-224 [FREE Full text] [doi: [10.1136/amiajnl-2011-000597](https://doi.org/10.1136/amiajnl-2011-000597)] [Medline: [22249968](https://pubmed.ncbi.nlm.nih.gov/22249968/)]
226. Wiese AD, Roumie CL, Buse JB, Guzman H, Bradford R, Zalimeni E, et al. Performance of a computable phenotype for identification of patients with diabetes within PCORnet: The Patient-Centered Clinical Research Network. *Pharmacoepidemiol Drug Saf* 2019 May;28(5):632-639 [FREE Full text] [doi: [10.1002/pds.4718](https://doi.org/10.1002/pds.4718)] [Medline: [30680840](https://pubmed.ncbi.nlm.nih.gov/30680840/)]
227. Williams BA, Geba D, Cordova JM, Shetty SS. A risk prediction model for heart failure hospitalization in type 2 diabetes mellitus. *Clin Cardiol* 2020 Mar;43(3):275-283 [FREE Full text] [doi: [10.1002/clc.23298](https://doi.org/10.1002/clc.23298)] [Medline: [31837035](https://pubmed.ncbi.nlm.nih.gov/31837035/)]
228. Wysham CH, Lefebvre P, Pilon D, Lafeuille M, Emond B, Kamstra R, et al. An investigation into the durability of glycemic control in patients with type II diabetes initiated on canagliflozin or sitagliptin: A real-world analysis of electronic medical records. *J Diabetes Complications* 2019 Feb;33(2):140-147 [FREE Full text] [doi: [10.1016/j.jdiacomp.2018.10.016](https://doi.org/10.1016/j.jdiacomp.2018.10.016)] [Medline: [30554838](https://pubmed.ncbi.nlm.nih.gov/30554838/)]
229. Yang F, Ma Q, Liu J, Ma B, Guo M, Liu F, et al. Prevalence and major risk factors of type 2 diabetes mellitus among adult psychiatric inpatients from 2005 to 2018 in Beijing, China: a longitudinal observational study. *BMJ Open Diabetes Res Care* 2020 Mar;8(1) [FREE Full text] [doi: [10.1136/bmjdr-2019-000996](https://doi.org/10.1136/bmjdr-2019-000996)] [Medline: [32139600](https://pubmed.ncbi.nlm.nih.gov/32139600/)]
230. Yue X, Wu J, Ruan Z, Wolden ML, Li L, Lin Y. The Burden of Hypoglycemia in Patients With Insulin-Treated Diabetes Mellitus in China: Analysis of Electronic Medical Records From 4 Tertiary Hospitals. *Value Health Reg Issues* 2020 May;21:17-21. [doi: [10.1016/j.vhri.2019.06.003](https://doi.org/10.1016/j.vhri.2019.06.003)] [Medline: [31634792](https://pubmed.ncbi.nlm.nih.gov/31634792/)]
231. Zheng L, Wang Y, Hao S, Shin AY, Jin B, Ngo AD, et al. Web-based Real-Time Case Finding for the Population Health Management of Patients With Diabetes Mellitus: A Prospective Validation of the Natural Language Processing-Based Algorithm With Statewide Electronic Medical Records. *JMIR Med Inform* 2016 Nov 11;4(4):e37 [FREE Full text] [doi: [10.2196/medinform.6328](https://doi.org/10.2196/medinform.6328)] [Medline: [27836816](https://pubmed.ncbi.nlm.nih.gov/27836816/)]
232. Zheng T, Xie W, Xu L, He X, Zhang Y, You M, et al. A machine learning-based framework to identify type 2 diabetes through electronic health records. *Int J Med Inform* 2017 Dec;97:120-127 [FREE Full text] [doi: [10.1016/j.ijmedinf.2016.09.014](https://doi.org/10.1016/j.ijmedinf.2016.09.014)] [Medline: [27919371](https://pubmed.ncbi.nlm.nih.gov/27919371/)]
233. Zhong VW, Obeid JS, Craig JB, Pfaff ER, Thomas J, Jaacks LM, et al. An efficient approach for surveillance of childhood diabetes by type derived from electronic health record data: the SEARCH for Diabetes in Youth Study. *J Am Med Inform Assoc* 2016 Nov;23(6):1060-1067 [FREE Full text] [doi: [10.1093/jamia/ocv207](https://doi.org/10.1093/jamia/ocv207)] [Medline: [27107449](https://pubmed.ncbi.nlm.nih.gov/27107449/)]
234. Zhong VW, Pfaff ER, Beavers DP, Thomas J, Jaacks LM, Bowlby DA, Search for Diabetes in Youth Study Group. Use of administrative and electronic health record data for development of automated algorithms for childhood diabetes case ascertainment and type classification: the SEARCH for Diabetes in Youth Study. *Pediatr Diabetes* 2014 Dec;15(8):573-584 [FREE Full text] [doi: [10.1111/pedi.12152](https://doi.org/10.1111/pedi.12152)] [Medline: [24913103](https://pubmed.ncbi.nlm.nih.gov/24913103/)]
235. Agrawal S, Kremsdorf R, Uysal S, Fredette ME, Topor LS. Nephrolithiasis: A complication of pediatric diabetic ketoacidosis. *Pediatr Diabetes* 2018 Mar;19(2):329-332. [doi: [10.1111/pedi.12559](https://doi.org/10.1111/pedi.12559)] [Medline: [28737266](https://pubmed.ncbi.nlm.nih.gov/28737266/)]
236. Cahn A, Altaras T, Agami T, Liran O, Touaty CE, Drahy M, et al. Validity of diagnostic codes and estimation of prevalence of diabetic foot ulcers using a large electronic medical record database. *Diabetes Metab Res Rev* 2019 Feb;35(2):e3094. [doi: [10.1002/dmrr.3094](https://doi.org/10.1002/dmrr.3094)] [Medline: [30378240](https://pubmed.ncbi.nlm.nih.gov/30378240/)]
237. Dong Y, Gao W, Zhang L, Wei J, Hammar N, Cabrera CS, et al. Patient characteristics related to metabolic disorders and chronic complications in type 2 diabetes mellitus patients hospitalized at the Qingdao Endocrine and Diabetes Hospital

- from 2006 to 2012 in China. *Diab Vasc Dis Res* 2017 Jan;14(1):24-32 [FREE Full text] [doi: [10.1177/1479164116675489](https://doi.org/10.1177/1479164116675489)] [Medline: [27941053](https://pubmed.ncbi.nlm.nih.gov/27941053/)]
238. DuBrava S, Mardekian J, Sadosky A, Bienen EJ, Parsons B, Hopps M, et al. Using Random Forest Models to Identify Correlates of a Diabetic Peripheral Neuropathy Diagnosis from Electronic Health Record Data. *Pain Med* 2017 Dec 01;18(1):107-115. [doi: [10.1093/pm/pnw096](https://doi.org/10.1093/pm/pnw096)] [Medline: [27252307](https://pubmed.ncbi.nlm.nih.gov/27252307/)]
239. Lee CS, Lee AY, Baughman D, Sim D, Akelere T, Brand C, UK DR EMR Users Group. The United Kingdom Diabetic Retinopathy Electronic Medical Record Users Group: Report 3: Baseline Retinopathy and Clinical Features Predict Progression of Diabetic Retinopathy. *Am J Ophthalmol* 2017 Aug;180:64-71 [FREE Full text] [doi: [10.1016/j.ajo.2017.05.020](https://doi.org/10.1016/j.ajo.2017.05.020)] [Medline: [28572062](https://pubmed.ncbi.nlm.nih.gov/28572062/)]
240. Martín-Merino E, Fortuny J, Rivero-Ferrer E, García-Rodríguez LA. Incidence of retinal complications in a cohort of newly diagnosed diabetic patients. *PLoS One* 2014;9(6):e100283 [FREE Full text] [doi: [10.1371/journal.pone.0100283](https://doi.org/10.1371/journal.pone.0100283)] [Medline: [24963628](https://pubmed.ncbi.nlm.nih.gov/24963628/)]
241. Song X, Waitman LR, Hu Y, Yu ASL, Robins D, Liu M. Robust clinical marker identification for diabetic kidney disease with ensemble feature selection. *J Am Med Inform Assoc* 2019 Mar 01;26(3):242-253. [doi: [10.1093/jamia/ocy165](https://doi.org/10.1093/jamia/ocy165)] [Medline: [30602020](https://pubmed.ncbi.nlm.nih.gov/30602020/)]
242. VanderWeele J, Pollack T, Oakes DJ, Smyrniotis C, Illuri V, Vellanki P, et al. Validation of data from electronic data warehouse in diabetic ketoacidosis: Caution is needed. *J Diabetes Complications* 2018 Jul;32(7):650-654. [doi: [10.1016/j.jdiacomp.2018.05.004](https://doi.org/10.1016/j.jdiacomp.2018.05.004)] [Medline: [29903409](https://pubmed.ncbi.nlm.nih.gov/29903409/)]
243. Abhyankar S, Demner-Fushman D, Callaghan FM, McDonald CJ. Combining structured and unstructured data to identify a cohort of ICU patients who received dialysis. *J Am Med Inform Assoc* 2014;21(5):801-807 [FREE Full text] [doi: [10.1136/amiajnl-2013-001915](https://doi.org/10.1136/amiajnl-2013-001915)] [Medline: [24384230](https://pubmed.ncbi.nlm.nih.gov/24384230/)]
244. Afzal Z, Schuemie MJ, van Blijderveen JC, Sen EF, Sturkenboom MCJM, Kors JA. Improving sensitivity of machine learning methods for automated case identification from free-text electronic medical records. *BMC Med Inform Decis Mak* 2013 Mar 02;13:30 [FREE Full text] [doi: [10.1186/1472-6947-13-30](https://doi.org/10.1186/1472-6947-13-30)] [Medline: [23452306](https://pubmed.ncbi.nlm.nih.gov/23452306/)]
245. Alcober-Morte L, Barrio-Ruiz C, Parellada-Esquius N, Subirana I, Comín-Colet J, Grau M, et al. Heart failure admission across glomerular filtration rate categories in a community cohort of 125,053 individuals over 60 years of age. *Hypertens Res* 2019 Dec;42(12):2013-2020. [doi: [10.1038/s41440-019-0315-6](https://doi.org/10.1038/s41440-019-0315-6)] [Medline: [31477871](https://pubmed.ncbi.nlm.nih.gov/31477871/)]
246. Broder A, Mowrey WB, Izmirly P, Costenbader KH. Validation of Systemic Lupus Erythematosus Diagnosis as the Primary Cause of Renal Failure in the US Renal Data System. *Arthritis Care Res (Hoboken)* 2017 Apr;69(4):599-604 [FREE Full text] [doi: [10.1002/acr.22972](https://doi.org/10.1002/acr.22972)] [Medline: [27390299](https://pubmed.ncbi.nlm.nih.gov/27390299/)]
247. Crawford DC, Bailey JNC, Miskimen K, Miron P, McCauley JL, Sedor JR, et al. Somatic T-cell Receptor Diversity in a Chronic Kidney Disease Patient Population Linked to Electronic Health Records. *AMIA Jt Summits Transl Sci Proc* 2018;2017:63-71 [FREE Full text] [Medline: [29888042](https://pubmed.ncbi.nlm.nih.gov/29888042/)]
248. Ernecoff NC, Wessell KL, Hanson LC, Lee AM, Shea CM, Dusetzina SB, et al. Electronic Health Record Phenotypes for Identifying Patients with Late-Stage Disease: a Method for Research and Clinical Application. *J Gen Intern Med* 2019 Dec;34(12):2818-2823 [FREE Full text] [doi: [10.1007/s11606-019-05219-9](https://doi.org/10.1007/s11606-019-05219-9)] [Medline: [31396813](https://pubmed.ncbi.nlm.nih.gov/31396813/)]
249. Fraccaro P, van der Veer S, Brown B, Prosperi M, O'Donoghue D, Collins GS, et al. An external validation of models to predict the onset of chronic kidney disease using population-based electronic health records from Salford, UK. *BMC Med* 2016 Jul 12;14:104 [FREE Full text] [doi: [10.1186/s12916-016-0650-2](https://doi.org/10.1186/s12916-016-0650-2)] [Medline: [27401013](https://pubmed.ncbi.nlm.nih.gov/27401013/)]
250. Hao S, Fu T, Wu Q, Jin B, Zhu C, Hu Z, et al. Estimating One-Year Risk of Incident Chronic Kidney Disease: Retrospective Development and Validation Study Using Electronic Medical Record Data From the State of Maine. *JMIR Med Inform* 2017 Jul 26;5(3):e21 [FREE Full text] [doi: [10.2196/medinform.7954](https://doi.org/10.2196/medinform.7954)] [Medline: [28747298](https://pubmed.ncbi.nlm.nih.gov/28747298/)]
251. Kitsos A, Peterson GM, Jose MD, Khanam MA, Castelino RL, Radford JC. Variation in Documenting Diagnosable Chronic Kidney Disease in General Medical Practice: Implications for Quality Improvement and Research. *J Prim Care Community Health* 2019;10:2150132719833298 [FREE Full text] [doi: [10.1177/2150132719833298](https://doi.org/10.1177/2150132719833298)] [Medline: [30879383](https://pubmed.ncbi.nlm.nih.gov/30879383/)]
252. Koynier JL, Adhikari R, Edelson DP, Churpek MM. Development of a Multicenter Ward-Based AKI Prediction Model. *Clin J Am Soc Nephrol* 2016 Nov 07;11(11):1935-1943 [FREE Full text] [doi: [10.2215/CJN.00280116](https://doi.org/10.2215/CJN.00280116)] [Medline: [27633727](https://pubmed.ncbi.nlm.nih.gov/27633727/)]
253. Magvanjav O, Cooper-DeHoff RM, McDonough CW, Gong Y, Segal MS, Hogan WR, et al. Antihypertensive therapy prescribing patterns and correlates of blood pressure control among hypertensive patients with chronic kidney disease. *J Clin Hypertens (Greenwich)* 2019 Jan;21(1):91-101 [FREE Full text] [doi: [10.1111/jch.13429](https://doi.org/10.1111/jch.13429)] [Medline: [30427124](https://pubmed.ncbi.nlm.nih.gov/30427124/)]
254. Malas MS, Wish J, Moorthi R, Grannis S, Dexter P, Duke J, et al. A comparison between physicians and computer algorithms for form CMS-2728 data reporting. *Hemodial Int* 2017 Jan;21(1):117-124. [doi: [10.1111/hdi.12445](https://doi.org/10.1111/hdi.12445)] [Medline: [27353890](https://pubmed.ncbi.nlm.nih.gov/27353890/)]
255. Malas MS, Wish J, Moorthi R, Grannis S, Dexter P, Duke J, et al. A comparison between physicians and computer algorithms for form CMS-2728 data reporting. *Hemodial Int* 2017 Jan;21(1):117-124. [doi: [10.1111/hdi.12445](https://doi.org/10.1111/hdi.12445)] [Medline: [27353890](https://pubmed.ncbi.nlm.nih.gov/27353890/)]
256. Meyers JL, Candrilli SD, Kovacs B. Type 2 diabetes mellitus and renal impairment in a large outpatient electronic medical records database: rates of diagnosis and antihyperglycemic medication dose adjustment. *Postgrad Med* 2011 May;123(3):133-143. [doi: [10.3810/pgm.2011.05.2291](https://doi.org/10.3810/pgm.2011.05.2291)] [Medline: [21566423](https://pubmed.ncbi.nlm.nih.gov/21566423/)]

257. Nadkarni GN, Gottesman O, Linneman JG, Chase H, Berg RL, Farouk S, et al. Development and validation of an electronic phenotyping algorithm for chronic kidney disease. *AMIA Annu Symp Proc* 2014;2014:907-916 [FREE Full text] [Medline: 25954398]
258. Robertson LM, Denadai L, Black C, Fluck N, Prescott G, Simpson W, et al. Is routine hospital episode data sufficient for identifying individuals with chronic kidney disease? A comparison study with laboratory data. *Health Informatics J* 2016 Jun;22(2):383-396 [FREE Full text] [doi: 10.1177/1460458214562286] [Medline: 25552482]
259. Salvador-González B, Rodríguez-Latre LM, Güell-Miró R, Álvarez-Funes V, Sanz-Ródenas H, Tovillas-Morán FJ. Estimation of glomerular filtration rate by MDRD-4 IDMS and CKD-EPI in individuals of 60 years of age or older in primary care. *Nefrologia* 2013;33(4):552-563 [FREE Full text] [doi: 10.3265/Nefrologia.pre2013.Apr.11929] [Medline: 23897188]
260. Schroeder EB, Powers JD, O'Connor PJ, Nichols GA, Xu S, Desai JR, SUPREME-DM Study Group. Prevalence of chronic kidney disease among individuals with diabetes in the SUPREME-DM Project, 2005-2011. *J Diabetes Complications* 2015 Jul;29(5):637-643. [doi: 10.1016/j.jdiacomp.2015.04.007] [Medline: 25936953]
261. Semler MW, Rice TW, Shaw AD, Siew ED, Self WH, Kumar AB, et al. Identification of Major Adverse Kidney Events Within the Electronic Health Record. *J Med Syst* 2016 Jul;40(7):167 [FREE Full text] [doi: 10.1007/s10916-016-0528-z] [Medline: 27234478]
262. Sun AZ, Shu Y, Harrison TN, Hever A, Jacobsen SJ, O'Shaughnessy MM, et al. Identifying Patients with Rare Disease Using Electronic Health Record Data: The Kaiser Permanente Southern California Membranous Nephropathy Cohort. *Perm J* 2020;24 [FREE Full text] [doi: 10.7812/TPP/19.126] [Medline: 32069207]
263. Anand V, Hyun C, Khan QM, Hall C, Hessefort N, Sonnenberg A, et al. Identification and Fibrosis Staging of Hepatitis C Patients Using the Electronic Medical Record System. *J Clin Gastroenterol* 2016 Sep;50(8):664-669. [doi: 10.1097/MCG.0000000000000519] [Medline: 26974763]
264. Atiemo K, Skaro A, Maddur H, Zhao L, Montag S, VanWagner L, et al. Mortality Risk Factors Among Patients With Cirrhosis and a Low Model for End-Stage Liver Disease Sodium Score ( $\leq 15$ ): An Analysis of Liver Transplant Allocation Policy Using Aggregated Electronic Health Record Data. *Am J Transplant* 2017 Oct;17(9):2410-2419 [FREE Full text] [doi: 10.1111/ajt.14239] [Medline: 28226199]
265. Bateman-Steel CR, Smedley EJ, Kong M, Ferson MJ. Hepatitis C enhanced surveillance: results from a southeastern Sydney pilot program. *Public Health Res Pract* 2015 Mar 30;25(2):e2521520 [FREE Full text] [doi: 10.17061/phrp2521520] [Medline: 25848738]
266. Corey KE, Kartoun U, Zheng H, Shaw SY. Development and Validation of an Algorithm to Identify Nonalcoholic Fatty Liver Disease in the Electronic Medical Record. *Dig Dis Sci* 2016 Mar;61(3):913-919 [FREE Full text] [doi: 10.1007/s10620-015-3952-x] [Medline: 26537487]
267. Cuthbert JA, Arslanlar S, Yepuri J, Montrose M, Ahn CW, Shah JP. Predicting short-term mortality and long-term survival for hospitalized US patients with alcoholic hepatitis. *Dig Dis Sci* 2014 Jul;59(7):1594-1602 [FREE Full text] [doi: 10.1007/s10620-013-3020-3] [Medline: 24445730]
268. Fialoke S, Malarstig A, Miller MR, Dumitriu A. Application of Machine Learning Methods to Predict Non-Alcoholic Steatohepatitis (NASH) in Non-Alcoholic Fatty Liver (NAFL) Patients. *AMIA Annu Symp Proc* 2018;2018:430-439 [FREE Full text] [Medline: 30815083]
269. Kaplan DE, Dai F, Aytaman A, Baytarian M, Fox R, Hunt K, VOCAL Study Group. Development and Performance of an Algorithm to Estimate the Child-Turcotte-Pugh Score From a National Electronic Healthcare Database. *Clin Gastroenterol Hepatol* 2015 Dec;13(13):2333-41.e1 [FREE Full text] [doi: 10.1016/j.cgh.2015.07.010] [Medline: 26188137]
270. Kartoun U, Corey KE, Simon TG, Zheng H, Aggarwal R, Ng K, et al. The MELD-Plus: A generalizable prediction risk score in cirrhosis. *PLoS One* 2017;12(10):e0186301 [FREE Full text] [doi: 10.1371/journal.pone.0186301] [Medline: 29069090]
271. Lai JC, Wong GL, Yip TC, Tse Y, Lam KL, Lui GC, et al. Chronic Hepatitis B Increases Liver-Related Mortality of Patients With Acute Hepatitis E: A Territorywide Cohort Study From 2000 to 2016. *Clin Infect Dis* 2018 Sep 28;67(8):1278-1284. [doi: 10.1093/cid/ciy234] [Medline: 30265321]
272. Loomis AK, Kabadi S, Preiss D, Hyde C, Bonato V, St Louis M, et al. Body Mass Index and Risk of Nonalcoholic Fatty Liver Disease: Two Electronic Health Record Prospective Studies. *J Clin Endocrinol Metab* 2016 Mar;101(3):945-952 [FREE Full text] [doi: 10.1210/jc.2015-3444] [Medline: 26672639]
273. Lu M, Chacra W, Rabin D, Rupp LB, Trudeau S, Li J, et al. Validity of an automated algorithm using diagnosis and procedure codes to identify decompensated cirrhosis using electronic health records. *Clin Epidemiol* 2017;9:369-376 [FREE Full text] [doi: 10.2147/CLEP.S136134] [Medline: 28744162]
274. Nguyen TA, DeShazo JP, Thacker LR, Puri P, Sanyal AJ. The Worsening Profile of Alcoholic Hepatitis in the United States. *Alcohol Clin Exp Res* 2016 Jun;40(6):1295-1303 [FREE Full text] [doi: 10.1111/acer.13069] [Medline: 27147285]
275. Singal AG, Rahimi RS, Clark C, Ma Y, Cuthbert JA, Rockey DC, et al. An automated model using electronic medical record data identifies patients with cirrhosis at high risk for readmission. *Clin Gastroenterol Hepatol* 2013 Oct;11(10):1335-1341.e1 [FREE Full text] [doi: 10.1016/j.cgh.2013.03.022] [Medline: 23591286]

276. Xu Y, Li N, Lu M, Myers RP, Dixon E, Walker R, et al. Development and validation of method for defining conditions using Chinese electronic medical record. *BMC Med Inform Decis Mak* 2016 Aug 20;16:110 [FREE Full text] [doi: [10.1186/s12911-016-0348-6](https://doi.org/10.1186/s12911-016-0348-6)] [Medline: [27542973](https://pubmed.ncbi.nlm.nih.gov/27542973/)]
277. Jamil K, Huang X, Lovelace B, Pham AT, Lodaya K, Wan G. The burden of illness of hepatorenal syndrome (HRS) in the United States: a retrospective analysis of electronic health records. *J Med Econ* 2019 May;22(5):421-429. [doi: [10.1080/13696998.2019.1580201](https://doi.org/10.1080/13696998.2019.1580201)] [Medline: [30724682](https://pubmed.ncbi.nlm.nih.gov/30724682/)]
278. Koola JD, Davis SE, Al-Nimri O, Parr SK, Fabbri D, Malin BA, et al. Development of an automated phenotyping algorithm for hepatorenal syndrome. *J Biomed Inform* 2018 Apr;80:87-95 [FREE Full text] [doi: [10.1016/j.jbi.2018.03.001](https://doi.org/10.1016/j.jbi.2018.03.001)] [Medline: [29530803](https://pubmed.ncbi.nlm.nih.gov/29530803/)]
279. Lin C, Karlson EW, Dligach D, Ramirez MP, Miller TA, Mo H, et al. Automatic identification of methotrexate-induced liver toxicity in patients with rheumatoid arthritis from the electronic medical record. *J Am Med Inform Assoc* 2015 May;22(e1):e151-e161 [FREE Full text] [doi: [10.1136/amiajnl-2014-002642](https://doi.org/10.1136/amiajnl-2014-002642)] [Medline: [25344930](https://pubmed.ncbi.nlm.nih.gov/25344930/)]
280. Wing K, Bhaskaran K, Smeeth L, van Staa TP, Klungel OH, Reynolds RF, et al. Optimising case detection within UK electronic health records: use of multiple linked databases for detecting liver injury. *BMJ Open* 2016 Sep 02;6(9):e012102 [FREE Full text] [doi: [10.1136/bmjopen-2016-012102](https://doi.org/10.1136/bmjopen-2016-012102)] [Medline: [27591023](https://pubmed.ncbi.nlm.nih.gov/27591023/)]
281. Feller DJ, Zucker J, Yin MT, Gordon P, Elhadad N. Using Clinical Notes and Natural Language Processing for Automated HIV Risk Assessment. *J Acquir Immune Defic Syndr* 2018 Feb 01;77(2):160-166 [FREE Full text] [doi: [10.1097/QAI.0000000000001580](https://doi.org/10.1097/QAI.0000000000001580)] [Medline: [29084046](https://pubmed.ncbi.nlm.nih.gov/29084046/)]
282. Felsen UR, Bellin EY, Cunningham CO, Zingman BS. Development of an electronic medical record-based algorithm to identify patients with unknown HIV status. *AIDS Care* 2014;26(10):1318-1325 [FREE Full text] [doi: [10.1080/09540121.2014.911813](https://doi.org/10.1080/09540121.2014.911813)] [Medline: [24779521](https://pubmed.ncbi.nlm.nih.gov/24779521/)]
283. Goetz MB, Hoang T, Kan VL, Rimland D, Rodriguez-Barradas M. Development and validation of an algorithm to identify patients newly diagnosed with HIV infection from electronic health records. *AIDS Res Hum Retroviruses* 2014 Jul;30(7):626-633. [doi: [10.1089/AID.2013.0287](https://doi.org/10.1089/AID.2013.0287)] [Medline: [24564256](https://pubmed.ncbi.nlm.nih.gov/24564256/)]
284. McInnes DK, Shimada SL, Midboe AM, Nazi KM, Zhao S, Wu J, et al. Patient Use of Electronic Prescription Refill and Secure Messaging and Its Association With Undetectable HIV Viral Load: A Retrospective Cohort Study. *J Med Internet Res* 2017 Feb 15;19(2):e34 [FREE Full text] [doi: [10.2196/jmir.6932](https://doi.org/10.2196/jmir.6932)] [Medline: [28202428](https://pubmed.ncbi.nlm.nih.gov/28202428/)]
285. Paul DW, Neely NB, Clement M, Riley I, Al-Hegelan M, Phelan M, et al. Development and validation of an electronic medical record (EMR)-based computed phenotype of HIV-1 infection. *J Am Med Inform Assoc* 2018 Feb 01;25(2):150-157 [FREE Full text] [doi: [10.1093/jamia/ocx061](https://doi.org/10.1093/jamia/ocx061)] [Medline: [28645207](https://pubmed.ncbi.nlm.nih.gov/28645207/)]
286. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004 Jan 1;32(Database issue):D267-D270 [FREE Full text] [doi: [10.1093/nar/gkh061](https://doi.org/10.1093/nar/gkh061)] [Medline: [14681409](https://pubmed.ncbi.nlm.nih.gov/14681409/)]
287. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. *Stud Health Technol Inform* 2006;121:279-290. [Medline: [17095826](https://pubmed.ncbi.nlm.nih.gov/17095826/)]
288. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc* 2010;17(5):507-513 [FREE Full text] [doi: [10.1136/jamia.2009.001560](https://doi.org/10.1136/jamia.2009.001560)] [Medline: [20819853](https://pubmed.ncbi.nlm.nih.gov/20819853/)]
289. Torii M, Waghlikar K, Liu H. Using machine learning for concept extraction on clinical documents from multiple data sources. *J Am Med Inform Assoc* 2011;18(5):580-587 [FREE Full text] [doi: [10.1136/amiajnl-2011-000155](https://doi.org/10.1136/amiajnl-2011-000155)] [Medline: [21709161](https://pubmed.ncbi.nlm.nih.gov/21709161/)]
290. Doering TA, Plapp F, Crawford JM. Establishing an evidence base for critical laboratory value thresholds. *Am J Clin Pathol* 2014 Dec;142(5):617-628. [doi: [10.1309/AJCPDIOFYZ4UNWEQ](https://doi.org/10.1309/AJCPDIOFYZ4UNWEQ)] [Medline: [25319976](https://pubmed.ncbi.nlm.nih.gov/25319976/)]
291. Kidney Disease: Improving Global Outcomes (KDIGO) CKD-MBD Work Group. KDIGO clinical practice guideline for the diagnosis, evaluation, prevention, and treatment of Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD). *Kidney Int Suppl* 2009 Aug(113):S1-130. [doi: [10.1038/ki.2009.188](https://doi.org/10.1038/ki.2009.188)] [Medline: [19644521](https://pubmed.ncbi.nlm.nih.gov/19644521/)]
292. Levey AS, Coresh J, Balk E, Kausz AT, Levin A, Steffes MW, National Kidney Foundation. National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Ann Intern Med* 2003 Jul 15;139(2):137-147. [Medline: [12859163](https://pubmed.ncbi.nlm.nih.gov/12859163/)]
293. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, eMERGE Team. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics* 2011 Jan 26;4:13 [FREE Full text] [doi: [10.1186/1755-8794-4-13](https://doi.org/10.1186/1755-8794-4-13)] [Medline: [21269473](https://pubmed.ncbi.nlm.nih.gov/21269473/)]
294. McCue ME, McCoy AM. The Scope of Big Data in One Medicine: Unprecedented Opportunities and Challenges. *Front Vet Sci* 2017;4:194 [FREE Full text] [doi: [10.3389/fvets.2017.00194](https://doi.org/10.3389/fvets.2017.00194)] [Medline: [29201868](https://pubmed.ncbi.nlm.nih.gov/29201868/)]
295. Sweet LE, Moulaison HL. Electronic Health Records Data and Metadata: Challenges for Big Data in the United States. *Big Data* 2013 Dec;1(4):245-251. [doi: [10.1089/big.2013.0023](https://doi.org/10.1089/big.2013.0023)] [Medline: [27447257](https://pubmed.ncbi.nlm.nih.gov/27447257/)]
296. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015;10(3):e0118432 [FREE Full text] [doi: [10.1371/journal.pone.0118432](https://doi.org/10.1371/journal.pone.0118432)] [Medline: [25738806](https://pubmed.ncbi.nlm.nih.gov/25738806/)]



297. Reimer AP, Milinovich A, Madigan EA. Data quality assessment framework to assess electronic medical record data for use in research. *Int J Med Inform* 2016 Jul;90:40-47 [[FREE Full text](#)] [doi: [10.1016/j.ijmedinf.2016.03.006](https://doi.org/10.1016/j.ijmedinf.2016.03.006)] [Medline: [27103196](#)]
298. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc* 2012;19(1):54-60 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000376](https://doi.org/10.1136/amiajnl-2011-000376)] [Medline: [22037893](#)]
299. Denaxas S, Gonzalez-Izquierdo A, Fitzpatrick N, Direk K, Hemingway H. Phenotyping UK Electronic Health Records from 15 Million Individuals for Precision Medicine: The CALIBER Resource. *Stud Health Technol Inform* 2019 Jul 04;262:220-223. [doi: [10.3233/SHTI190058](https://doi.org/10.3233/SHTI190058)] [Medline: [31349307](#)]
300. Birtwhistle RV. Canadian Primary Care Sentinel Surveillance Network: a developing resource for family medicine and public health. *Can Fam Physician* 2011 Oct;57(10):1219-1220 [[FREE Full text](#)] [Medline: [21998241](#)]

## Abbreviations

**CHF:** congestive heart failure

**eMERGE:** Electronic Medical Records and Genomics

**EMR:** electronic medical records

**ICD:** International Classification of Diseases

**ICPC:** International Classification of Primary Care

**ML:** machine learning

**NLP:** natural language processing

**PPV:** positive predictive value

**PRISMA-ScR:** Preferred Reporting Items for Systematic reviews and Meta-analyses extension for Scoping Reviews

*Edited by C Lovis; submitted 28.08.20; peer-reviewed by F Lau, L Kosowan; comments to author 20.09.20; revised version received 20.11.20; accepted 05.12.20; published 01.02.21*

*Please cite as:*

*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*

*URL: <https://medinform.jmir.org/2021/2/e23934>*

*doi: [10.2196/23934](https://doi.org/10.2196/23934)*

*PMID: [33522976](https://pubmed.ncbi.nlm.nih.gov/33522976/)*

©Seungwon Lee, Chelsea Doktorchik, Elliot Asher Martin, Adam Giles D'Souza, Cathy Eastwood, Abdel Aziz Shaheen, Christopher Naugler, Joon Lee, Hude Quan. Originally published in JMIR Medical Informatics (<http://medinform.jmir.org>), 01.02.2021. 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 Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.