A General Statistical Framework for Quality Assessment of Electronic Health Records Data

MSR137

Jiapeng Xu [1] Hui Wang [1] Ying Lu [2]

[1]Lumbrita LLC, 16185 Los Gatos Blvd, Los Gatos, CA 95032 [2]Department of Biomedical Data Science, School of Medicine, Stanford University, 94305, Stanford, CA, USA

Introduction

- There are concerns about the reliability and the quality of EHR data for use in clinical research
- To address this challenge, we propose a quantitative, score-based QA framework for categorical and continuous EHR data
- We will present demo results for sex and weight in this poster

QA for Categorical Variable (Sex)

Methods

Data

- Self-reported sex can have inconsistency with the patient's diagnosis history, e.g., male patients have pregnancy code, etc.
- Input data: patient-level observed counts of female/male restrictive ICD and RxNorm code: N_f and N_m
- A Bayesian model to assess the reliability of self-reported sex in EHR.

Model:

- Counts of female and male restrictive ICD and RxNorm codes follow a binomial distribution, respectively, $N_f \sim Binomial(N_s, p_f)$ and $N_m \sim Binomial(N_s, p_m)$
- The prior probability p_f and p_m follows a Beta distribution: $p_f \sim Beta(a_{0f}, b_{0f})$, and $p_m \sim Beta(a_{0m}, b_{0m})$.
- The posterior distribution of p_f and p_m is: $p_f|N_f, N_m \sim Beta(a_{0f} + N_f, b_{0f} + N_m)$, and $p_m|N_f, N_m \sim Beta(a_{0m} + N_f, b_{0m} + N_m)$.
- Prior parameters a_{0f} and b_{0f} were estimated from the empirical distribution of female/male restrictive ICD and RxNorm codes.

Decision Rule and Scoring:

- CDF of posterior probability of $p_f > 0.8$: $P(p_f > 0.8)$
- Female/male Likelihood Ratio: $\frac{P(p_f>0.8)}{P(p_m>0.8)}$

This framework enables probabilistic inference for the likelihood of a patient's sex based on the observed pattern of sex-restrictive diagnosis codes, integrating prior knowledge about the distribution of diagnosis and medication history among known male and female populations.

Results

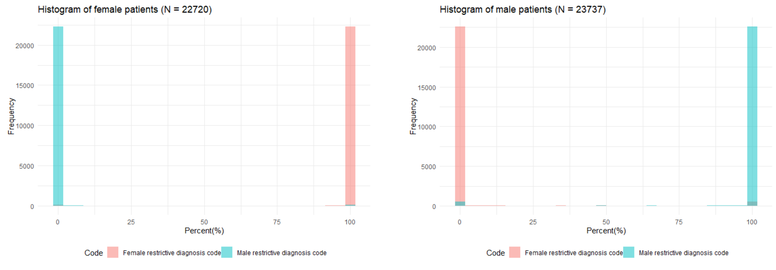

Figure 1. Histogram of female/male restrictive diagnosis code for female patients (left) and male patients (right)

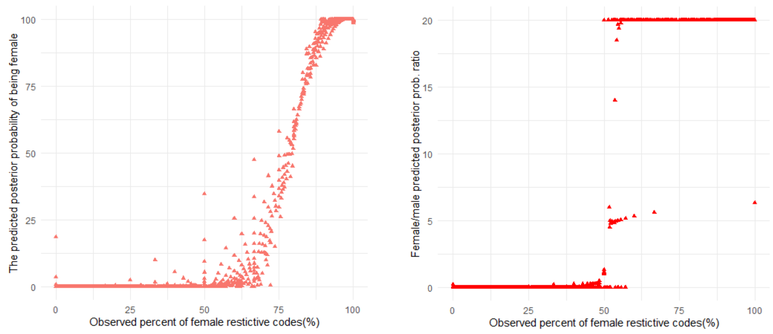

Figure 2. CDF of posterior probability and female/male likelihood ratio

- Confusion matrix of female/male classification using likelihood ratio: $\frac{P(p_f>0.8)}{P(p_m>0.8)}$;
  Decision Rule: Female if $\frac{P(p_f>0.8)}{P(p_m>0.8)} > 10$; Male if $\frac{P(p_f>0.8)}{P(p_m>0.8)} < 3$; Otherwise, unknown

	Posterior-inferred sex		
Self-reported gender	Female	Male	Unknown
Female	20684	**316**	**1719**
Male	**1632**	21515	**590**
Unknown	**39**	**34**	**4**

Table 1. Confusion matrix of female/male classification using likelihood ratio

QA for continuous Variable (Weight)

Methods

Input data: observed weight in each patient over time. We analyzed weight data (a total of 25,548,357 weight records) in an MVP (Million Veteran Program) cohort of 496,311 patients using VA EHR data between year 2000 and 2016.

Model

- Traditional method for continuous variables QA is thresholding
- We introduce a method based on longitudinal observations at patient-level using the Exponentially Weighted Moving Average (EWMA)
- The EWMA $\bar{y}_{i,EWMA}$ for a measurement y_i taken at time t_i, $t_1 \leq t_i \leq t_n$, is defined as a weighted average over the entire sequence:

$$\bar{y}_{i,EWMA} = \frac{\sum_{j=1}^{n} w_j y_j}{\sum_{j=1}^{n} w_j}$$

- EWMA detects implausible deviations by comparing each measurement against a weighted average of nearby values in time
- EWMA accounts for temporal dependencies: weights decaying exponentially based on time gaps
- Longitudinal QA score (Q_R):

$$Q_R = 2(1 - \Phi(\frac{|y_i - \bar{y}_{i,EWMA}|}{SE})$$

- Thresholding QS score (Q_S):

$$Q_S = 2(1 - \Phi(\frac{|y_i - \bar{y}|}{SE})$$

Validation

We randomly selected 100 patients whose proportion of flagged measurements were between 0 and 20% using a p-value cutoff of 0.05. The data of these patients were manually reviewed independently by two biostatisticians for identification of problematic measurements. All discrepancies between the two reviewers were reviewed and called independently by a third biostatistician. All reviews were blinded to the results of the algorithm.

Results

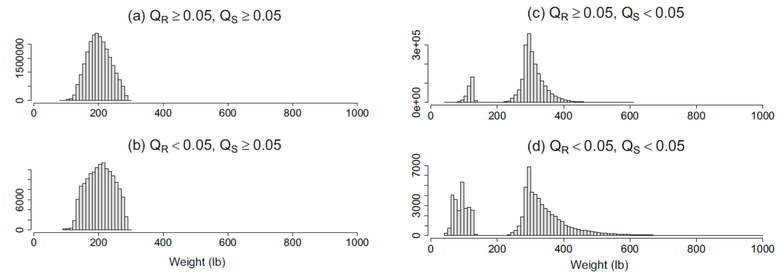

Figure 3. Histograms of weight measurements (lb) stratified by the agreements between Q_R and Q_S. (Q_R: longitudinal QA score; Q_S: thresholding QA score)

	Longitudinal (Q_R)	**Thresholding (Q_S)**
False positive rate (FPR)	0.3%	12.9%
Power	75.4%	28.0%
Positive predictive value (PPV)	71.4%	2.2%
Negative predictive value	99.7%	99.1%

Table 2. Validation results for weight records (Q_R and Q_S cutoff = 0.05)

Use case: To demonstrate how the quality of weight data impacts BMI classification, we computed BMI in our MVP cohort using all the data without any QA, data QAed with the thresholding method (QS), and data QAed with the longitudinal method (QR). The proportion of incidence of a patient who ever falls into each BMI class were then calculated. Table 3 compares these proportions among the three methods.

BMI class	**No QA**	**Thresholding QA**	**Longitudinal QA**
N	10,377,511	9,098,710	9,606,933
Underweight	4.54%	1.47%	2.24%
Normal to overweight	75.56%	76.46%	73.07%
Obese Class I/II	64.36%	61.68%	61.06%
Obese Class III	14.54%	6.69%	11.39%

Underweight: BMI<18.5; Normal to overweight: BMI≥18.5 and <30; Obese Class I/II: BMI≥30 and <40; Obese Class III: BMI≥40.

Table 3. Proportion of subjects who ever had any BMI in the listed BMI classes

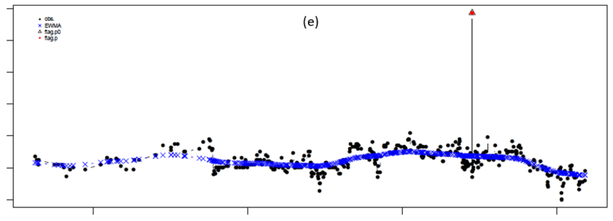

Figure 4. An illustration of EWMA longitudinal QA with weight observations in a patient

Reference

Wang, H., et.al. (2021). A statistical quality assessment method for longitudinal observations in electronic health record data with an application to the VA million veteran program. *BMC Medical Informatics and Decision Making*, 21, 1–8.