AMERICAN ACADEMY OF OPHTHALMOLOGY®

Advancing Toward a Common Data Model in Ophthalmology: Gap Analysis of General Eye Examination Concepts to Standard Observational Medical Outcomes Partnership (OMOP) Concepts

Cindy X. Cai, MD, MS,[1,*] *William Halfpenny, MB, BChir, MEng,*[2,*] *Michael V. Boland, MD, PhD,*[3] *Harold P. Lehmann, MD, PhD,*[4,5] *Michelle Hribar, PhD,*[6,7,8] *Kerry E. Goetz, MS,*[6] *Sally L. Baxter, MD, MSc*[2,9]

Purpose: Evaluate the degree of concept coverage of the general eye examination in one widely used electronic health record (EHR) system using the Observational Health Data Sciences and Informatics Observational Medical Outcomes Partnership (OMOP) common data model (CDM).

Design: Study of data elements.

Participants: Not applicable.

Methods: Data elements (field names and predefined entry values) from the general eye examination in the Epic foundation system were mapped to OMOP concepts and analyzed. Each mapping was given a Health Level 7 equivalence designation—*equal* when the OMOP concept had the same meaning as the source EHR concept, *wider* when it was missing information, *narrower* when it was overly specific, and *unmatched* when there was no match. Initial mappings were reviewed by 2 graders. Intergrader agreement for equivalence designation was calculated using Cohen's kappa. Agreement on the mapped OMOP concept was calculated as a percentage of total mappable concepts. Discrepancies were discussed and a final consensus created. Quantitative analysis was performed on *wider* and *unmatched* concepts.

Main Outcome Measures: Gaps in OMOP concept coverage of EHR elements and intergrader agreement of mapped OMOP concepts.

Results: A total of 698 data elements (210 fields, 488 values) from the EHR were analyzed. The intergrader kappa on the equivalence designation was 0.88 (standard error 0.03, $P < 0.001$). There was a 96% agreement on the mapped OMOP concept. In the final consensus mapping, 25% (1% fields, 31% values) of the EHR to OMOP concept mappings were considered *equal*, 50% (27% fields, 60% values) *wider*, 4% (8% fields, 2% values) *narrower*, and 21% (52% fields, 8% values) *unmatched*. Of the *wider* mapped elements, 46% were missing the laterality specification, 24% had other missing attributes, and 30% had both issues. *Wider* and *unmatched* EHR elements could be found in all areas of the general eye examination.

Conclusions: Most data elements in the general eye examination could not be represented precisely using the OMOP CDM. Our work suggests multiple ways to improve the incorporation of important ophthalmology concepts in OMOP, including adding laterality to existing concepts. There exists a strong need to improve the coverage of ophthalmic concepts in source vocabularies so that the OMOP CDM can better accommodate vision research.

Financial Disclosure(s): Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. *Ophthalmology Science 2023;3:100391* © *2023 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).*

Supplemental material available at www.ophthalmologyscience.org.

The widespread adoption of electronic health records (EHRs) has generated a tremendous volume of digital data available for analysis. Nationwide surveys show increasing EHR use among ophthalmologists, from 32% in 2011[1] to 72% in 2016.[2] Alongside this growth in data is growing awareness of the importance of data standardization.[3–8] Common data models (CDMs) are a means by which to organize and standardize data by imposing a common structure and vocabulary.[9–11] For example, a CDM can specify how and where visual acuity data are stored. Standardization allows disparate datasets, such as those from different EHRs, to be combined. Large-scale data aggregation has many applications including patient care, secondary use of EHR data for research, population health

https://doi.org/10.1016/j.xops.2023.100391
ISSN 2666-9145/23

surveillance, and quality improvement, among others.[12] Despite the advantages of CDMs, ophthalmology data are often not well represented in data models.

The Observational Health Data Sciences and Informatics (OHDSI, pronounced "Odyssey") Observational Medical Outcomes Partnership (OMOP) CDM is a frequently used model that represents observational data (e.g., medical observations, conditions, measurements, drug exposures, procedures) in person-centric tables.[12,13] Several large, biomedical research initiatives have been standardized to the OMOP CDM, including the National Institutes of Health *All of Us* Research Program, the Veterans Affairs Millions Veterans Program, and the National COVID Cohort Collaborative. Unfortunately, these large data repositories do not include detailed eye examination data because discrete elements from a recorded eye examination were not mapped during standardization. For example, in a typical data *extract-transform-load* (ETL) process that maps source data to the OMOP CDM, every data element in the source data (e.g., a local EHR system) is first *extracted*, then *transformed* to an OMOP standard concept, then *loaded* into the target destination (e.g., an OMOP-based data warehouse). Standard concepts in OMOP primarily draw from existing vocabularies and terminologies such as the International Classification of Diseases, Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT), Current Procedural Terminology, Logical Observation Identifiers Names and Codes (LOINC), or RxNorm.[13,14] If there is no clear mapping or logic to guide the transformation from the source to the target, or if the information is collected in a full text field, then the data element is dropped; it is simply not included in the ETL (specifically *load*) process. This exclusion is the current circumstance for eye examination data in many of these databases. Not including these eye examination data has important implications—eye examination findings cannot be used to define clinically relevant cohorts, nor can we ascertain clinically important research outcomes.

Successful ETL requires the appropriate vocabulary (or standard concepts) as well as structure (e.g., the correct table or location to which to add the data). In this study, we focused on evaluating the availability of vocabulary or terminology to represent ophthalmic eye examination concepts. We mapped data elements from the general eye examination module of a commonly used EHR system to OMOP standard vocabularies and analyzed the current gaps in covering eye examination data. By this mapping, we aimed to lay groundwork for future data transformations from source EHRs to the OMOP CDM. Additionally, the gaps identified will inform subsequent efforts to develop new concepts needed to improve representation of eye examination data.

Methods

The University of California San Diego institutional review board determined that the study protocol did not constitute human subjects research.

Source Concept Selection

The source concepts were the data elements from the main examination of the Epic ophthalmology module (Kaleidoscope from the May 2022 Epic Foundation System, Epic Systems). The ophthalmology module from the Foundation System is available to all Epic users who have Kaleidoscope implemented and is customizable for each institution. The data elements consisted of the variable name for each field and any associated predefined permissible entry value that could be selected from a preset list (Figure 1). Only concepts specific to right eyes were chosen to prevent duplication. The predefined values associated with results from visual acuity testing (e.g., 20/60) were also excluded as these are largely numeric and not discrete concepts. The examination components associated with specialty areas, including contact lens, strabismus, and retinopathy of prematurity located separately from the main eye examination, were not included in the analysis.

Mapping Process

Figure 2 depicts an overview of the tiered mapping process and subsequent analysis. A detailed protocol was created to standardize mapping source EHR concepts to target OMOP concepts (Supplemental Material). In the primary mapping, performed by W.H., each source data element from the EHR was mapped to an OMOP standard concept using 2 open source tools, Automated Terminology Harmonization, Extraction and Normalization for Analytics (Athena), and USAGI (Fig 2, panels 1A and 1B). Athena is a web application used to browse vocabulary in the OMOP CDM.[15] USAGI is software that uses a term-similarity approach to map terms from a source system into standard OMOP vocabulary.[16] USAGI and Athena were used in the initial mapping and Athena was used to verify the initial mappings. The source EHR concept, target concept with OMOP concept identification (ID) number, OMOP source vocabulary (SNOMED-CT, LOINC, or RxNorm), and source vocabulary ID number were recorded.[13,16] The OMOP CDM primarily draws from existing source vocabularies and has limited de-novo vocabulary. Source vocabularies typically have a specific purpose. For example, SNOMED-CT is an international clinical oncology for medical diagnoses and examination findings, LOINC focuses on laboratory tests and clinical observations, and RxNorm provides normalized names for clinical drugs. There can be overlap between terminologies. When a concept could be mapped to multiple terminologies, the closest match was chosen. If multiple terminologies provided similar matches (e.g., all would be considered *equal*, or all considered *wide*), preference was given to selecting the SNOMED-CT term based on our protocol.[14]

Each mapping was categorized based on how well the OMOP concept represented the EHR source element. The concept map equivalence designation was based on the Health Level 7 Fast Healthcare Interoperability Resources concept-map equivalence.[17] (Fig 2, panel 1C) *Equal* is defined as OMOP mappings that directly represented the source EHR element, *wider* indicated mappings that represented the source element with some information loss, *narrower* indicated mappings that introduced additional potentially inaccurate representation, and *unmatched* indicated no mapping was possible. To help identify the etiology of *wider* and *unmatched* mappings, these equivalence designations were further grouped into categories. *Wider* mappings could be missing the laterality concept (for example, specifying the right or left eye), other modifiers (for example, the specification of checking visual acuity with a pinhole occluder), or both. *Unmatched* mappings could be a true no match (for example, the source concept was deemed critical and there was no appropriate

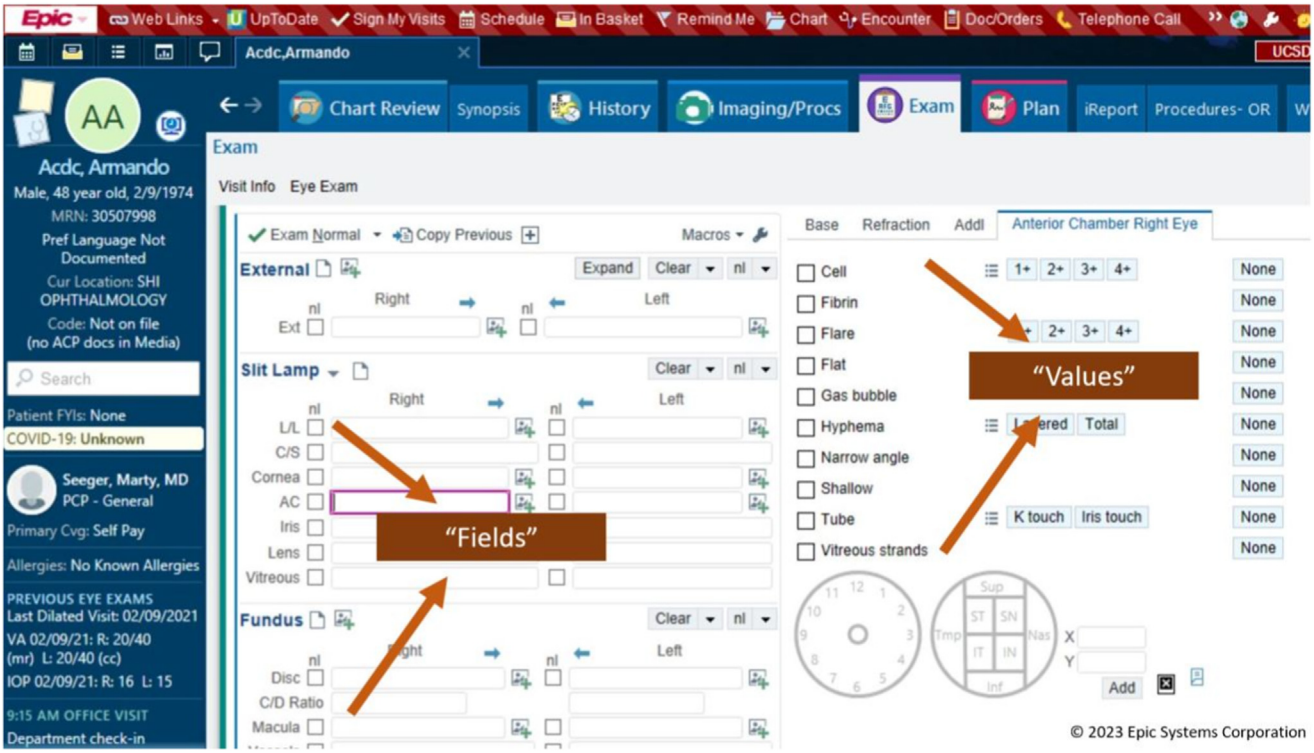


Figure 1. Example of data elements extracted from the source electronic health record which consist of names of fields (e.g., lids/lashes, macula) and predefined entry values (e.g., cell, fibrin, vitreous strands). Note that this interface is from a simulated environment and not from a real patient's record. Copyright 2023 Epic Systems Corporation; image shared with permission from the Epic Content Sharing team.

OMOP standard concept) or other (for example, the source concept was deemed not important in an ETL process).

The results of the first round of mapping were exported into Microsoft Excel and independently reviewed and updated by 2 secondary graders (C.X.C. and S.L.B., faculty-level ophthalmologists with informatics training). After these mappings were compared, the 2 graders discussed discrepancies to form the final consensus mapping.

Statistical Analysis

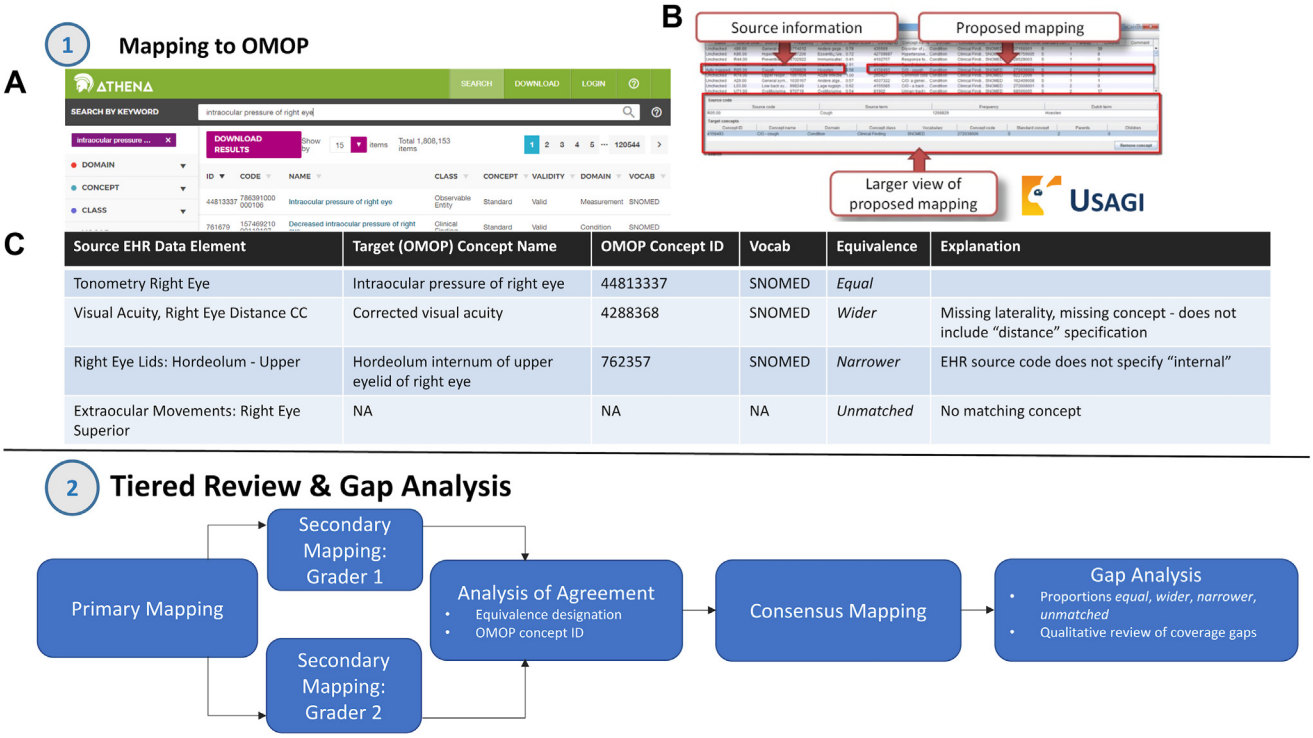
Summary statistics were used to describe the consensus mapping by equivalence designation. Agreement between the 2 secondary graders for equivalence designation was calculated using Cohen's kappa.[18] Agreement for OMOP concept ID selection was calculated as the proportion of OMOP concept IDs that were identical between the 2 graders divided by total mappable concepts (excluding concepts that were *unmatched*). Concepts that were *wider* and *unmatched* were qualitatively reviewed and grouped into categories. All statistical analyses were performed using Stata (StataCorp, 2019, *Stata Statistical Software: Release 16.*) and Python using the *pandas* library (available on GitHub).[19]

Results

A total of 698 data elements were included in the analysis, consisting of fields (n = 210) and predefined entry values (n = 488). These included data elements such as visual acuity testing method, intraocular pressure, extraocular movements, pupils, refraction, gonioscopy, slit lamp examination, and dilated fundus examination, among others (Table 1).

Here, we provide some concrete examples to illustrate the mapping process. For example, in recording visual acuity, the field name specified how the visual acuity was checked (e.g., right eye distance without correction, right eye distance without correction with pinhole). The predefined entry value in the chart was used for the measurement (e.g., Snellen). There was an *equal* mapping for the field of "right eye distance without correction" with a LOINC term ("visual acuity far uncorrected right eye"). However, there was only a *wider* mapping for the field "right eye distance with correction" with the SNOMED-CT term "corrected visual acuity." This mapping was designated as *wider* because the SNOMED-CT term was missing the laterality (in this case "right eye") and the specification of "distance." The field "right eye distance with correction +/−," where the number of additional or missing letters seen is typically recorded, was designated as *unmatched* because there were no similar SNOMED-CT codes. Many of the pre-defined values for visual acuity testing (e.g., Snellen - linear, Snellen - single, Snellen - blocked) were designated as *wider*. The SNOMED-CT term "Snellen chart assessment" does not have the same granularity as the source Epic element that also distinguished whether the Snellen letters were presented in a linear, single, or blocked fashion. Additional examples of *equal*, *wider*, *narrower*, and *unmatched* can be found in Table 1.

The first round of grading was reviewed by the 2 secondary graders, and 10% (n = 70) of equivalence labels or



Source EHR Data Element	Target (OMOP) Concept Name	OMOP Concept ID	Vocab	Equivalence	Explanation
Tonometry Right Eye	Intraocular pressure of right eye	44813337	SNOMED	*Equal*	
Visual Acuity, Right Eye Distance CC	Corrected visual acuity	4288368	SNOMED	*Wider*	Missing laterality, missing concept - does not include "distance" specification
Right Eye Lids: Hordeolum - Upper	Hordeolum internum of upper eyelid of right eye	762357	SNOMED	*Narrower*	EHR source code does not specify "internal"
Extraocular Movements: Right Eye Superior	NA	NA	NA	*Unmatched*	No matching concept

Figure 2. Overview of the mapping process and subsequent gap analysis. Panel 1 depicts the mapping process from the source electronic health record (EHR) to the target Observational Medical Outcomes Partnership (OMOP) common data model, with use of the Athena tool (**A**), USAGI tool (**B**), and final generation of mappings (**C**) for each EHR data element. These mappings then underwent a tiered review process (Panel 2), followed by evaluation of intergrader agreement, development of consensus mappings, and completion of a gap analysis based on the consensus mapping. CC = with correction; ID = identification; NA = not available; SNOMED = Systematized Nomenclature of Medicine.

OMOP standard concept IDs were changed by grader 1 and 14% (n = 100) by grader 2. In the consensus mapping, 15% (n = 107) were updated compared to the primary mapping. Overall, there was excellent agreement between the 2 secondary graders on the equivalence designation with a kappa statistic of 0.88 (standard error 0.03, $P < 0.001$). Agreement for the exact mapped OMOP concept ID was also excellent at 96%.

After the 2 graders reached agreement on discrepant data elements in the consensus mapping step, 25% (n = 177) of the mappings were considered *equal* (1%, n = 28, of fields, and 31%, n = 149, of values), 50% (n = 348) *wider* (27%, n = 57, of fields, and 60%, n = 291, of values), 4% (n = 25) *narrower* (8%, n = 16, of fields, and 2%, n = 9, of values), and 21% (n = 148) *unmatched* (52%, n = 109, of fields, and 8%, n = 39, of values) (Fig 3). Of the *wider* data mappings, 46% (n = 160) had missing laterality specification, 24% (n = 85) had other missing specifications, and 30% (n = 103) had both issues. Of the *unmatched* data elements, 47% (n = 70) did not have a match, and 53% (n = 78) were considered either not relevant in an ETL process (e.g., mapping the field name of right eye anterior chamber when the values within the field will be mapped, for example right eye anterior chamber cell), or combined with other data elements in an ETL process (e.g., results of the extraocular movement examination would be interpreted together and mapped to a single OMOP concept rather than mapping each field) (Table 1). Of the mappings that had a match (whether *equal*, *wider*, or *narrower*), 90% (n = 492) were mapped to SNOMED-CT source vocabulary, 9% (n = 47) to LOINC, and 2% (n = 47) to RxNorm.

On qualitative review of the mappings that were designated as *wider* with missing other concepts and *unmatched*, there were data elements from all components of the ophthalmic examination that had imprecise (or not *equal*) matches to OMOP concepts (Table 2, Table S3).

Discussion

We systematically mapped data elements related to the general eye examination from the foundation system of a widely used EHR system, Epic, to standard vocabularies in the OMOP CDM, an increasingly utilized data model for harmonizing disparate observational health data.[20–23] There were substantial gaps in coverage of ophthalmology concepts in the OMOP CDM in all areas of the general eye examination. Only a quarter of the source data elements had a completely matching OMOP standard concept. This gap analysis demonstrated several opportunities for improving representation of ophthalmic examination data in the OMOP CDM.

Table 1. Sample Data Elements Mapped to the OMOP CDM

Type	Examination Area	Data Element	OMOP Concept ID	OMOP Concept Name	HL7 Equivalence Designation	OMOP Source Vocabulary	Source Concept ID	Explanation
Field	Visual acuity	Right eye distance SC	1989004	Visual acuity far uncorrected right eye	*Equal*	LOINC	98505-1	
Field	Visual acuity	Right eye distance CC	4288368	Corrected visual acuity	*Wider*	SNOMED-CT	397536007	Missing laterality, missing concept - does not include "distance" specification
Field	Tonometry	Right eye	44813337	Intraocular pressure of right eye	*Equal*	SNOMED-CT	786391000000106	
Field	Tonometry	Target right eye	0	No matching concept	*Unmatched*	None		
Field	Extraocular movements	Right eye superior	0	No matching concept	*Unmatched*	None		
Value	Visual acuity	Method: Snellen - linear	4102368	Snellen chart assessment	*Wider*	SNOMED-CT	252973004	Missing concept - does not include "linear" specification
Value	Visual acuity	Method: Snellen - blocked	4102368	Snellen chart assessment	*Wider*	SNOMED-CT	252973004	Missing concept - does not include "blocked" specification
Field	Main examination	Right eye conjunctiva normal	4201560	Conjunctiva normal	*Wider*	SNOMED-CT	301926003	Missing laterality
Value	Main examination	Right eye conjunctiva: bleb - seidel negative	4087942	Bulbar conjunctival drainage bleb	*Wider*	SNOMED-CT	246883008	Missing laterality, missing concept - does not include "Seidel negative" specification
Value	Main examination	Right eye lids: hordeolum - upper	762357	Hordeolum internum of upper eyelid of right eye	*Narrower*	SNOMED-CT	331601000119103	EHR source code does not specify "internal"
Value	Main examination	Right eye lids: hordeolum - lower	36684643	Internal hordeolum of right lower eyelid	*Narrower*	SNOMED-CT	334171000119102	EHR source code does not specify "internal"

CC = with correction; CDM = common data model; HL7 = Health Level 7; ID = identification; LOINC = Logical Observation Identifiers Names and Codes; OMOP = Observational Outcomes Medical Partnership; SC = without correction; SNOMED-CT = Standardized Nomenclature of Medicine Clinical Terms.
Definitions of equivalence designation: *Equal*: mappings that directly represent the source element; *Wider*: mappings that represent the source element with some information loss; *Narrower*: mappings that introduce additional, potentially inaccurate information; *Unmatched*: no mapping was possible.
The table includes the source data elements from the electronic health record on the left with the type of data element, area of the eye examination the data element is from, and the name of the data element. The results of the mapping are shown on the right with the mapped OMOP standard concept ID number, OMOP concept name, HL7 concept-map equivalence designation, OMOP source vocabulary, source vocabulary ID, and additional explanations.

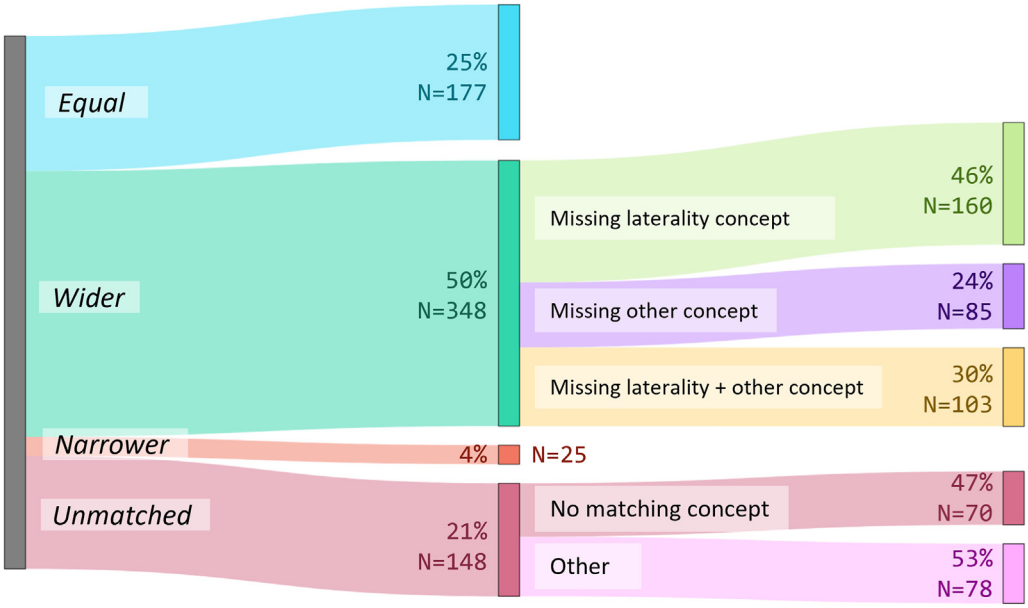


Figure 3. Sankey diagram depicting the proportions of concept-map equivalence designation for the consensus mapping. Only a quarter of the source data elements had a completely matching Observational Medical Outcomes Partnership standard concept.

One common reason for imprecise matches was the issue of missing laterality—many concepts in OMOP do not specify laterality while the source ophthalmic examination did. Because many analyses in research require eye-specific metrics, laterality information is critical. Beyond laterality, a substantial proportion of data elements in all components of the general eye examination had some degree of missing clinical granularity in the closest-matching OMOP standard concept. First, there were missing modifiers that related to the method and type of the measured value. For example, among concepts related to visual acuity, there were gaps relating to the distance associated with the measurement, whether pinhole was used, and granular details regarding the type of vision testing chart used. Appropriate interpretation and analysis of visual acuity data rely on how the measurement was obtained, thus, adequately representing the permutations of visual acuity testing is a major need. Second, lack of specific anatomic location was another major gap. In the funduscopic examination, the retina can be divided into multiple sections including macula and periphery. The presence of drusen in the peripheral retina does not have the same implication for impact on visual acuity as the presence of drusen in the macula. Finally, there were some data elements that were not represented at all by OMOP concepts, such as geographic atrophy and examination findings related to gonioscopy. These represent major opportunities for expanding the OMOP CDM as well as source vocabularies.

Our results demonstrated high levels of intergrader agreement, engendering a high level of confidence regarding the gaps identified. The mapping process involved 2 steps, with the first round of mapping performed by a single grader and the second round by 2 graders. We acknowledge that the degree to which the subsequent graders were influenced by the first round of grading could inflate the agreement. However, due to the time-consuming nature of the mapping process, this 2-tiered grading system was necessary to optimize clinician involvement. Additionally, as another informal step to consensus building, we had additional clinical and research experts in the OMOP Eye Care & Vision Research workgroup comment on our gap analysis and they agreed that the gaps we identified were clinically relevant.[24] Having multiple graders and reviewers of the mappings reduced the risk of human error in the ETL process. Even using a semi-automated tool such as USAGI did not eliminate the possibility of human errors. During our consensus mapping, we noticed some of the disagreements were simply due to overlooking matching concepts. Our experience points to the need for having either multiple iterations or multiple graders involved in the mapping process to limit human error. This tiered mapping approach and transparency regarding the mappings themselves (included in the Supplement) represent key strengths of this study.[25–29]

This study was focused on evaluating the availability of vocabulary to adequately represent clinical concepts (i.e., semantic representation) and not on the structure of the CDM needed to accommodate the terminology. An underlying assumption of our approach is that we need precoordinated concepts. The OMOP CDM has limited capacity for incorporating postcoordinated concepts and, for the cases where postcoordination may be possible, precoordination remains the approach generally recommended by the OHDSI Community, a discussion of which is outside the scope of this study.[30,31] We acknowledge that allowing postcoordination would change the designation of some concepts specifically from *wider* to *equal*. However, allowing postcoordination does not change the other conceptual gaps identified in this study (e.g., those identified as *unmatched*).

Table 2. Qualitative Review and Categorization of Data Elements that had *Wider* or *Unmatched* OMOP Concepts with Examples

	Area of Conceptual Gap	
Component of ophthalmic examination	Partial Matches in OMOP (designated *wider* with missing concepts)	No Matches in OMOP (designated *unmatched*)
VA	Type of chart used for VA testing: Snellen - Linear Snellen - Single Snellen - Blocked	Type of chart used for VA testing: Numbers - Linear Numbers - Single Numbers - Blocked
IOP	Method of measurement: Tonopen Palpation	Target IOP right eye Maximum IOP right eye
Pupil examination	Grading of pupillary response Grading of afferent pupillary defect	
Extraocular movements		Findings by eye (e.g., right eye nasal upshoot)
Confrontational visual field	Method of testing: Count fingers Toys	
Refraction	Type of cycloplegic refraction: Subjective Overrefraction Autorefraction Retinoscopy	
Glasses and contact lens	Type of final glasses prescription: Single vision lens Progressive addition lenses Trifocal Bifocal	Expiration date of final glasses prescription
Color vision		Method of testing: Stilling Hardy-Rand-Rittler
Keratometry		Method of testing: Automated Manual
Others	Light setting used for brightness acuity testing: Off Low Medium High	Schirmer's testing with anesthesia
Gonioscopy	Grading of gonioscopic findings (e.g. wide open angles)	Type of mirror used for gonioscopy (e.g., Sussmann, 4 mirror)
Anterior slit lamp examination	Examination findings such as cystic bleb	Examination findings such as glaucoma drainage device implant
Fundoscopic examination	Examination findings such as size, characteristic, and location of drusen	Examination findings such as geographic atrophy, lacquer crack, retinal pigment epithelium mottling, and normal right macula

IOP = intraocular pressure; OMOP = Observational Outcomes Medical Partnership; VA = visual acuity.

We chose to evaluate both the variable names for each field and associated pre-defined permissible entry values even though these are not equivalent data types—fields are relatively fixed, while entry values are customizable at each institution. This choice was made to evaluate potential examination findings. For example, it is not helpful for downstream research to only ETL the field name of "right eye conjunctiva" without the potential examination finding of "neoplasm." Our methodology of assessing both the field and value allows us to evaluate whether the precoordinated combined concept of "right eye conjunctiva: neoplasm" is available in OMOP standard vocabularies. Our work here does not cover additional pre-defined entry values that each institution might customize on the Epic Foundation System.

To address the identified concept gaps, several workgroups have formed in partnership with various stakeholders, including the American Academy of Ophthalmology, OHDSI (which governs the OMOP CDM), and SNOMED International (which governs SNOMED-CT, one of the source vocabularies included in the OMOP CDM). Efforts are underway to delve more deeply into these concept gaps, identify which gaps are most important and urgent to address, and submit new standards to these organizations to fill the gaps. Additionally, the workgroups within the OHDSI community are working to develop conventions to guide ETL of ophthalmic data elements. Broad engagement with eye care providers and vision researchers is needed to inform use cases and to help prioritize

standards development for data elements of greatest utility for clinical and research applications.

The gap analysis performed here was in the representation of a general eye examination by a single EHR system. We also do not assert that this effort represents a gap analysis of all relevant ophthalmic data. For example, clinical trials and research studies have a wide array of structural and functional endpoints that may not be in routine clinical use and were not analyzed here. We also did not examine imaging-related metrics or subspecialty-specific examination components. Leveraging components of the general eye examination in defining cohorts can be advantageous. Prior work has demonstrated a substantial increase in the number of cases, for example, of pseudoexfoliation glaucoma, that can be identified when findings in the slit lamp examination were used.[32] This study is an initial effort to characterize gaps in the coverage of a general eye examination and certainly more work is needed.

There are several limitations of this study. First, this study focuses on the data elements from a single EHR. Although this EHR is widely used, the degree to which these results are generalizable to other EHR representations of the ophthalmic examination is unknown. A major strength of this work is that we used the foundation system of the EHR for maximal generalizability across institutions that use Epic Systems. Second, the predefined entry values for the data field elements (for example, what the user can enter for the right eye anterior chamber finding) are not based on standard vocabulary or concepts. Instead, these predefined entry values originated from various institutions working with the EHR vendor, where they became incorporated into the "Foundation system." There is a need in the field of ophthalmology to develop best practices and consensus around standardized description of examination elements. Third, it should be recognized that not all of the data elements included in this study are likely to have equal utility for observational research or clinical practice. This study did not address the usefulness of different data elements, which would be an important pragmatic consideration when implementing an ETL. Fourth, we also took a strict standard to determining whether a term was mapped for not. For example, "corrected visual acuity" was designated as a *wider* mapping, and one could argue that distance is implied by convention unless otherwise stated. Although the risk of ambiguity could be low in this instance, we chose the stricter methodology to make concepts as explicit as possible. Since OMOP data warehouses are shared across specialties in each institution, it is possible that nonophthalmic researchers will have access to the data. To prevent potential misuse of data elements, we thought it was best to explicitly represent concepts. This was also aimed at reducing subjectivity in the mapping process. Finally, there are possibilities for human error in the specific mappings, as alluded to earlier, although this is inherent in any ETL process and was mitigated by having multiple graders involved.

In conclusion, we present a comprehensive gap analysis of the OMOP CDM in representing general eye examination elements from a widely used EHR system. Gaps in representation were present across essentially all components of the eye examination. This analysis will inform future standards development and advance efforts toward improved representation of ophthalmic data elements in the OMOP CDM. Greater representation means greater availability of practice-sourced data in research, which can lead to more effective ophthalmic care for patients.

Footnotes and Disclosures

Originally received: March 23, 2023.
Final revision: August 16, 2023.
Accepted: August 21, 2023.
Available online: August 25, 2023. Manuscript no. XOPS-D-23-00063.

[1] Wilmer Eye Institute, Johns Hopkins School of Medicine, Baltimore, Maryland.

[2] Division of Biomedical Informatics, Department of Medicine, University of California San Diego, La Jolla, California.

[3] Massachusetts Eye and Ear, Harvard Medical School, Boston, Massachusetts.

[4] Division of Health Sciences Informatics, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland.

[5] Biomedical Informatics and Data Science, Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland.

[6] Office of Data Science and Health Informatics, National Eye Institute, National Institute of Health, Bethesda, Maryland.

[7] Department of Ophthalmology, Casey Eye Institute, Portland, Oregon.

[8] Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, Oregon.

[9] Division of Ophthalmology Informatics and Data Science, Viterbi Family Department of Ophthalmology and Shiley Eye Institute, University of California San Diego, La Jolla, California.

*Co-first authors.

Disclosures:

All authors have completed and submitted the ICMJE disclosures form.

The authors made the following disclosures:

S.L.B.: Equipment support − Optomed, Topcon outside the submitted work. C.X.C.: Grant support −Regeneron Pharmaceuticals, Inc.; M.V.B.: Consulting fees − Carl Zeiss Meditec, Topcon Healthcare, Janssen, Allergan.

Data Availability: Due to the proprietary nature of the source data elements, the data generated and used in the current study are available to other Epic users from the corresponding author upon reasonable request.

The study was supported by the National Institutes of Health (NIH, Bethesda, MD) Grants DP5OD029610 (S.L.B.), R01EY034146 (S.L.B.), P30EY022589 (S.L.B.), T35EY033704 (S.L.B.), K23EY033440 (C.X.C.), and unrestricted departmental grants from Research to Prevent Blindness (New York, NY). Dr Cai is the Jonathan and Marcia Javitt Rising Professor of Ophthalmology. The sponsors or funding organizations had no role in the design or conduct of this research. Michelle Hribar, PhD, an editor of this journal, was recused from the peerreview process of this article and had no access to information regarding its peer review.

HUMAN SUBJECTS: No human subjects were included in this study. The University of California San Diego institutional review board determined that the study protocol did not constitute human subjects research.

Author Contributions:

Conception and design: Cai, Halfpenny, Boland, Lehmann, Baxter

Analysis and interpretation: Cai, Halfpenny, Boland, Lehmann, Hribar, Goetz, Baxter; Data collection: Cai, Halfpenny, Baxter

Obtained funding: Cai, Baxter

Overall responsibility: Cai, Halfpenny, Baxter

Meeting Presentation: The 2023 Association for Research in Vision and Ophthalmology (ARVO) Meeting, New Orleans, Louisiana.

Abbreviations and Acronyms:
CDM = common data model; **EHR** = electronic health record; **ETL** = extract-transform-load; **ID** = identification; **LOINC** = Logical Observation Identifiers Names and Codes; **OHDSI** = Observational Health Data Sciences and Informatics; **OMOP** = Observational Medical Outcomes Partnership; **SNOMED-CT** = Systematized Nomenclature of Medicine Clinical Terms.

Keywords:
Data model, Data standards, OHDSI, OMOP, Terminologies.

Correspondence:
Cindy X. Cai, MD, MS, Wilmer Eye Institute, 1800 Orleans Street, Maumenee Building, Room 711, Baltimore, MD 21287. E-mail: ccai6@jhmi.edu.

References

1. Boland MV, Chiang MF, Lim MC, et al. Adoption of electronic health records and preparations for demonstrating meaningful use: an American Academy of Ophthalmology survey. *Ophthalmology*. 2013;120:1702−1710.
2. Lim MC, Boland MV, McCannel CA, et al. Adoption of electronic health records and perceptions of financial and clinical outcomes among ophthalmologists in the United States. *JAMA Ophthalmol*. 2018;136:164−170.
3. He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25:30−36.
4. Halfpenny W, Baxter SL. Towards effective data sharing in ophthalmology: data standardization and data privacy. *Curr Opin Ophthalmol*. 2022;33:418−424.
5. Wang SY, Pershing S, Lee AY. Big data requirements for artificial intelligence. *Curr Opin Ophthalmol*. 2020;31:318−323.
6. Baxter SL, Lee AY. Gaps in standards for integrating artificial intelligence technologies into ophthalmic practice. *Curr Opin Ophthalmol*. 2021;32:431−438.
7. Boland MV, Schuman JS, Mattox CG. American Glaucoma Society position statement: electronic data standards for clinical practice. *J Glaucoma*. 2013;22:174−175.
8. Sheehan J, Hirschfeld S, Foster E, et al. Improving the value of clinical research through the use of common data elements. *Clin Trials*. 2016;13:671−676.
9. Kent S, Burn E, Dawoud D, et al. Common problems, common data model solutions: evidence generation for health technology assessment. *Pharmacoeconomics*. 2021;39:275−285.
10. Reisinger SJ, Ryan PB, O'Hara DJ, et al. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. *J Am Med Inform Assoc*. 2010;17:652−662.
11. Weeks J, Pardee R. Learning to share health care data: a brief timeline of influential common data models and distributed health data networks in U.S. Health Care Research. *EGEMS (Wash DC)*. 2019;7:4.
12. Hripcsak G, Duke JD, Shah NH, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;216:574−578.
13. Observational Health Data Sciences, Informatics. The book of OHDSI. https://ohdsi.github.io/TheBookOfOhdsi/; 2021. Accessed January 17, 2023.
14. Hoskins Jr HD, Hildebrand PL, Lum F. The American Academy of Ophthalmology adopts SNOMED CT as its official clinical terminology. *Ophthalmology*. 2008;115:225−226.
15. GitHub - OHDSI/Athena: web application for distributing and browsing the standardized vocabularies for all instances of an OMOP CDM. GitHub. https://github.com/OHDSI/Athena. Accessed January 17, 2023.
16. Usagi. http://ohdsi.github.io/Usagi/. Accessed January 17, 2023.
17. Codesystem-concept-map-equivalence - FHIR v4.3.0. https://www.hl7.org/fhir/codesystem-concept-map-equivalence.html. Accessed January 17, 2023.
18. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37−46.
19. GitHub - will-0/epic-omop-map-22: repository containing code for analysis of code mappings from EPIC kaleidoscope content coverage analysis. GitHub. https://github.com/will-0/epic-omop-map-22. Accessed March 10, 2023.
20. Park K, Cho M, Song M, et al. Exploring the potential of OMOP common data model for process mining in healthcare. *PLoS One*. 2023;18:e0279641.
21. Junior EPP, Normando P, Flores-Ortiz R, et al. Integrating real-world data from Brazil and Pakistan into the OMOP common data model and standardized health analytics framework to characterize COVID-19 in the Global South. *J Am Med Inform Assoc*. 2023;30:643−655.
22. Papez V, Moinat M, Voss EA, et al. Transforming and evaluating the UK biobank to the OMOP common data model for COVID-19 research and beyond. *J Am Med Inform Assoc*. 2022;30:103−111.
23. Klann JG, Joss MAH, Embree K, Murphy SN. Data model harmonization for the All of Us research program: transforming i2b2 data into the OMOP common data model. *PLoS One*. 2019;14:e0212463.
24. Upcoming workgroup calls. https://www.ohdsi.org/upcoming-working-group-calls/. Accessed January 17, 2023.
25. Biedermann P, Ong R, Davydov A, et al. Standardizing registry data to the OMOP common data model: experience from three pulmonary hypertension databases. *BMC Med Res Methodol*. 2021;21:238.
26. Cho S, Sin M, Tsapepas D, et al. Content coverage evaluation of the OMOP vocabulary on the transplant domain focusing on concepts relevant for kidney transplant outcomes analysis. *Appl Clin Inform*. 2020;11:650−658.
27. Garza M, Del Fiol G, Tenenbaum J, et al. Evaluating common data models for use with a longitudinal community registry. *J Biomed Inform*. 2016;64:333−341.
28. Matcho A, Ryan P, Fife D, Reich C. Fidelity assessment of a clinical practice research datalink conversion to the OMOP common data model. *Drug Saf*. 2014;37:945−959.
29. Zhou X, Murugesan S, Bhullar H, et al. An evaluation of the THIN database in the OMOP Common Data Model for active drug safety surveillance. *Drug Saf*. 2013;36:119−134.

30. Representing laterality + general questions about post-coordination. https://forums.ohdsi.org/t/representing-laterality-general-questions-about-post-coordination/16872/7. Accessed July 1, 2022.
31. Belenkaya R, Gurley MJ, Golozar A, et al. Extending the OMOP common data model and standardized vocabularies to support observational cancer research. *JCO Clin Cancer Inform*. 2021;5:12–20.
32. Stein JD, Rahman M, Andrews C, et al. Evaluation of an algorithm for identifying ocular conditions in electronic health record data. *JAMA Ophthalmol*. 2019;137:491–497.