

Lessons Learned from Implementation of ETL Conventions for Best Recorded Visual Acuity (BRVA) in the OMOP Common Data Model at an Academic Institution.

Authors and Affiliations

Sarah R. Wellik, MD¹, Cindy X Cai, MD, MS^{2,3}, Michelle Hribar, PhD^{4,5}, Kerry Ashby⁶, Sally L. Baxter, MD, MSc⁷, Nitin Parikh, MS^{8,9}, The OHDSI Eye Care and Vision Research Workgroup

¹*Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL, USA*

²*Wilmer Eye Institute, Johns Hopkins School of Medicine, Baltimore, MD, USA*

³*Biomedical Informatics and Data Science, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA*

⁴*University of Illinois at Chicago, Department of Ophthalmology and Visual Sciences, Chicago, IL, USA*

⁵*University of Illinois at Chicago, School of Medicine, AI.Health4All, Chicago, IL, USA*

⁶*Topcon Healthcare Inc, San Diego, CA, USA*

⁷*University of California San Diego, San Diego, CA, USA*

⁸*University of Miami Health System, Miami, FL, USA*

⁹*University of Miami Clinical and Translational Science Institute, Miami, FL, USA*

Background

Visual acuity (VA) is a fundamental clinical measure used to assess the spatial resolution of the visual system. However, VA data captured in electronic health records (EHRs) are frequently heterogeneous and inconsistently documented, with variability in chart types, measurement distance, and clinical qualifiers. These inconsistencies limit reproducibility and comparability across observational datasets. To address this gap, the Observational Health Data Sciences and Informatics (OHDSI) Eye Care and Vision Research Workgroup developed standardized extract-transform-load (ETL) conventions to derive Best Recorded Visual Acuity (BRVA) for the right eye (OD), left eye (OS), and both eyes (OU) using the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The approach was informed by a survey of multi-institutional datasets, which include: the American Academy of Ophthalmology IRIS Registry^{1,2}, Epic Cosmos³, the Sight Outcomes Research Collaborative (SOURCE)⁴, INSIGHT Health Data Research Hub for Eye Health,⁵ Cerner/Oracle Health⁶, the Pediatric Eye Disease Investigator Group (PEDIG)⁷, and a multi-institutional clinical data warehouse in Korea.⁸ To our knowledge, there have been no previous formal specifications to inform standardized representation of BRVA in the OMOP CDM. The purpose of this study is to describe the ETL process for BRVA from Epic, the EHR of the University of Miami Health System eye care data, sharing our experience and lessons learned from this process.

Methods

In early 2025, members of the Eye Care Vision and Research Workgroup conducted a landscape survey of visual acuity ETL conventions in existing ophthalmic registries.⁹ The ETL conventions used here are in line with the existing conventions, voted upon and approved by the Eye Care and Vision Research Workgroup. We included VA from standard charts (Snellen for distance, Jaeger near card, and Early

Treatment of Diabetic Retinopathy Study [ETDRS] charts, as well as common low vision readings when patients cannot read standard charts) and in addition, other aspects under which the vision was measured are important qualifiers in interpreting the results of visual acuity testing (with and without correction, pinhole, etc.) and lastly additional partial lines read are often documented as +/- additional letters after the VA reading. Although the EHR contains the VA testing results, some important details can be poorly or inconsistently documented. Given the current inconsistencies of VA data in the EHR, we sought to create ETL rules that did not introduce erroneous assumptions about the testing conditions. The ETL rules identify the “best” recorded VA value (BRVA) across multiple measures. The ETL conventions can be found in Eye Care Vision and Research Workgroup Github.⁹

During extraction, raw VA values (e.g., 20/20) and associated source field names were captured. Transformation converted VA values into logMAR (Logarithm of the Minimum Angle of Resolution) . This scale is a continuous decimal value using standardized formulas that range from 0.00 (equivalent to 20/20) to 1.00 (equivalent to 20/200), and higher numerical values for even lower levels of vision, up to 4.0 which represents no light perception vision. Adjustments were applied for additional letters read (e.g., “+2” letters corresponding to slight improvement in logMAR). Data cleaning procedures remove extraneous strings such as measurement distance. For each encounter, we used the distance and near VA corrected and uncorrected, the pinhole and manifest VA. From these, BRVA is defined as the lowest logMAR value recorded per eye (OD, OS, OU). Results were loaded into the OMOP CDM measurement table with eye-specific concept identifiers, numeric logMAR values, and preserved source values to maintain traceability.

Results

We used the accepted ETL procedure at the University of Miami to calculate the best recorded visual acuity stored in the Epic EHR system from 1/1/2020 - 1/1/ 2026. The transformation rules converted VA values into logMAR (calculated $LogMAR = -\log_{10}(Snellen\ visual\ acuity\ as\ decimal)$) and stored in the *measurement* table of the OMOP database. Non-Snellen values were also stored as concepts by converting diverse VA measurements into logMAR and storing them in the *value* as number field in the *measurement* table. This ETL procedure supported consistent interpretation and cross-site comparability. Any values that could not be converted at this time because of non-standard charts or measurements received a *value_as_concept_id* of 0 and a null in the *value_as_number* field in the *measurement* table. The original EHR value (ex. 20/20) was stored in the *value_source_value* of the *measurement* table. The source variable name (ex. Dist VA OD sc) was stored in the *measurement_source_value* field.

In the Epic EHR, the Snellen line and additional letters read are typically (but not always) stored in separate data fields. During the ETL process, these paired fields needed to be processed together. Furthermore, multiple of the same type of measurements for the same eye (i.e., multiple distance right eye corrected VA) are stored in separate rows in the Epic database. Adjusting the ETL to account for different documentation patterns and items stored across multiple fields presented a challenge. We include in Table 1 the preliminary total counts in each field examined with the number of distinct BCVA logMAR values.

We are also developing data quality metrics for BCVA and hope to incorporate these rules into OHDSI Data Quality Dashboard.

Table 1. Preliminary ETL counts are provided below.

Measurement Concept ID	Measurement Concept Name	Distinct Visit Count
723167	BRVA OD chart agnostic, letters read, distance or near, with or without pinhole	3163350
723168	BRVA OS chart agnostic, letters read, distance or near, with or without pinhole	3157417
723169	BRVA OU chart agnostic, letters read, distance or near, with or without pinhole	85411
Both 723167 and 723168		3011310

Conclusion

This work presents a standardized and reproducible ETL framework for deriving BRVA within the OMOP CDM. The BRVA framework used ensures that the most clinically relevant value is retained per encountered improves consistency, scalability, and comparability of ophthalmic data across institutions. Key strengths include alignment with existing OMOP standards, preservation of source data, and support for multiple VA measurement formats. There were many challenges encountered during this process, including how to deal with the multitude of different methods of VA testing, ways of recording and concatenation of appended modifiers to each reading. Conversion to uniform logMAR decimal values presented an additional step in the transformation process. This framework will enable robust eye care research within OHDSI and will support future standardized, multi-site observational studies that involve VA, which is arguably the most important measurement in vision data analysis.

References

1. Brant A, Kolomeyer N, Goldberg JL, Haller J, Lee CS, Lee AY, Lorch AC, Miller JW, Hyman L, Pershing S. Evaluating Visual Acuity in the American Academy of Ophthalmology IRIS® Registry. *Ophthalmol Sci.* 2024;4(1):100352.

2. Verana Health. Verana Health and the American Academy of Ophthalmology Launch the Verana Research Network, an IRIS Registry Initiative. veranahealth.com. December 7, 2022. Accessed August 3, 2025. <https://veranahealth.com/verana-health-and-the-american-academy-of-ophthalmology-launch-the-verana-research-network-an-iris-registry-initiative/>
3. Epic Cosmos. Accessed August 3, 2025. <https://cosmos.epic.com>
4. About. Accessed August 3, 2025. <https://www.sourcecollaborative.org/about>
5. INSIGHT Health Data Research Hub for Eye Health. INSIGHT. Accessed August 3, 2025. <https://www.insight.hdrhub.org>
6. Oracle Health (formerly Cerner) – <https://www.oracle.com/health/>
7. [Pediatric Eye Disease Investigator Group - Public Site](#)
8. <https://www.hira-research.or.kr//journal/view.html?doi=10.52937/hira.21.1.2.179>
9. https://docs.google.com/document/d/1anv1s9EixdIHFMh3G2E2DJ8JS1Q5wlxY-JTgl_wG_1w/edit?tab=t.0EyeCareVisionWG/etl-resources/best_recorded_visual_acuity_at_main_OHDSI/EyeCareVisionWG_GitHub - OHDSI Github documentation folder