

## **A VAC4EU Systematic Review to Summarize and Critically Appraise Existing Phenotype Libraries Using Electronic Health Records**

Sima Mohammadi<sup>1\*</sup>, Cori Campbell<sup>2,3\*</sup>, Miriam C.J.M. Sturkenboom<sup>1</sup>, Tiago A. Vaz<sup>1 †</sup>

1. Department of Data Science & Biostatistics, University Medical Center Utrecht (UMCU), Utrecht, the Netherlands

2. Department of Health Services Research and Policy, The London School of Hygiene and Tropical Medicine, London, UK

3. Nuffield Department of Medicine, University of Oxford, Oxford, UK

\* Joint first authors

† Corresponding author: Tiago A. Vaz

### **Keywords:**

Phenotype, Library, Algorithm, Electronic Health Records

**Abstract word: 300**

**Main text word: 3050**

## **Abstract**

### **Background**

Pharmacoepidemiology and population health studies using secondary analysis of electronic health care records (EHR) must define study variables through available electronic data. Defining a study variable starts with the identification of a phenotype, which is a defined set of criteria used to identify specific traits or medical conditions. In the real-world data perspective, a phenotype library is a collection of code lists or algorithms that standardize these sets of criteria. We conducted a systematic review of existing phenotype libraries to appraise their attributes, accessibility, interoperability, and portability.

### **Methods**

We systematically searched three databases (Scopus, PubMed, and Web of Science) until June 2024, to identify studies on key characteristics of phenotype libraries. The search combined MeSH terms related to "electronic health records," "phenotype algorithm," and "phenotype library". Study parameters extracted included: library size, vocabularies, phenotype construction tools, validation and library management process, and portability in different sites.

### **Findings**

Of 134 articles, 26 met eligibility criteria, leaving nine articles related to eight unique phenotype libraries including CALIBER (Health Data Research UK (HDR UK) Phenotype Library or CALIBER), Centralized Interactive Phenomics Resource (CIPHER), ClinicalCodes Library, Manitoba Centre for Health Policy (MCHP) Concept Dictionary, Observational Health Data Sciences and Informatics (OHDSI) ATLAS, Open CodeLists, Phenotype Execution and Modeling Architecture (PhEMA) Workbench, Phenotype KnowledgeBase (PheKB). These libraries varied largely in size and vocabularies. Each library created rule-based phenotypes, though OHDSI and CIPHER also utilized machine learning. All libraries are both human and machine-readable. Validation processes varied and were only applied to some libraries. All libraries utilized a web-based platform and met at least the minimum requirements for library management, including phenotype definitions, metadata (if applicable), and version control.

## **Interpretations**

We observed large variations in library features including phenotype construction. Transparency about phenotypes and creating computable phenotypes enhance portability and streamline the effective reuse of phenotypes for different systems.

## **Funding**

This investigation was supported by a Fellowship awarded by VAC4EU (Vaccine Collaboration for Europe) Phenotype Representation Model: An International and Streamlined Approach to Enhance RWE Studies (grant nr 2023/0001).

## **Research in context**

Evidence before this study.

Electronic health data have been used extensively in epidemiology and health data science research for decades, as they offer a wealth of detailed real-world data which may be used to address important evidence gaps. Importance of such data sources has been strongly highlighted following the COVID-19 pandemic, which saw a massive increase in the number of scientific investigations utilizing electronic data to rapidly produce evidence to guide health policy. Recent development of multiple phenotype libraries has presented an important advancement in the field. Libraries serve as repositories for the construction and re-use of phenotypes built with electronic health data including diagnostic codes, laboratory values and demographic information. To our knowledge, a systematic review to identify and describe all existing phenotype libraries has not been undertaken following the COVID-19 pandemic undertaken following the COVID-19 pandemic.

Added value of this study.

This study provides a comprehensive systematic review which identifies and describes all currently existing phenotype libraries. We summarize and compare phenotype construction processes, data sources, user interfaces, portability and algorithm validation practices across 8 individual phenotype libraries. We highlight how these libraries facilitate robust and transparency, open scientific practices in digital health research, and identify potential opportunities for innovation . This systematic review serves as an important benchmark study, providing a central documentation and description of phenotype libraries built to date.

Implications of all the available evidence.

The use of phenotype libraries and collaboratively constructed and validated phenotypes in health research may greatly improve the robustness and impact of health research. We describe how libraries can be currently be used to improve research practice, as well as how existing libraries

## Introduction

Digitalizing individual-level patient data is one of the most impactful health technology innovations in recent decades, enabling the storage, sharing, and re-use of medical information for research purposes to draw valuable insights<sup>1</sup>. Electronic Health Records (EHRs) may include all electronic aspects of a patient's medical background, such as previous medical history, diagnoses, medication history and prescriptions, procedures, and vaccination records<sup>2</sup>. In recent years, there has been increasing emphasis on EHR use to generate real-world evidence (RWE) for various medical research fields, while pharmacoepidemiologists have been drawing RWE insights from EHR data for decades. RWE provides essential insights into medical interventions' efficacy, safety, and utility in everyday clinical settings. However, since these data are not collected for research purposes but primarily collected for healthcare delivery, billing, and public health monitoring, they may not be fit for the purpose of answering research questions. To overcome data quality & evidence issues important advances have been made, such as standardization of metadata of the data sources, data quality assessment tools, and increased transparency of protocols and reporting<sup>3-5</sup>. In clinical trials, variables (which have been specified pre-hoc) are observed for study participants to answer the research question. In contrast, in Real World Data (RWD)/RWE, study variables must be created using available data in the EHR systems requiring identification and transformation steps, meaning that transparency of data processing and analysis methods is crucial<sup>3</sup>.

A phenotype library is a curated collection of algorithms representing phenotypes using diagnostic codes, medications, and other criteria. These libraries contain transparent, standardized formats that are both machine- and human-readable<sup>6,7</sup>. In terms of transparency and reproducibility, it is necessary to report phenotypes in studies<sup>4</sup>. Phenotype libraries ensure support the need for transparency by incorporating metadata to help structure the various components of a phenotype, supporting the generation of robust evidence from RWD to guide clinical decision-making and health policy<sup>7</sup>. Reusing and sharing phenotypes becomes particularly important when working with EHR in multiple countries, which may operate in different healthcare systems, languages, and vocabularies. Therefore, phenotype libraries should capture data source-specific algorithms to reach standardized documentation of phenotype algorithms based on a common clinical definition.

Phenotype libraries must also be adaptable and undergo regular maintenance, as well as flexible to evolve with updates in clinical practice standards, the emergence of novel disease phenotypes, and advancements in medical knowledge<sup>8</sup>. Routine curation and updating are necessary to maintain the accuracy and relevance of

phenotypes<sup>9</sup>. A phenotype library should be developed by combining clinical expertise and scientific programming. Machine learning approaches using EHR data and expert knowledge have also been proposed to develop phenotyping models, potentially reducing the time and cost involved in manual development<sup>7</sup>.

Computable phenotypes are machine-readable definitions designed for programmatic execution using standardized workflows and logic. In addition, portability focuses on the ability to transfer and apply these phenotypes across different systems or environments. Designing phenotypes to be both computable and portable enhances their reusability<sup>10</sup>. This appraisal aims to describe the landscape of existing phenotype libraries that are available. We systematically review them and summarize their size, supported coding systems, the process for algorithm definitions, requirements in terms of knowledge and tools, maintenance, management, and portability through data models.

## **Methods**

### **Database search strategy**

This study is a systematic review following the Preferred Reporting Items for Systematic Reviews (PRISMA) requirements<sup>11</sup>. We systematically searched three literature databases (Scopus, PubMed, and Web of Science) up to June 2024 to identify studies that detailed the development, validation, and/or utilization of EHR-based phenotype libraries. Our search strategy combined terms related to "electronic health records," "phenotype algorithm," and "phenotype library" with both controlled vocabulary and free-text terms. Medical Subject Headings (MeSH) search terms used to search literature databases are available in Table 1. Search terms were constructed by two authors (TV, CC), and articles were screened and selected for inclusion by three authors (authors (SM, TV, CC). Articles identified for full-text review were double-reviewed, each assessed by two authors independently, and conflicts regarding article inclusion were resolved collectively. Our inclusion criteria focused on articles that provide descriptions of the phenotype creation and curation processes, including the methodologies and tools used for phenotype development. The web-based interface of each library was accessed to gather data on the library size, vocabulary, and other relevant information, where possible. We excluded conference proceedings, editorials, unpublished or grey literature, and articles that focused on genotypes. We excluded articles that reported on libraries including only specific disease sub-phenotypes, for example, libraries restricted to only cardiovascular disease. Additional hand-searching of reference lists from included articles was undertaken to identify potential articles for inclusion. No restrictions for language were applied to database search results.

### **Data extraction and reporting**

Summary characteristics for included articles were extracted by three authors (SM, TV, CC). Characteristics extracted for each article included: phenotype library name and size, medical vocabularies included, knowledge used to construct phenotype (medical, programming, or other), whether the phenotype algorithms are readable by a human and/or computer, phenotype validation process, library maintenance, portability, management, and user interface. Library characteristics were updated as of October 2024.

### **Role of the funding source**

Funding for this study was provided by the Vaccine Monitoring Collaboration for Europe (VAC4EU), a European not-for-profit association that monitors vaccine coverage, safety, and effectiveness by leveraging real-world data. The study funder approved the study design but did not play a role in the collection, analysis, and interpretation of data. The funders were allowed to review and comment on the manuscript, but the authors retained full control of the final decisions, in compliance with the requirements of the ENCePP Code of Conduct<sup>12</sup>.

## Results

### Study characteristics

Our systematic search identified a total of 134 articles (**Figure 1**). After deduplication, 112 articles were available for title and abstract screening. Following title/abstract screening, 26 articles remained for full-text reviews. 17 were excluded as they were not EHR-based phenotype libraries. A total of nine articles were identified for final inclusion following a full-text review, with each article reporting on a unique phenotype library<sup>13-21</sup>. **Table 2** includes summary characteristics for each included phenotype library.

In total, eight individual phenotype libraries were identified, including the Centralized Interactive Phenomics Resource (CIPHER)<sup>14,21</sup>, the ClinicalCodes<sup>20</sup>, the Health Data Research UK (HDR UK) Phenotype Library also referred to as CALIBER for its data platform<sup>17</sup>, the Manitoba Centre for Health Policy (MCHP) Concept Dictionary<sup>15</sup>, the Open CodeLists<sup>16</sup>, the Observational Health Data Sciences and Informatics (OHDSI) ATLAS<sup>18</sup>, the Phenotype KnowledgeBase (PheKB)<sup>13</sup>, and the Phenotype Execution and Modeling Architecture (PhEMA) Workbench<sup>19</sup>. The PhEMA workbench includes collections of pre-defined phenotypes PheKB phenotypes and is primarily designed to facilitate the construction and sharing of these phenotypes, rather than functioning as a library<sup>19,22</sup>. We report a diversity of crossway libraries in size and features, with the number of phenotypes or concepts per library ranging from <100, as in PheKB and PhEMA Workbench<sup>22</sup>, to thousands of phenotypes, as in CIPHER<sup>23</sup> and CALIBER<sup>24</sup>.

Medical vocabularies used to identify coded concepts varied within libraries, with most libraries including ICD10, ICD10CM, ICD9CM, SNOMED, procedural, and medicinal codes. Most vocabularies belong to the Unified Medical Language System (UMLS) vocabulary collection. Some libraries also included regional vocabularies<sup>25</sup>, free-text<sup>14</sup>, and laboratory tests<sup>15,18,19</sup>. Types of phenotypes included in the libraries vary, encompassing not only diseases or syndromes but also extending to biomarkers, demographics and lifestyle factors, treatments, and medical procedures.



## **Phenotype construction process**

The processes by which phenotypes are constructed varies by library. All libraries support the creation of rule-based phenotype construction, where diagnostic codes and relevant criteria are combined using Boolean operators ('AND,' 'OR,' 'NOT'). These algorithms are available in human- and computer-readable formats and include comprehensive metadata. Metadata offer details on the data elements and algorithm construction. All libraries support phenotype construction using one (or more) tools or programming languages. For instance, CALIBER translates phenotype algorithms into Structured Query Language (SQL) queries and offers an open-source R library for manipulating clinical terminologies<sup>17</sup>. The MCHP library utilizes SAS codes to implement measures associated with defined concepts, enabling the reuse of research measures in future projects<sup>15</sup>. The Open CodeLists Platform utilizes the electronic health records query language (ehrQL), a specialized language designed for querying data within the OpenSAFELY database and extracting datasets from it. For instance, in its deployment with TPP (Thrombosis with Thrombocytopenia), ehrQL allows consistent queries to be executed in different healthcare research databases, even if their underlying structures differ<sup>16</sup>. CIPHER<sup>14</sup> and OHDSI ATLAS<sup>18,25,26</sup> phenotype libraries support the use of machine learning (ML) for phenotype discovery. OHDSI employs the APHRODITE R package, which enables the incorporation of machine learning methods to support the construction of phenotype classifiers using training datasets<sup>18</sup>. PheKB supports the use of Natural Language Processing (NLP) for the identification of relevant information/features from free-text fields; However, PheKB itself does not directly implement or perform NLP. Instead, researchers can integrate NLP tools into their workflows when using PheKB to extract relevant data from unstructured text within medical records. While most electronic phenotype algorithms on PheKB are presented in human-readable formats requiring manual translation into executable code for local clinical repositories, some include computable representations, such as those developed with Konstanz Information Miner (KNIME ) which is a tool aid in developing computable phenotypes, enhancing the re-usability and portability<sup>13</sup>. The PhEMA Workbench, powers Clinical Quality Language (CQL) to create human-readable designations for phenotypes, effectively capturing inclusion and exclusion criteria. PhEMA Workbench is planning to enhance the system's capabilities by incorporating NLP (using CQL4NLP47) and ML<sup>17</sup>.

Most libraries allow external users (i.e., individuals not belonging to the institution that created and maintains the phenotype library) to contribute and utilize phenotypes<sup>14,23,24,26</sup>. External users can register as contributors and must adhere to contribution guidelines. They are obliged to follow a standardized template when submitting a phenotype, requiring a clear definition of the variables in the studies, the logic or algorithm used to define it, a list of relevant codes, and any available validation processes. After submission, phenotypes typically go through a peer review process conducted by the library's maintainers<sup>26</sup>. HDR UK Phenotype Library also allows contributors to upload pre-existing code lists<sup>24</sup>. All libraries provide advanced search functionality, enabling users to locate specific phenotypes efficiently.

### **Phenotype validation**

All libraries deployed a form of phenotype validation, with processes differing across libraries (**Table 2**). Most libraries validated a subset of phenotypes rather than all phenotypes available in each library. Phenotype validation was most often conducted by verifying individuals identified by specific phenotype algorithms against a clinically observed “gold standard”, often involving manual review of patient charts or medical records to ensure that the actual clinical diagnosis of the patients matches the standard definition of the phenotype in the library. CIPHER<sup>14</sup> reported positive predictive values (PPVs), sensitivities, and specificities of validated phenotypes against gold standards. CALIBER offers up to six validation methods including source concordance, case note review, consistency of risk factor-disease associations based on non-EHR studies, consistency with findings from previous research, genetic association consistency, and comparisons with external population<sup>17</sup>. In the MCHP phenotype library, several existing MCHP concepts for specific diseases and procedures required redefinition and validation; however, detailed information about the validation methods was not available<sup>15</sup>. OHDSI uses APHRODITE (Automated Phenotype Routine for Observational Definition, Identification, Training, and Evaluation) to facilitate the creation and validation of phenotype definitions, and has developed PheValuator as a framework to evaluate the phenotype algorithm performance. PheKB performs validation for some phenotypes, at different sites and by reviewing electronic medical records against gold standards, evaluating algorithms using metrics such as PPV and sensitivity<sup>13</sup>. The PhEMA Workbench includes a validation protocol designed to facilitate the development and validation of phenotypes against clinical gold standards, with an accompanying an SQL script for randomly selecting cases and non-cases from the output

generated during the execution step. The script extracts relevant data and a data entry form for reviewers to document which criteria were satisfied by each cohort study. To date in PhEMA Workbench thrombotic events have been validated across three sites<sup>19</sup> (Table 3).

### **Library portability, management, and maintenance**

All libraries include minimal portability whereby code lists may be downloaded for library users (Table 2). Most libraries are built and managed by the institutions at which they were built; the one exception was the ClinicalCodes library<sup>20,27</sup>, which has now migrated to the HDR UK library<sup>17</sup>. All libraries reported version control measures that minimally include phenotype authorship, versioning and dating. Libraries are designed to facilitate phenotype updating. The CIPHER workflow provides phenotype definitions, code lists, metadata, authorship, creation date, and code list version control, cataloging versions for reuse. Phenotypes are identified by the phenotype name or author name. The CIPHERplatform allows sharing of resources like programming code or links to public repositories, aiding accessibility and collaboration amongst a primary audience including principal scientists, project managers, clinicians, statisticians, and data scientists, and a secondary audience including healthcare administrators utilising phenotypes for policy and guideline development<sup>21</sup>. ClinicalCodes serves as a repository for code lists, providing citation details if linked to an article, along with associated metadata. Both code lists and metadata can be downloaded or uploaded, and users have the option to comments on the code lists. ClinicalCodes employs JSON files to link metadata and code lists from published articles, with the rClinicalCodes R package facilitating automated downloads of code lists. Its technical framework incorporates PostgreSQL, Python, and JavaScript<sup>20</sup>. CALIBER platform enables storing phenotype definition, algorithms, metadata and tools, and allows linkage to studies using the platform<sup>24</sup>. The MCHP library provides detailed operational definitions and programming code for measures used in MCHP studies and promotes consistency in documenting research methodologies<sup>15</sup>. The OpenSAFELY workflow supports maintaining code lists, defining criteria for their creation, sharing, and maintaining version control. Each code list is given a URL for citation, facilitating reproducible practices in code list usage and integration with GitHub which enables complete analytic pipeline testing, code sharing, review, versioning, and ensuring reusability<sup>16</sup>. OHDSI Atlas offers an interface for defining logical criteria for cohorts, and cohort definitions may be exported in JSON format for sharing with other Observational Medical Outcomes Partnership (OMOP) users<sup>19</sup>. PheKB supports the cataloging, sharing, and uploading of phenotypes along with metadata, inclusion and exclusion criteria used

for phenotype definitions, and other supporting documents. It provides interfaces for discussion of phenotype level as well as searching phenotypes or metadata using keywords, housing both descriptive documents and executable logic that allows researchers to create, share, and transport phenotypes effectively. PheKB utilizes the Drupal content management system to manage phenotype metadata, optimize algorithm workflows, and enhance search functionality. For technical validation of data, a tool written as a Ruby web service has been integrated via a custom Drupal module to identify structural and content errors in data dictionaries and data files associated with a given phenotype. Users upload data dictionaries for a phenotype and files containing implementation results, and data structure and values must abide by rules established in the data dictionary. PheKB tools also enable the creation of computable phenotypes which have improved portability<sup>13</sup>.

### **Library interface**

All libraries offer web-based interfaces for accessing phenotypes. CIPHER offers a platform for developing, validating, and sharing phenotype algorithms, integrated with various data sources and leveraging standardized vocabularies<sup>14</sup>. OHDSI uses ATLAS as an interface to design and execute phenotype definitions, using the OMOP Common Data Model<sup>18</sup>. CALIBER and OHDSI support a standard API, enabling interaction between different software applications for developers<sup>24,26</sup>. OpenSAFELY utilizes Docker, a secure analytics platform to establish reproducible computational environments<sup>16,28</sup>. PhEMA uses GitHub to manage its codebase, share tools, and facilitate collaboration on phenotype execution<sup>22</sup>.

## Discussion

In this systematic review, we have identified and described eight phenotype libraries that support the construction and use of phenotypes for secondary use of electronic health data. Phenotype libraries allow users to identify disease occurrence/presence, laboratory parameters, demographics, medicine, and treatment, for the creation of study variables in a transparent manner. These libraries capture and allow the sharing of phenotype algorithm metadata and also facilitate data harmonization and reuse of phenotypes for various systems and frameworks<sup>13,14</sup>. This review provides evidence that various vocabularies were utilized for phenotype construction, with most phenotypes comprising standard vocabularies, including ICD10, ICD9, SNOMED, READ, and OPCS vocabularies. Construction of rule-based algorithmic phenotypes requires clinical knowledge of concepts, and building computable phenotypes requires programming knowledge to code and implement phenotype identification in a data source. Utilizing phenotype algorithms and ensuring the completeness and validity of electronic information in the database require a combination of clinical expertise and programming skills.

We report that all libraries provide an interface or platform that facilitates the management of phenotypes by storing them in a structured and accessible format and enables users to efficiently search, retrieve, and utilize phenotypes along with their linked metadata. All the libraries meet the minimum requirements for portability such as using various standardized vocabularies, using metadata, and offering the potential for both human and computer readability. Notwithstanding these advanced features, ensuring accurate metadata continues to be imperative, as it describes the versioning aspects of a phenotype definition; automation may facilitate advances in making phenotype metadata more readily available<sup>8</sup>. Our findings align with the study from Chapman et al<sup>8</sup>, highlighting that phenotype definitions may exist in various formats including code lists, algorithms (incorporating logical connectives), complex data elements derived through NLP, and trained classifiers. High-quality phenotypes are defined by their reproducibility, portability, and validity, and these criteria may serve as a basis for evaluating libraries<sup>8</sup>. Our systematic review demonstrates that ClinicalCodes and PhEKB enable

discussion at both the code list and phenotype levels. Phenotype libraries should foster social interactions between authors and users, facilitating feedback and discussion<sup>8</sup>.

We have identified key areas for library innovation and improvement . These include the storage of detailed validation information, facilitating community feedback and recognition of contributors, supporting diverse content types such as phenotypes based on Natural Language Processing (NLP) or Machine Learning (ML), and ensuring the inclusion of fully-computable phenotype definitions<sup>29</sup>. Our study demonstrates that all libraries consist of computable phenotypes, enhancing the reusability of phenotypes. However, libraries may have several phenotype algorithms that are restricted to human readability<sup>13</sup>. A previous study by Kashyap et al<sup>18</sup> reported that using the APHRODITE tool to develop and share ML-based phenotype classifiers was more efficient rule-based definitions<sup>18</sup>, but sharing of such classifiers may be constrained by diversity in EHR data collection for different countries<sup>18</sup>.

The US Food and Drug Administration (FDA) recommends validation of study variables by patient chart review<sup>30</sup>. Our study highlights the diverse validation methods are employed across libraries, including the likes of chart reviews cross-EHR concordance utilized by CALIBER (comparing risk factor-disease associations across EHR and non-EHR studies<sup>17</sup>), and more automated methods such as PheValuator utilised by OHDSI<sup>18</sup>. Further investigation of variable validation methods may warrant further investigation to quantify precision and efficacy in validating phenotypes.<sup>31</sup>

## **Conclusion**

Major advancements have been made in recent decades with the development of multiple large-scale and accessible phenotype libraries that allow for transparency in the identification of study variables when re-using health care data sources. Many libraries are only partially validated, and therefore enhanced validation processes are essential to ensure the accuracy and reliability of data. Detailed algorithm metadata and fully computable phenotypes (supported by programming tools) may enhance portability and support the reuse of phenotypes for different systems and frameworks.. Collaborative efforts for transparency of phenotype algorithms across libraries are essential for producing robust results and fostering greater user engagement for creating larger and more standard libraries.

## Acknowledgments

We wish to acknowledge Judit Riera Arnau who contributed her expert knowledge of concepts.

## Declaration of interests

CC declares doctoral funding awarded from the University of Oxford and GSK UK (2019-2023). MCJMS is head of Data Science and Biostatistics Department at University Medical Center Utrecht, which conducts studies for the European Medicines Agency and several Vaccine manufacturers. All according to ENCEPP code of conduct.

## Author contributions

CC and TAV were responsible for the study conception and design. CC, TAV and SM were responsible for drafting search terms, carrying out systematic searches, screening and identifying articles for inclusion and data extraction. SM, CC, TAV and MCJMS were responsible for manuscript drafting and reviewing

## References

1. Kasemsap K. Mastering electronic health record in global health care. *Healthcare Ethics and Training: Concepts, Methodologies, Tools, and Applications*: IGI Global; 2017: 1091-111.
2. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *International journal of medical informatics* 2008; **77**(5): 291-304.
3. Wang SV, Schneeweiss S, Berger ML, et al. Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. *Pharmacoepidemiol Drug Saf* 2017; **26**(9): 1018-32.

4. Gini R, Pajouheshnia R, Gardarsdottir H, et al. Describing diversity of real world data sources in pharmacoepidemiologic studies: The DIVERSE scoping review. *Pharmacoepidemiology and drug safety* 2024; **33**(5): e5787.
5. Wang SV, Pottegård A, Crown W, et al. HARmonized Protocol Template to Enhance Reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: a good practices report of a joint ISPE/ISPOR task force. *Value in Health* 2022; **25**(10): 1663-72.
6. Robinson JR, Wei WQ, Roden DM, Denny JC. Defining Phenotypes from Clinical Data to Drive Genomic Research. *Annu Rev Biomed Data Sci* 2018; **1**: 69-92.
7. Almowil ZA, Zhou S-M, Brophy S. Concept libraries for automatic electronic health record based phenotyping: a review. *International Journal of Population Data Science* 2021; **6**(1).
8. Chapman M, Mumtaz S, Rasmussen LV, et al. Desiderata for the development of next-generation electronic health record phenotype libraries. *Gigascience* 2021; **10**(9): giab059.
9. Feder SL. Data quality in electronic health records research: quality domains and assessment methods. *Western journal of nursing research* 2018; **40**(5): 753-66.
10. Chapman M, Rasmussen LV, Pacheco JA, Curcin V. Phenoflow: A Microservice Architecture for Portable Workflow-based Phenotype Definitions. *AMIA Jt Summits Transl Sci Proc* 2021; **2021**: 142-51.
11. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021; **372**: n71.
12. Gini R, Fournie X, Dolk H, et al. The ENCePP Code of Conduct: A best practise for scientific independence and transparency in noninterventional postauthorisation studies. *Pharmacoepidemiol Drug Saf* 2019; **28**(4): 422-33.
13. Kirby JC, Speltz P, Rasmussen LV, et al. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. *Journal of the American Medical Informatics Association* 2016; **23**(6): 1046-52.
14. Honerlaw J, Ho Y-L, Fontin F, et al. Framework of the Centralized Interactive Phenomics Resource (CIPHER) standard for electronic health data-based phenomics knowledgebase. *Journal of the American Medical Informatics Association* 2023; **30**(5): 958-64.
15. Smith M, Turner K, Bond R, Kawakami T, Roos LL. The concept dictionary and glossary at MCHP: tools and techniques to support a population research data repository. *International Journal of Population Data Science* 2019; **4**(1).
16. Nab L, Schaffer AL, Hulme W, et al. OpenSAFELY: A platform for analysing electronic health records designed for reproducible research. *Pharmacoepidemiology and Drug Safety* 2024; **33**(6): e5815.
17. Denaxas S, Gonzalez-Izquierdo A, Direk K, et al. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. *Journal of the American Medical Informatics Association* 2019; **26**(12): 1545-59.
18. Kashyap M, Seneviratne M, Banda JM, et al. Development and Validation of Phenotype Classifiers across Multiple Sites in the Observational Health Sciences and Informatics (OHDSI) Network. *bioRxiv* 2019: 673418.
19. Brandt PS, Pacheco JA, Adekkanattu P, et al. Design and validation of a FHIR-based EHR-driven phenotyping toolbox. *Journal of the American Medical Informatics Association* 2022; **29**(9): 1449-60.
20. Springate DA, Kontopantelis E, Ashcroft DM, et al. ClinicalCodes: an online clinical codes repository to improve the validity and reproducibility of research using electronic medical records. *PLoS one* 2014; **9**(6): e99825.
21. Honerlaw J, Ho Y-L, Fontin F, et al. Centralized Interactive Phenomics Resource: an integrated online phenomics knowledgebase for health data users. *Journal of the American Medical Informatics Association* 2024; **31**(5): 1126-34.



22. PheMA. PheMA/phema-workbench-app. 2024. <https://github.com/PheMA/phema-workbench-app> (accessed October 2024).
23. Kelly Cho JH, Anne Ho, Francesca Fontin, Ashley Galloway, Jeff Gosian, Monika Maripuri, Michael Murray, Rahul Sangar, Tiffany Sim, Joanne Sordillo, Vidisha Tanukonda, Edward Zielinski. Centralized Interactive Phenomics Resource (CIPHER). 2024. <https://phenomics.va.ornl.gov/web/> (accessed October 2024).
24. Library HUP. HDR UK Phenotype Library. 2024. <https://phenotypes.healthdatagateway.org/> (accessed October 2024).
25. Rasmussen LV, Brandt PS, Jiang G, et al. Considerations for improving the portability of electronic health record-based phenotype algorithms. AMIA Annual Symposium Proceedings; 2019: American Medical Informatics Association; 2019. p. 755.
26. Rao G. PhenotypeLibrary: The OHDSI Phenotype Library. 2024. <https://ohdsi.github.io/PhenotypeLibrary/>. (accessed October 2024).
27. ClinicalCodes. ClinicalCodes.org. 2024. <https://clinicalcodes.rss.mhs.man.ac.uk/> (accessed October 2024).
28. Oxford BlfADSatUo. OpenCodelists. 2024. <https://www.phc.ox.ac.uk/research/covid-19/projects/opensafely-a-new-secure-analytics-platform-for-electronic-health-records> (accessed September 2024 2024).
29. Thayer DS, Mumtaz S, Elmessary MA, et al. Creating a next-generation phenotype library: the health data research UK Phenotype Library. *JAMIA Open* 2024; **7**(2).
30. Girman CJ, Ritchey ME, Lo Re V, 3rd. Real-world data: Assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products. *Pharmacoepidemiol Drug Saf* 2022; **31**(7): 717-20.
31. Joel Swerdel, Seng Chan You, Ray Chen & Patrick Ryan. Clinical Validity. The Book of OHDSI. Living document. *Observational Health Data Sciences and Informatics*; 2021. p. 16.4. Available from: <https://ohdsi.github.io/TheBookOfOhdsi/ClinicalValidity.html#phevaluator>.

## Tables and panels

**Table 1. Search terms used in a systematic search of Embase, MEDLINE, and Web of Science databases. Updated searches run to 16 May 2024.**

<b>Database</b>	<b>Search terms</b>	<b>Number of articles identified</b>
MEDLINE and Embase (searched via Ovid)		
PubMed	("phenotyp*" [All Fields]) AND (("librar*" [All Fields] AND "algorithm*" [All Fields]) OR "authoring" [All Fields]) AND ("electronic health records" [MeSH Terms] OR ("electronic" [All Fields] AND "health" [All Fields] AND "records" [All Fields]) OR	23

It is made available under a [CC-BY-NC-ND 4.0 International license](#) .

"electronic health records"[All Fields])

Scopus	( TITLE-ABS-KEY ( "phenotyp*" ) AND TITLE-ABS-KEY ( ( "librar*" AND "algorithm*" ) OR "authoring" ) AND TITLE-ABS-KEY ( "electronic health records" OR ( "electronic" AND "health" AND "records" ) OR "electronic health records" ) )	13
Web of Science	(ALL=(phenotyp*)) AND ((ALL=(librar*) AND ALL=(algorithm*)) OR ALL=(authoring)) AND (ALL=(electronic health records) OR (ALL=(electronic) AND ALL=(health) AND ALL=(records)) OR ALL=(electronic health records))	98

---

**Table 2. Summary characteristics of the existing phenotype libraries**

Library	CIPHER	MCHP	Open CodeLists	HDR UK	OHDSI ATLAS	PheKB	PhEMA	ClinicalCodes
Features		Concept Dictionary		phenotype library (CALIBER)			Workbench	
<b>Size</b>	6656 Phenotypes <sup>+</sup>	>373 Code lists	~ over 1700 Code lists	1093 phenotypes, 2159 Code lists	24,923 concepts under 88 phenotypes, 1317 phenotypes		32 phenotypes	670 Code lists, migrating to CALIBER
<b>Coding system</b>	ICD10CM, ICD10- PCS, CPT, Clinical stop code, free-text, LONIC, NDC	ICD10-CA, ICD9CM, CCI, Tariff codes, procedural, prescriptions and, laboratory records	SNOMEDCT, ICD10, READ CTV3, BNF	ICD9CM, ICD10,READ ICD11, OPCS-4, SNOMEDCT, BNF, GPRD product, OXMIS, Multilex codes, CTV3, ICPC2	ICD10, ICD10CN, ICD10CM, ICD9CM, SNOMED, LONIC, and 45 other vocabularies	ICD9, CPT, NLP lab and medication codes	RXNORM, ICD9CM, CPT,ICD9P, LONIC,	READ, OXMIS, SNOMED, CPRD product/medical code, BNF code, OXMIS, ICD-9, ICD-10
<b>Phenotype</b>	Rule-based	Rule-based	Rule-based	Rule-based	Rule-based	Rule-based	Rule-based	Rule-based

<b>creation</b>	phenotypes and ML	phenotypes	phenotypes	phenotypes	phenotypes and ML	phenotypes	phenotypes	
<b>Tools/languages used to create phenotypes</b>	ML	SAS	ehrQL,	R-packages , SQL	R, SAS, R-packages of APHRODITE	supports integrating NLP into the phenotype algorithm, KNIME for creating computable phenotypes	CQL, SQL,	
<b>Phenotype validation</b>	validation Performance evaluated using PPV, NPV, specificity and/or AUC	Phenotype validation perform for part of phenotypes	NA	Six validation methods	validation against the gold standard and automated methods by APHRODITE and PheValuator	Clinical validation against the gold standard.	Chart review validation against gold standard for one phenotype	NA
<b>Interface</b>	Web-based interface that offers a user-friendly and flexible	Web-based interfaces for accessing detailed health-related concepts	Web-based interfaces in the OpenSAFELY platform using Docker to generate	web-based phenotype library Uses standard API to facilitate research	OHDSI uses ATLAS as an interface supporting various data sources through the OMOP Common	Web-based interface using Drupal content management system	The Workbench web-based API processes the complete	Web-based interface, a virtual machine developed at the University of Manchester.

---

platform	and definitions.	reproducible	interoperability	Data Model. It uses	FHIR-based	Its technical
		computational		standard API to	phenotype	framework
		environments.		facilitate research	definition,	incorporates
				interoperability	translates it to	PostgreSQL,
					the	Python, and
					appropriate	JavaScript.
					representation,	
					using CQL on	
					OMOP	

~ A Code list is a list of specific codes (e.g., ICD, SNOMED, or other standardized vocabulary codes) that are used to represent specific components of a phenotype.

+ A phenotype could be more than a code lists phenotype. The phenotype could be an algorithm using logical rules and criteria that combine code lists.

AUC: area under the ROC curve;

BNF, British National Formulary

CCI, Canadian Classification of Health Interventions.

CPRD, Clinical Practice Research Datalink.

CPT, current procedural terminology;

CQL, Cassandra Query Language;

CTV3, Clinical Terms Version 3

ehrQL, Electronic Health Record Query Language,

GPRD, General Practice Research Database

HCPCS, Healthcare Common Procedure Coding System;

ICD, International Classification of Diseases;

Java, Java Programming Language;

KNIME, Konstanz Information Miner;

LOINC, Logical Observation Identifiers Names and Codes;

ML, Machine Learning;

NDC, National Drug Code;

NLP, Natural Language Processing;

OXMIS, Oxford Medical Information System

R, R Programming Language;

READ Codes, Read Clinical Classification

RXNORM, RxNorm (Standardized Nomenclature for Clinical Drugs)

SAS, Statistical Analysis System;

SNOMED C, Systematized Nomenclature of Medicine - Clinical Terms

SQL, Structured Query Language;

STATA, Stata Statistical Software;



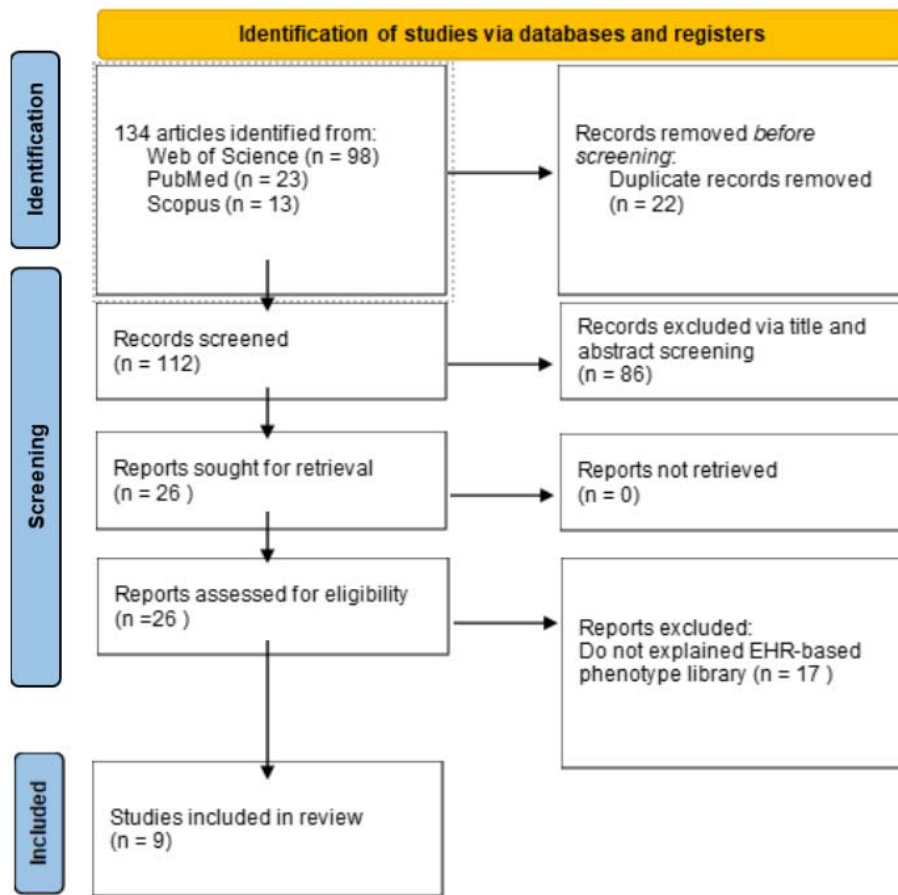
**Table 3. Description of phenotype validation schemes for included phenotype libraries.**

Library	Validation process	Performance metrics
<b>CIPHER</b>	1-Chart review as a gold standard 2-Comparing against patient-reported data- 3-Replicating a known association (such as replicating expected results from a genome-wide association study [GWAS]).	Sensitivity, PPV Specificity, AUC NPV
<b>HDR UK phenotype library or CALIBER</b>	CALIBER offers six validation methods: 1- Cross-EHR source concordance: quantified the percentage of cases identified in each source, quantified the overlap between sources, and evaluated per-source completeness and diagnostic validity 2- Case note review: review patient record 3- Consistency of risk factor–disease associations from non-EHR studies: produced and reported hazard ratios and CI of known risk factors from Cox proportional hazards models adjusted for age, sex, and other	PPV NPV Compare effect estimate

	<p>covariates and compared with consistent estimates of risk factor associations derived from non-EHR research-driven studies.</p> <p>4- Consistency with prior prognosis research: produced Kaplan–Meier cumulative incidence curves at appropriate time intervals and endpoints and stratified by EHR source and compare the survival patterns between patients diagnosed with the disease in a EHR primary care setting that never hospitalized against patients diagnosed in HES.</p> <p>5- External populations: comparing the consistency of results in the evaluation criteria in different data sources</p>	
<p><b>MCHP</b></p> <p><b>Concept</b></p> <p><b>Dictionary</b></p>	<p>In the MCHP phenotype library, a number of existing MCHP concepts for specific diseases and procedures required redefinition and validation; however, detailed information about the validation methods is not provided.</p>	<p>NA</p>
<p><b>PheKB</b></p>	<p>Clinical validation of phenotype algorithms at different sites using eMERGE data only for set of phenotypes.</p> <p>For technical validation of data, a validation tool written as a Ruby web</p>	<p>Sensitivity, Specificity, PPV, NPV</p>

	<p>service was integrated via a custom Drupal module to identify structural and content errors and in data dictionaries and data files associated with a given phenotype.</p> <p>Users will upload data dictionaries for a phenotype, and data files containing implementation results, and the system ensures that the data structure and values fit the rules established in the data dictionary.</p>	
<b>OHDSI ATLAS</b>	<p>1- comparing some or all persons in a defined cohort to a reference 'gold standard'</p> <p>2-PheValuator uses an automated process that builds probabilistic models to estimate the likelihood that patients have a condition by leveraging available healthcare data and known associations to create a "synthetic" gold standard, then evaluate the performance of the model by comparing the predicted probability to the test model<sup>31</sup>.</p>	<p>Sensitivity: <math>\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}</math></p> <p>Specificity: <math>\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}</math></p> <p>PPV: <math>\text{PPV} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}</math></p> <p>NPV: <math>\text{NPV} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Negatives}}</math></p>
<b>PhEMA Workbench</b>	<p>The PhEMA Workbench includes a validation protocol along with a SQL script to randomly select cases and non-cases, and extract all relevant data. The validation process involved manual chart reviews against gold standards.</p>	

<b>ClinicalCodes</b>	NA	
<b>Open-Codelists</b>	NA	



**Figure 1.** PRISMA Flow chart of selection of studies eligible for inclusion.

