External validation of type II diabetes electronic phenotyping algorithms

Jitendra Jonnagaddala, PhD[1], Selva Muthu Kumaran Sathappan, MSc[2], Su Chi Lim, MBBS, PhD[3], E Shyong Tai, MB ChB, PhD[4] , Mengling Feng, PhD[2,*], Siaw-Teng Liaw, MBBS, PhD[1,*]

1. School of Public Health & Community Medicine, UNSW Sydney, Australia
2. Saw Swee Hock School of Public Health, National University of Singapore, Singapore
3.Khoo Teck Puat Hospital, Singapore
4. National University Hospital, Singapore

WHO Collaborating Centre on eHealth (AUS-135)

Never Stand Still | Medicine

Background and Objective

- Electronic Medical Record Systems (EMRs) provide unique opportunity to identify cohorts of interest. However, the identification of cohorts using EMRs is challenging, as they are not designed for clinical and epidemiological research, rather focused primarily on clinical care [1].
- In this study, we propose to use electronic phenotyping algorithms to identify type II diabetes (T2DM) patients of interest automatically. Though various algorithms have already been designed, developed and validated to identify T2DM patients in literature, majority of these algorithms are not validated internationally across different countries.
- Additionally, there is also a research gap in portability and performance of these algorithms on EHR data sources that are converted into Observational Medical Outcomes Partnership (OMOP) common data model [2]. To validate the T2DM electronic phenotyping algorithms developed by Ritchie.et.al [3] and Liaw.et.al [4] in both Australian and Singaporean settings in OMOP converted data sources.

Methods

- We implemented the rule-based algorithms developed by Ritchie.et.al [3] and Liaw.et.al [4], both of which uses three data elements such as diagnosis, medication and abnormal pathology for T2DM patient identification.
- We applied both the algorithms to OMOP CDM standard datasets from Australia and Singapore, which comprised of lab measurements, medical conditions, demographic details, hospital visit details, medications and determined accuracy by quantifying the percentage of correctly identified T2DM cases to the total T2DM cases.

Results

- T2DM dataset from Singapore had a total of 5187 patients, out of which 4897 met the criteria for T2DM and the algorithm by Ritchie.et.al [3] identified 65.9 % while the algorithm by Liaw.et.al [4] identified 92.05 %.
- On the other hand, AU-ePBRN T2DM cohort had a total of 4336 T2DM patients; out of which the algorithms identified 71.5% and 94.2% respectively. Table 1 below shows the details of the dataset used and algorithm performance.

Conclusions

- We were able to assess the performance of phenotype algorithms across national boundaries and across distinct data sources at ease by using OMOP CDM standard datasets and tools.
- In addition to understanding the regional variability behind algorithm performance, this study also proved that large- scale observational studies could benefit through a common data model.

Phenotype Algorithm	Dataset description	T2DM patient Count	OMOP Vocabulary Version	OMOP CDM Version	Accuracy (%)
PheKB	SG-T2DM cohort	4897	v5.0 JAN 2020	v5.3.1	65.9
PheKB	AU-ePBRN T2DM cohort (Aug 2019)	4336	v5.0 01-AUG-19	v5.3.1	71.5
ePBRN	SG-T2DM cohort	4897	v5.0 JAN 2020	v5.3.1	92.5
ePBRN	AU-ePBRN T2DM cohort (Aug 2019)	4336	v5.0 01-AUG-19	v5.3.1	94.2

References

1. Lasko, T.A., J.C. Denny, and M.A. Levy, Computational Phenotype Discovery Using Unsupervised Feature Learning over Noisy, Sparse, and Irregular Clinical Data. PLOS ONE, 2013. 8(6): p. e66341.
2. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform. 2015;216: 574–578. pmid:26262116
3. Ritchie, M.D., et al., Robust Replication of Genotype-Phenotype Associations across Multiple Diseases in an Electronic Medical Record. The American Journal of Human Genetics, 2010. 86(4): p. 560-572.
4. Liaw, S. T., Taggart, J., Yu, H., de Lusignan, S., Kuziemsky, C., & Hayen, A. (2014). Integrating electronic health record information to support integrated care: practical application of ontologies to improve the accuracy of diabetes disease registers. Journal of Biomedical Informatics, 52, 364

Correspondence to: Prof. Siaw-Teng Liaw siaw@unsw.edu.au or Asst.Prof Mengling Feng ephfm@nus.edu.sg