Quantifying EHR Continuity in the All of Us Research Program

Lina Sulieman, PhD[1], Xinzhuo Jiang, MS[2], Karthik Natarajan, PhD[2]
[1]Vanderbilt University Medical Center, Nashville, TN; [2]Columbia University Irving Medical Center, New York, NY

Background

Electronic Health Records (EHR) are a powerful resource in biomedical research. However, data quality issues, especially those related to completeness, can hinder research studies and compromise the credibility of the trained models[1]. Data completeness is a challenging quality problem. Assessing data completeness for studies' subjects is essential to account for possible biases, especially when EHR data is collected from different healthcare provider organizations (HPOs). Loyalty score is an EHR continuity and data completeness metric that quantifies the existence of routine clinical care[2]. The *All of Us* Research Program is a national initiative collecting biomedical data from consented participants, including EHR, to create a diverse biomedical research repository that researchers have used since 2020. The EHR data are collected from 64 HPOs, standardized, and converted into the OMOP Common Data Model (CDM). Hence, quantifying EHR continuity can help the All of Us researchers to adjust for any missing data, especially if the missingness varies between EHR sites. The loyalty algorithm calculates the score for datasets modeled to i2b2 CDM. This study's objective is to convert the variables of the loyalty score from i2b2 to OMOP, apply the loyalty score to the All of Us EHR dataset, and stratify the score based on EHR HPO, participants' demographics, and socioeconomic factors.

Methods

To create a loyalty score executable on OMOP datasets, we extracted the i2b2 concepts and queries originally used to calculate the loyalty score. Using the concept relationship table, we mapped i2b2 concepts to analogous standard OMOP concepts and reviewed the mappings to exclude incorrect matches. OMOP queries were constructed to extract relevant loyalty score variables, including sex, mammography, pap smear, prostate-specific antigen test, colonoscopy, fecal occult blood test, influenza vaccine, pneumococcal vaccine, body mass index, A1C, prescribed one medication, prescribed two medications, general medical exam, and emergency visits. The loyalty score requires specifying a lookback year, defining the period during which these variables are extracted. Using OMOP-formatted queries, we calculated the loyalty score for lookback periods ranging from 1 to 7 years to assess variations in the score.

Participants with at least one visit from the *All of Us* 8th released dataset (CDR-8) were included in the analysis. We created three models: the original model, the non-weighted model, and the weighted model. The original model used the original coefficients from i2b2 study. The unweighted model used all HPO EHR data, while the weighted model applied weights to ensure balanced representation of participants across HPO sites. We stratified loyalty scores by HPO sites to identify potential missing data across sites. Two predictive models were trained to forecast the likelihood of a visit in the following year, serving as a silver standard for adjusting model coefficients as Klann applied in his study[3]. Performance metrics include area under the curve (AUC), precision-recall (PR) curve, Brier score, and expected calibration error (ECE). Additionally, we stratified the loyalty score by participants' demographics, including age at study entry, self-reported race, income, and education.

Results

The CDR-8 included 426,438 participants with EHR data, 60.69% were female and 54.42% were white. During our concept mappings from i2b2 to OMOP, we corrected concepts for PSA, pap smear, and

influenza that were mapped to general concepts such as primary care visit and expanded the list of concepts. The median loyalty scores for 1 to 7 lookback years were 0.42, 0.49, 0.51, 0.53, 0.56, 0.57, and 0.57. Among the sites, 7 (10.9%) sites had a median score lower than 0.2 using 3 lookback years, while 2 (3.13%) sites had a median score lower than 0.2 using 7 lookback years. Participants who were in the age group 50-to-64 and 65 years and older had the highest score in all lookback years with median values 0.58 and 0.63 for the 7 years lookback (See Table 1). Participants who self-reported race as white had the highest score with 0.61 for a 7-year lookback, while participants who self-reported black and Asian races had 0.51 and 0.52, respectively. Participants who reported their annual income less than 10k had the lowest score in all lookback years with 0.49 using 7-year lookback. Moreover, participants who earned GED, did not graduate high school, or have unknown education level had the lowest scores with 0.51 for a 7-year lookback. The AUC and PR for the original model were similar to weighted and non-weighted methods (Table 2). The weighted and non-weighted methods had better calibration with 0.1242 for non-weighted vs 0.2657 for the original model (See Table 2). The majority of HPO had lower Brier scores when using non-weighted methods compared to the original model. We only showed the comparison for non-weighted vs original model using 3-year lookback period.

Table 1. Median loyalty score by lookback periods stratified by demographic and socio-economic factors

Demographics	**Subcategories**	**1 yr**	**2 yr**	**3 yr**	**4 yr**	**5 yr**	**6 yr**	**7 yr**	**Percent**
Age group	18-34	0.2065	0.3468	0.4150	0.4328	0.4891	0.4921	0.4921	13.67%
	35-49	0.3399	0.4269	0.4921	0.4941	0.5099	0.5119	0.5188	22.39%
	50-64	0.4150	0.4921	0.5099	0.5375	0.5662	0.5702	0.5751	27.52%
	65+	0.4921	0.5257	0.5692	0.5870	0.6067	0.6206	0.6294	36.41%
Gender	Female	0.4150	0.4921	0.5099	0.5346	0.5662	0.5692	0.5751	60.92%
	Male	0.4042	0.4783	0.5040	0.5168	0.5435	0.5692	0.5692	37.12%
	Non-Binary	0.4180	0.5010	0.5296	0.5672	0.5692	0.5761	0.5781	0.39%
	Transgender	0.4160	0.4921	0.5069	0.5099	0.5188	0.5336	0.5405	0.17%
	Unknown	0.3755	0.4486	0.4921	0.5099	0.5227	0.5435	0.5613	1.41%
Race	Asian	0.3498	0.4328	0.4921	0.5010	0.5099	0.5099	0.5188	3.04%
	Black/African American	0.2698	0.4042	0.4269	0.4783	0.4921	0.4931	0.5069	15.50%
	Other	0.3389	0.4170	0.4792	0.4921	0.5099	0.5099	0.5208	6.47%
	Unknown	0.2767	0.4150	0.4328	0.4921	0.4990	0.5089	0.5099	16.15%
	White	0.4486	0.5099	0.5662	0.5800	0.5929	0.5949	0.6067	58.84%
Income	less than 10k	0.1571	0.3286	0.3982	0.4269	0.4664	0.4921	0.4921	11.28%
	10k to 25k	0.3498	0.4328	0.4921	0.5069	0.5188	0.5435	0.5613	10.54%
	25k to 35k	0.3636	0.4575	0.4921	0.5099	0.5385	0.5642	0.5692	6.81%
	35k to 50k	0.4150	0.4921	0.5099	0.5524	0.5692	0.5781	0.5860	8.14%
	50k to 75k	0.4328	0.5099	0.5524	0.5692	0.5850	0.5929	0.5929	11.35%
	75k to 100k	0.4694	0.5158	0.5692	0.5840	0.5929	0.5978	0.6107	9.14%
	100k to 150k	0.4891	0.5336	0.5702	0.5870	0.5939	0.6107	0.6117	11.61%
	150k to 200k	0.4921	0.5346	0.5731	0.5929	0.5998	0.6107	0.6146	5.54%

	more than 200k	0.4921	0.5346	0.5751	0.5929	0.6018	0.6107	0.6196	7.65%
	Unknown	0.3409	0.4190	0.4664	0.4921	0.5099	0.5099	0.5188	17.95%
Education	Less than GED	0.2184	0.3419	0.4150	0.4328	0.4911	0.4921	0.4921	7.54%
	Twelve Or GED	0.2747	0.4150	0.4328	0.4921	0.4931	0.5069	0.5099	16.97%
	College One to Three	0.4150	0.4921	0.5040	0.5247	0.5662	0.5692	0.5702	25.79%
	College Graduate	0.4269	0.5049	0.5346	0.5692	0.5791	0.5870	0.5929	23.99%
	Advanced Degree	0.4891	0.5257	0.5702	0.5929	0.5958	0.6107	0.6117	23.58%
	Unknown	0.3261	0.4150	0.4328	0.4921	0.4921	0.5040	0.5099	2.12%

Table 2. The performance of original model vs non-weighted model vs weighted model

Method	**AUC**	**AUPR**	**Brier Score**	**Expected Calibration Error**
Original	0.8138	0.8254	0.2657	0.2066
Non-weighted	0.8148	0.8128	0.1242	0.1497
Weighted	0.8178	0.8151	0.1288	0.15

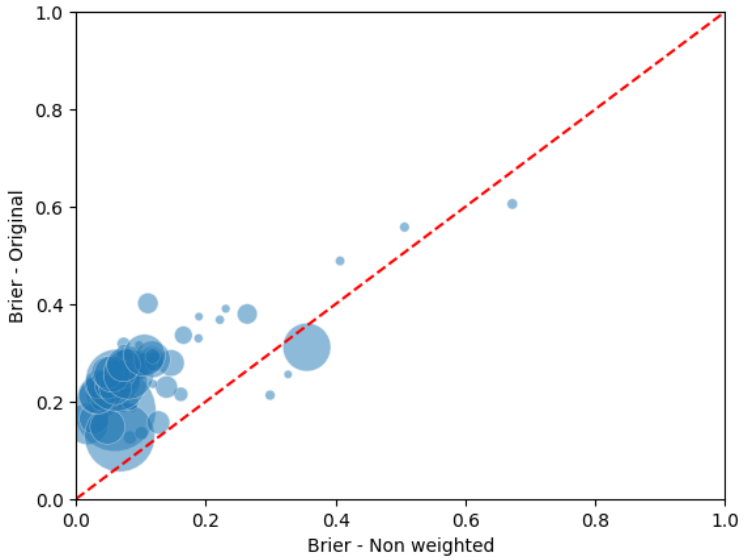

Figure 1. Comparing Brier between the original model with original coefficient and the retrained non-weighted model for all HPO sites

Discussion and Conclusion

In this study, we adapted the loyalty score from i2b2 to the OMOP Common Data Model. Our script enabled the measurement of EHR continuity across any OMOP-compliant dataset. We observed wide variation in EHR continuity between sites, regardless of the lookback period. This variability suggests that the completeness of preventive care documentation differs across sites—an important factor to consider when conducting studies using the All of Us dataset.

On average, larger sites had higher loyalty scores compared to smaller sites. This may reflect differences in OMOP mapping quality across institutions. However, low loyalty scores at large sites may be due to EHR fragmentation, potentially indicating that participants receive specialty care rather than primary or preventive care at those institutions. We also found that models retrained on the All of Us dataset were better calibrated than models using the original parameters, highlighting the benefit of adapting models to the target dataset.

Our results revealed substantial differences in EHR continuity between white and non-white participants, as well as across socioeconomic indicators (e.g., income and education). These differences may reflect disparities in healthcare access and utilization, which could in turn influence the validity of downstream analyses. The loyalty score can help researchers adjust for such biases in their studies. Furthermore, it can be employed in multi-site research to identify institutions with low EHR continuity and guide efforts to improve data completeness at those sites.

References

1. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. eGEMs (Generating Evidence & Methods to improve patient outcomes). 2016;
2. Lin KJ, Rosenthal GE, Murphy SN, Mandl KD, Jin Y, Glynn RJ, et al. External validation of an algorithm to identify patients with high data-completeness in electronic health records for comparative effectiveness research. Clin Epidemiol. 2020;12.
3. Klann JG, Henderson DW, Morris M, et al. A broadly applicable approach to enrich electronic-health-record cohorts by identifying patients with complete data: a multisite evaluation. J Am Med Inform Assoc. 2023;30(12):1985-1994. doi:10.1093/jamia/ocad166