
Database Notes

Development of a repository of computable phenotype definitions using the clinical quality language

Pascal S. Brandt ¹, Jennifer A. Pacheco², and Luke V. Rasmussen ²

¹Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, Washington, USA, and ²Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

Corresponding Author: Pascal S. Brandt, PhD, Department of Biomedical Informatics and Medical Education, University of Washington, Box 358047, Seattle, WA 98195, USA; psbrandt@uw.edu

Received 15 July 2021; Revised 21 September 2021; Editorial Decision 21 October 2021; Accepted 29 October 2021

ABSTRACT

Objective: The objective of this study is to create a repository of computable, technology-agnostic phenotype definitions for the purposes of analysis and automatic cohort identification.

Materials and Methods: We selected phenotype definitions from PheKB and excluded definitions that did not use structured data or were not used in published research. We translated these definitions into the Clinical Quality Language (CQL) and Fast Healthcare Interoperability Resources (FHIR) and validated them using code review and automated tests.

Results: A total of 33 phenotype definitions met our inclusion criteria. We developed 40 CQL libraries, 231 value sets, and 347 test cases. To support these test cases, a total of 1624 FHIR resources were created as test data.

Discussion and Conclusion: Although a number of challenges were encountered while translating the phenotypes into structured form, such as requiring specialized knowledge, or imprecise, ambiguous, and conflicting language, we have created a repository and a development environment that can be used for future research on computable phenotypes.

Key words: FHIR, CQL, EHR-driven phenotyping, cohort identification

Lay Summary

The process of conducting biomedical research almost always involves extracting patient data from electronic health record systems. Identifying the right patients in these systems is challenging due to the nature of clinical data and the complexity of diseases. Further, the process is usually manual, which is not very efficient and can lead to mistakes. In this work, we present a standard format that can be used to describe the patients of interest. This description can then be used to automatically extract the appropriate patient data. We create a database of 33 descriptions using this standard format and describe a method that can be used by anyone who wants to create additional standardized patient cohort descriptions. We hope that using the proposed format and methods will increase the rate at which clinical research can be conducted, and in turn, the rate of biomedical knowledge generation.

INTRODUCTION

Sets of criteria that are used to identify cohorts of patients for clinical research are referred to as *phenotype definitions*, or *phenotypes* for brevity. Phenotype definitions must be executed by implementers, often by manually translating textual descriptions of selection criteria into executable code, in order to identify patient cohorts. The heterogeneity of both the representation of the logic, as well as the data model that underlies the logic, is a key barrier to evaluating phenotype implementations to gain insight into the process of phenotype development.

Systems such as Informatics for Integrating Biology and the Bedside (i2b2)¹ and the Observational Health Data Sciences and Informatics (OHDSI) platform^{2,3} provide computable phenotypes that are bound to their respective data models. These computable representations are automatically translated into queries for a specific database system when the phenotype is used for cohort identification. However, i2b2 and OHDSI phenotypes cannot be directly shared between platforms, nor are they comparable without some translation.

Within the electronic clinical quality measure space, Clinical Quality Language (CQL; <https://cql.hl7.org/>) is used for the representation of similar criteria but is technology-agnostic. This means that it is not coupled to any specific software implementation. Furthermore, CQL has been shown to be a feasible logical expression language for representing clinically validated phenotype definitions.²⁻⁸ It supports a wide range of Boolean, temporal, aggregate, and other operations. The language is data model independent but does require the selection of a data model when writing CQL. It works out of the box with Fast Healthcare Interoperability Resources (FHIR; <https://hl7.org/FHIR/>), which is widely used and has recently become the legally required standard for clinical data exchange in the United States. Additionally, the Common Data Model Harmonization project⁹ provides mappings from FHIR to many other common healthcare data models, maximizing potential utilization of CQL-based phenotype definitions.

OBJECTIVES

In this work, we developed a database of phenotype definitions represented in a structured, unambiguous, computable, technology-agnostic standardized format. This representation would allow automated computational analysis and cohort identification against data platforms that support FHIR, and CQL or for which another mapping exists. We additionally provide a suite of test cases that, together with the provided testing configuration, can be used to validate the correctness of the phenotype definitions. Targeted to informaticians and research data analysts, we provide these phenotypes to the clinical research informatics community as an initial repository of computable technology-agnostic phenotype definitions, which we hope we and others will extend over time, using the same methods and tools and development environment.

MATERIALS AND METHODS

Data set

We selected phenotype definitions from the PheKB phenotype repository,¹⁰ as it is the most mature and widely used in the United States. PheKB was initiated in 2012 and has had phenotypes contributed from many projects and collaborative groups, most notably the electronic Medical Records and Genomics Network.¹¹⁻¹³ The reposi-

tory is continuously growing, with over 100 phenotypes in various stages of development.

We extracted all public phenotype definitions available on May 22, 2020. We automatically selected from the list of publicly available phenotypes those with a status of “Final.” From this collection, we reviewed each definition and excluded any phenotype entry that (1) did not make use of structured data [ie, was entirely natural language processing (NLP)-based], (2) that was a generic repository for submitting data, or (3) was not used in a published research study. These criteria were chosen to ensure that our final data set consisted only of completed and clinically validated phenotype definitions. For each phenotype that met these criteria, we downloaded all files linked to the phenotype definition. Source code for this step of the process is available on GitHub (<https://github.com/PheMA/phekb-export>).

Translation

Two authors (PSB and LVR) independently translated each of the phenotype definitions using the available metadata and artifacts downloaded from PheKB. One author was primarily responsible for the translation of each phenotype into CQL and FHIR, but the authors were not entirely blinded. Group discussion amongst all authors was used to confirm interpretation of phenotype definitions that were ambiguous.

Standard terminologies such as the International Classification of Diseases versions 9 (ICD-9) and 10 (ICD-10), Current Procedural Terminology (CPT), Logical Observation Identifiers Names and Codes (LOINC), and RxNorm were used for structured data and were represented using FHIR *ValueSet* resources. These terminologies were usually specified in the phenotype definitions, but where they were not, we used the recommended default terminologies from the FHIR standard. We developed an open-source tool to translate value sets in various formats into FHIR resources, and built an interface to allow web-based interaction with the tool (<https://github.com/PheMA/terminology-manager>). The tool supports Comma Separated Values files, as well as concept sets exported from the OHDSI platform, into *ValueSet* resources. It also supports searching, inspecting, and importing value sets directly from the Value Set Authority Center (VSAC),¹⁴ using the VSAC FHIR server.

For each phenotype, we created a single CQL library that contained the logic required to identify matching patients. Logic shared between phenotypes was authored in shared libraries that were imported using the CQL *include* operator. The shared libraries were iteratively refined and expanded (with already developed phenotypes refactored to incorporate new library functions) during the course of the project.

We did not implement NLP logic, as there is no widely accepted standard representation or implementation of this type of logic. To our knowledge, there is currently no way to natively express NLP constructs using FHIR or CQL, although this is an active area of research.^{15,16}

We adopted a number of conventions for the standards-based representation. First, in this work, we only represent phenotype cases, and not controls, suspected cases, or subtypes. Case definitions usually contain the most, and most varied, criteria; thus, they serve as a good basis for analysis or extension. We adopted the convention of creating a CQL statement in each library called “Case,” which represents the entry point for evaluating the phenotype definition. Additionally, unless explicitly stated otherwise, we modeled

drugs based on their RxNorm ingredient name and lab values based on the highest ranked appropriate LOINC code.

Development environment

We made use of several open-source tools during the phenotype translation process and published this development environment on GitHub (<https://github.com/PheMA/phekb-phenotypes>). We used Visual Studio Code (<https://code.visualstudio.com>) as our primary CQL authoring environment, and for syntax highlighting we used the *language-cql* plugin (<https://github.com/Jonnokc/Clinical-Quality-Language>).

To translate CQL into the equivalent machine-readable representation, known as the Expression Logical Model (ELM), we used the reference implementation of the CQL to ELM translator (https://github.com/cqframework/clinical_quality_language). For testing, we used the CQL Testing Framework (CTF) developed by the Agency for Healthcare Research and Quality (<https://github.com/AHRQ-CDS/CQL-Testing-Framework>). The CTF provides a mechanism to specify test data, which are materialized as FHIR resources, using a simple YAML file. The CTF also provides a configurable test runner, which can run a specific CQL library against the test data generated by the YAML specification, and assert that the results match what is expected. This allows authors to carefully create test data to make sure phenotypes correctly identify potentially tricky edge and corner cases.

Validation

We used 2 methods to ensure that our CQL phenotype definitions were correctly translated from the artifacts available in PheKB. First, each phenotype was translated by a single author, and then verified using a code review process. The primary author created a pull request on GitHub (a way of isolating code for a given purpose, in this case representing a single phenotype definition), and the second author reviewed the code to make sure it accurately represented the phenotype definition as described in PheKB.

Secondly, we used an approach from software engineering called test-driven development to ensure that our translations of phenotype logic and value sets were correct. We made use of the CTF to implement this approach. In addition to allowing the CQL author to express both test cases and FHIR data using YAML, the CTF integrates with the Mocha JavaScript testing framework (<https://mochajs.org>) in order to evaluate phenotype logic using the given data to assert that results produced are correct. This evaluation is done using the ELM representation of the phenotype, and the open-source JavaScript CQL engine (<https://github.com/cqframework/cql-execution>).

All tests were run automatically on each code commit to ensure no regressions were introduced. The full development and validation pipeline is shown in [Figure 1](#).

RESULTS

At the time of our extraction, there were a total of 71 publicly available phenotype definitions in PheKB with a status of “Final.” We excluded 2 definitions that were not actually phenotypes. One was used as a placeholder to publish new value sets, and one was the description of a risk model. We excluded 3 more that used only NLP criteria. Finally, from the remaining phenotypes, we identified only those with associated publications. This selection process resulted in 33 total phenotype definitions and is illustrated in [Figure 2](#).

We created a total of 40 CQL libraries—one for each phenotype, and 7 helper libraries, totaling 3327 lines of CQL code. A total of 231 value sets were assembled, of which 216 were manually created and 15 were imported from VSAC. These value sets consist of 17 948 individual codes, of which 13,340 are unique. Additionally, 347 test cases were written that collectively contain 2044 test assertions. To support these test cases, 347 patients, 96 encounters, 101 procedures, 335 medication orders, 385 conditions, and 360 observations were manually created as FHIR resources using the CTF.

DISCUSSION

While building this repository, we observed numerous advantages to using popular healthcare-specific standards such as FHIR and CQL. Advantages include convenient conceptual models, mechanisms for verification, and the availability of tools, documentation and expertise to provide assistance during development. This repository will provide a collection of diverse, computable, verified, and standardized phenotype definitions that will aid automated analysis.

In addition, we note other benefits above our primary objective. First, the methods and tools we used are documented within the repository and can be adopted by other researchers and developers. Also, our use of a logical representation that is technology and data model independent may facilitate automated execution by allowing implementation sites to implement their own data providers for existing phenotype definitions. Similarly, the use of common standard terminologies may enable automatic mapping during local execution.

During this work, we also experienced first hand many of the challenges that face phenotype implementers. We encountered numerous occurrences of ambiguity, underspecify, and imprecise language. For example, the *Clopidogrel Poor Metabolizers* phenotype uses the phrase “within 30 days,” but does not specify whether the interval boundary should be inclusive or not, or whether 30 days both before or after the event should be considered. In each case, the primary CQL developer had to confer with the other authors in order to determine the exact semantics of the phenotype definition. Even then, we would occasionally rely on subjective decisions regarding the intent. This resulted in a considerable slowdown in implementation. We note that a benefit of having translated the narrative definition to CQL is that the up-front investment in time has removed the ambiguities for all subsequent users.

Additionally, some phenotype definitions relied heavily on domain or tribal knowledge not specified within the definition itself. This makes it difficult for non-clinicians or healthcare outsiders to replicate research or use existing phenotype definitions for new research. For example, the *High-Density Lipoproteins* phenotype requires that a cohort member have at least one “random glucose test,” but does not specify how these tests are to be identified. We also encountered contradictory criteria definitions, for example, the *Bone Scan Utilization* phenotype requires that a cohort member be both >35 and ≥ 35 years old. The creation of this repository demonstrates a step forward for these phenotypes. Although a formal representation may not eliminate all these issues, it would require phenotype authors to be more precise at the definition phase, which would reduce the cognitive load on implementers.

Although we did not formally track the amount of time to implement each phenotype, authoring a formal definition in CQL and providing confirmatory tests does require an additional investment in time and resources for phenotype authors. Furthermore, it

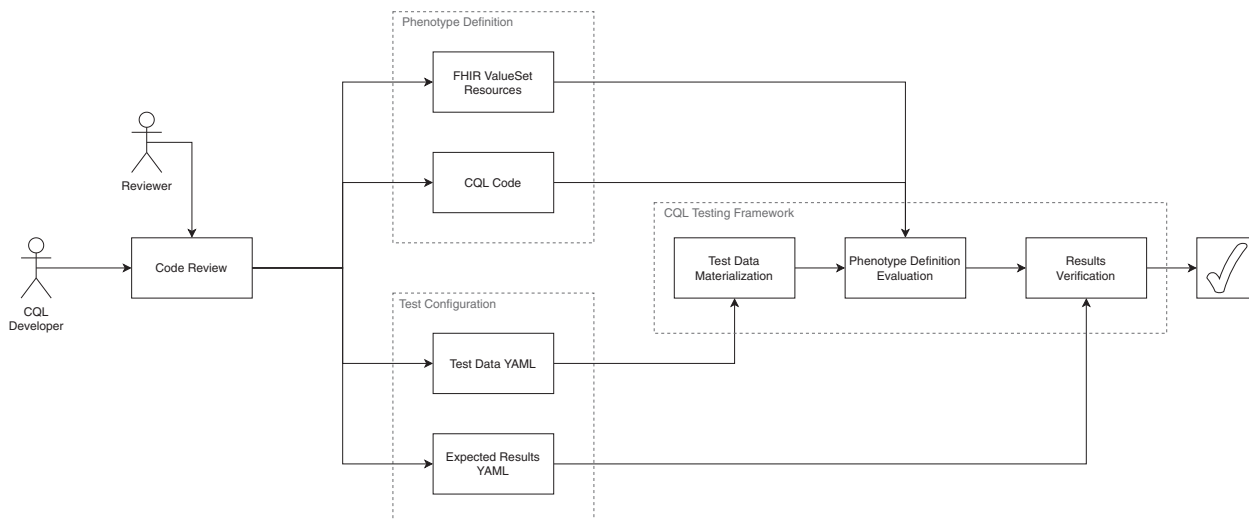


Figure 1. Phenotype definition development and verification pipeline.

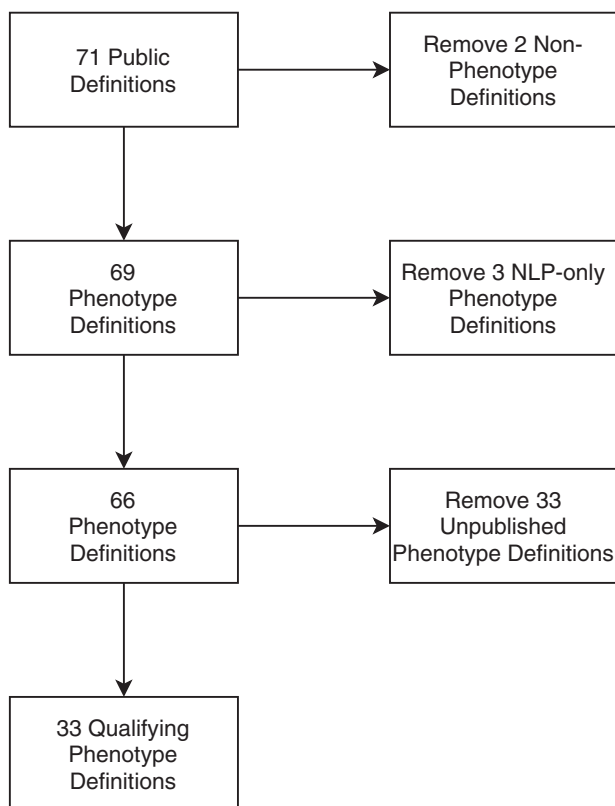


Figure 2. Phenotype definition selection process.

requires specialized informatics knowledge that has its own learning curve. However, we believe that the reduction in ambiguity benefits all phenotype authors (informaticians and research data analysts), and that when a phenotype is planned for reuse or broader dissemination, the time spent by the phenotype author in formalizing its representation using CQL has a cumulative payback each time the phenotype is reused.

This work demonstrates the realization of previous desiderata for computable phenotypes, including supporting human and computable formats, set operations and relational algebra, using

well-defined temporal relationships, using standard terminologies, and supporting standards-compliant interfaces to external software.¹⁷ Given the benefits of a concrete, unambiguous phenotype definition, we hope that phenotype repository managers will encourage the inclusion of computable definitions, in addition to providing APIs to allow integrating with their repositories.¹⁸

We acknowledge the following limitations of this work. First, given the subjective nature of interpreting narrative phenotype definitions, we cannot guarantee fidelity of the intended definition. The only way to determine semantic correctness would have been to reach out to the original phenotype definition authors, who may not have a definitive answer (given elapsed time from when some phenotypes were authored). Additionally, our CQL-based phenotype definitions were not clinically validated on actual datasets, although they are derived from clinically valid phenotype definitions.

CONCLUSION

This repository of structured phenotype definitions provides clear definitions of phenotype algorithms, represented in a format that facilitates automation of cohort identification within supported data platforms. We believe that the provided data set and development environment can be a resource for clinical informatics practitioners and researchers who want to study phenotype definitions or identify cohorts of patients for biomedical knowledge discovery.

In the future, we plan to evaluate the phenotype definitions in the translated dataset. This includes evaluating a single phenotype definition (represented using the standard proposed in this work) at 3 large academic medical centers, with performance being evaluated using manual chart review.

FUNDING

The Fulbright Foreign Student Program and the South African National Research Foundation (to PSB). In part by NIH grant R01GM105688 and by the NHGRI through the grant U01HG011169 (Northwestern University; to LVR and JAP).

AUTHOR CONTRIBUTIONS

All authors helped select the set of phenotypes to translate. PSB and LVR developed the representation standard, the development environment, and translated the PheKB phenotypes into the standard format. JAP was consulted to help resolve ambiguities during translation. PSB wrote the first draft of the manuscript, and all authors helped refine and edit the final version.

CONFLICT OF INTEREST STATEMENT

PSB is a consultant for Commure, Inc. LVR and JAP have no competing interests to disclose.

DATA AVAILABILITY

The raw data underlying this article are available in PheKB at <https://phekb.org/>, and the phenotypes available in GitHub are available at <https://github.com/PheMA/phekb-phenotypes>.

REFERENCES

- Murphy SN, Weber G, Mendis M, *et al.* Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010; 17 (2): 124–30.
- Hripcsak G, Duke JD, Shah NH, *et al.* Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015; 216: 574–8.
- Jiang G, Prud'Hommeaux E, Xiao G, Solbrig HR. Developing a semantic web-based framework for executing the clinical quality language using FHIR. *CEUR Workshop Proc* 2017; 2042: 1–5.
- Gulden C, Mate S, Prokosch H-U, Kraus S. Investigating the capabilities of FHIR search for clinical trial phenotyping. *Stud Health Technol Inform* 2018; 253: 3–7.
- Karim MR, Nguyen B-P, Zimmermann L, *et al.* A distributed analytics platform to execute FHIR-based phenotyping algorithms. In: *Proceedings of the 11th International Conference Semantic Web Applications and Tools for Healthcare and Life Sciences*; December 3–6 2018; Antwerp, Belgium.
- Meineke FA, Stäubert S, Löbe M, Uciteli A, Löffler M. Design and concept of the SMITH phenotyping pipeline. *Stud Health Technol Inform* 2019; 267: 164–72.
- McClure RC, Macumber CL, Skapik JL, Smith AM. Igniting harmonized digital clinical quality measurement through terminology, CQL, and FHIR. *Appl Clin Inform* 2020; 11 (1): 23–33.
- Brandt PS, Kiefer RC, Pacheco JA, *et al.* Toward cross-platform electronic health record-driven phenotyping using. *Clinical Quality Language. Learn Heal Syst* 2020; 4 (4): 1–9.
- HL7 International-Biomedical Research, and Regulation Work Group. Common Data Models Harmonization. <https://build.fhir.org/ig/HL7/cdmh/cdmh-overview.html>. Accessed March 26, 2021.
- Kirby JC, Speltz P, Rasmussen LV, *et al.* PheKB: A catalog and workflow for creating electronic phenotype algorithms for transportability. *J Am Med Inform Assoc* 2016; 23 (6): 1046–52.
- McCarty CA, Chisholm RL, Chute CG, *et al.*; eMERGE Team. The eMERGE Network: A consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics* 2011; 4 (1): 13.
- Gottesman O, Kuivaniemi H, Tromp G, *et al.*; eMERGE Network. The electronic medical records and genomics (eMERGE) network: past, present, and future. *Genet Med* 2013; 15 (10): 761–71.
- Zouk H, Venner E, Lennon NJ, *et al.* Harmonizing clinical sequencing and interpretation for the eMERGE III Network. *Am J Hum Genet* 2019; 105 (3): 588–605.
- Bodenreider O, Nguyen D, Chiang P, *et al.* The NLM value set authority center. *Stud Health Technol Inform* 2013; 192: 1224.
- Hong N, Wen A, Stone DJ, *et al.* Developing a FHIR-based EHR phenotyping framework: a case study for identification of patients with obesity and multiple comorbidities from discharge summaries. *J Biomed Inform* 2019; 99: 103310.
- Wen A, Rasmussen LV, Stone D, *et al.* CQL4NLP: development and integration of FHIR NLP extensions in clinical quality language for EHR-driven phenotyping. *AMIA Jt Summits Transl Sci Proc* 2021; 2021: 624–33.
- Mo H, Thompson WK, Rasmussen LV, Pacheco JA, *et al.* Desiderata for computable representations of electronic health records-driven phenotype algorithms. *J Am Med Inform Assoc* November 2015; 22 (6): 1220–30.
- 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.