MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI230942

A Post-Marketing Drug Evaluation Framework Based on Real-World Electronic Health Records Data

Yu WANG[a], Shuang MA[a], Hua RU[a], Hongyi NI[a] and Jingsong LI[a, b ,1]
[a] *Research Center for Healthcare Data Science, Zhejiang Laboratory, Hangzhou, China*
[b] *Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education, Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China*

Abstract. Real-world data (RWD) could be a new way to evaluate the safety and efficacy of post-marketing drugs, while there is no common method for how to use RWD for drug evaluation. In this paper, we present a framework for real-world drug evaluation based on electronic medical record (EHR) data. We designed a data model customized for post-marketing drug evaluation and a unified post-marketing drug evaluation pipeline. The proposed framework can be applied to drug evaluations with different study paradigms for different purposes by flexible use of the proposed data model and pipeline. A prototype system has been developed according to the framework. Real-world EHRs in a tertiary hospital in China between 2010 and 2020 were converted to the proposed data model, and as a test case, we conducted a research on the risk of allergic reactions to cefodizime and ceftriaxone using the prototype system.

Keywords. Post-marketing drug evaluation, real-world data, electronic health records, common data model

1. Introduction

Post-marketing drug evaluation is an important tool to promote rational clinical use and effective control of pharmaceutical costs. The use of randomized controlled trials (RCTs) revolutionized medicine in the 20th century and has become the gold standard for evaluating drug efficacy. However, RCTs are not always feasible, especially for serious or rare diseases where subject recruitment is difficult and slow. [1] The volume and availability of electronically collected routine medical data, also known as real-world data (RWD), has grown rapidly in recent years and could provide a new way to assess the safety and efficacy of post-marketing drugs. Compared with RCTs, the use of RWD for drug evaluation can provide evidence of the real clinical use of post-marketing drugs and promote the safe and effective use of drugs by patients [2]; and it can also provide inspiration for the development of new drugs and drug repositioning. Although countries around the world have published policies in recent years calling for the use of RWD to

[1] Corresponding Author: Jingsong LI, Email address: ljs@zju.edu.cn.

support the development and marketing of new medicines [3], there is still no common method for drug evaluation using RWD. Currently, claims data are the primary real-world data source used for post-marketing drug evaluation, including the *Sentinel Initiative* in the United States [4], *EU-ADR* in Europe [5] and *AsPEN* in Asia [6]. In a recent analysis of the capabilities of the *Sentinel Initiative*, some of the most commonly cited reasons for the inability to use the current claims-based Sentinel for safety investigations include: lack of clinical detail to accurately identify health outcomes, lack of or inaccurate measurement of important confounding variables, etc. Electronic health record (EHR) data contain more clinical detail than claims data, and the introduction of large-scale EHR data would greatly enhance the ability to assess the efficacy and safety of post-marketing drugs. However, there are few studies on a unified framework for post-marketing drug evaluation based on EHR data.

This paper proposes a framework for post-marketing drug evaluation based on large-scale EHR data. The framework includes a drug evaluation data model as the data foundation and a unified pipeline for detection and causal evaluation of the association between drugs and health outcomes of interest (HOIs) in different scenarios of post-marketing drug evaluation.

2. Methods

The overview of the proposed post-marketing drug evaluation framework is shown in Figure 1. The Drug Evaluation Data Model (DEEM) is the data foundation of this framework, and the Drug Evaluation Pipeline (DEP) describes a unified post-marketing drug evaluation pipeline with 4 key steps, and different types of drug evaluation applications, including drug utilization description analysis, drug effectiveness and safety evaluation analysis under different clinical trial paradigms, were enabled by flexible use of DEEM and DEP.

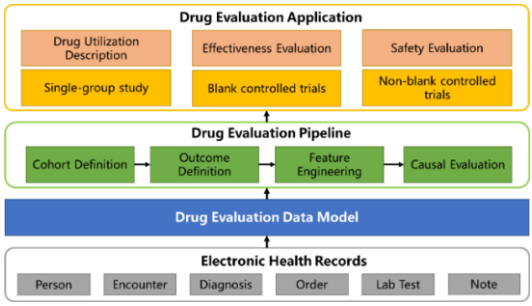

Figure 1. Overview of the post-marketing drug evaluation framework based on EHRs.

2.1 Drug Evaluation Data Model

DEDM is divided into 4 main domains: *Concept*, *Clinical Data*, *Research Data*, and *Derived Table*, as shown in Figure 2. The *Concept* domain provides standardized terminology codes for the entire framework. Based on the existing open source standard terminology data, terminologies were integrated and classified based on the definition of existing standard terminology sets. EHR data are cleaned and transformed by ETL tools and mapped to standardized concepts. Similar to OMOP Common Data Model[8] and Sentinel Common Data Model, DEDM stores clinical data in the *Clinical Data* domain;

demographic information, visits, diagnoses, prescriptions, procedures, lab tests, etc. are stored in a patient-centered manner. The demographic information, visits, diagnoses, prescriptions, lab test results and procedures are derived from the structured data in EHR, while the information that we have extracted from the text data using natural language processing tools is stored in the *Clinical Finding* table, such as "tumor location" and "tumor size". *Research Data* domain is centered on *Research* table, and configurations and results of cohort definition, outcome definition feature engineering and causal estimation models, 4 core steps in DEP, are also stored in *Research Data* domain. To flexibly adapt drug evaluation studies and bridge the gap between clinical and research data, DEDM contains two *Derived tables*: *Treatment_Period* and *HOI*. *Treatment_Period* is derived from *Clinical Data* according to user definitions in *Research data* and stores the information of complete therapy period, which is crucial to drug evaluation study. And *HOI* table stores the definitions of commonly used HOIs in drug evaluation studies, including efficacy outcomes and safety outcomes, and how they are derived from *Clinical Data*.

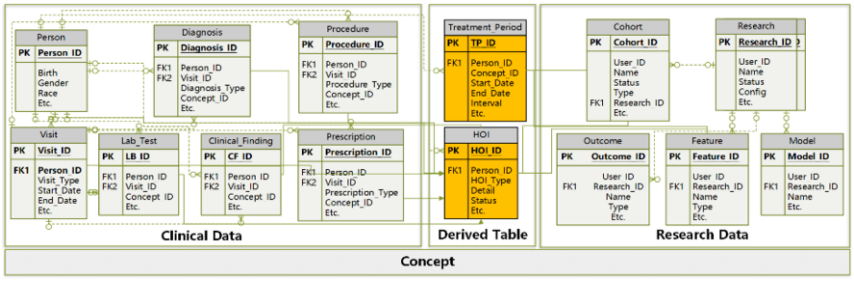

Figure 2. Entity-relation diagram of drug evaluation data model.

2.2 Drug Evaluation Pipeline

There are 4 steps in the proposed DEP: *Cohort definition*, *Outcome definition*, *Feature engineering* and *Causal estimation*, and the 4 steps are to be followed sequentially. *Cohort definition* requires users to develop a set of eligible criteria and provides screening results for the enrolled population. The eligibility criteria are stored in DEDM as the form of SQL statements. *Outcome definition* defines the outcome events of the study, either as user-defined clinical outcomes based on *Clinical Data*, or as those pre-defined in the *HOI* table. In *Feature Engineering* step confounding variables used in the study are defined and calculated, and the derived paths of both outcome events and features are stored as SQL statements as well. Finally, *Causal estimation* step enables comparing the outcomes between groups by a series of statistic or machine learning tools, and propensity score methods are employed for controlling the baseline and driving causal inference.

2.3 Drug Evaluation Application

DEDM and DEP can be applied to various clinical trial paradigms, such as single-group trials, blank-controlled trials, and non-blank-controlled trials, etc. Based on this framework, users can perform a variety of drug evaluation applications, including drug utilization description analysis, drug efficacy analysis, drug safety analysis, etc.

Based on the above framework, we have developed a prototype drug evaluation system. The presentation layer uses HTML and VUE, the business layer uses JAVA technology stack, data is stored in Oracle, and data analysis is performed by Python.

3. Results

We converted the EHR data of more than 9.5 million patients between 2010.1 and 2020.5 from a tertiary hospital in China into DEDM. The concept domain covered 985,404 concepts from more than 15 standard concept categories and more than 15 different types of standard terminology systems. Concepts not included in the existing terminology but used in Chinese clinical practice were also retained and coded, and reviewed by clinical physicians.

The prototype drug evaluation system was developed according to DEP. As a test case, we conducted a research on the risk of allergic reactions to cefodizime and ceftriaxone based on the converted data, which was a non-blank controlled trial for safety evaluation. The research design and the automically generated evaluation report in the prototype are shown in Figure 3.

Figure 3. Entity-relation diagram of drug evaluation data model.

4. Discussion

In this paper, we proposed a framework for post-marketing drug evaluation based on real-world EHR data, the core of which is a data model tailored for drug evaluation. This data model not only includes clinical data necessary for drug evaluation, but also incorporates research designs, which are usually conducted offline manually, into the data model. On the other hand, the framework embeds HOIs commonly used in drug evaluation studies into DEDM to facilitate the user to generate drug-efficacy/safety relationships in bulk, improving the efficiency of post-marketing proactive evaluation.

The FDA's Sentinel Program has become an important strategic resource for the FDA, but its operation is largely dependent on the needs of the FDA and its coordinating centers. OHDSI's predecessor, OMOP, was established for drug evaluation, and has a number of tools to help users conduct drug evaluations, such as Atlas. However, the toolkit provided by Atlas requires users to have a lot of knowledge about the OHDSI components, i.e. vocabulary, concept sets, etc. Key steps of drug evaluation in Atlas are not organized into a unified pipeline. The framework proposed in this paper specifies a pipeline from converting EHR data into DEDM, to cohort construction customized for different clinical trial paradigms, then to confounder selection and extraction, and finally

to causality assessment of association, making the drug evaluation process more coherent and centralized.

Although the utility of our proposed work has been addressed, some challenges exist, and further work is needed. Mapping EHR data to standardized concepts is essential to the effectiveness of drug evaluation analysis. Efforts are needed for continuously optimizing the algorithm for term mapping. On the other hand, the trial paradigms evolved in the proposed framework are limited, more paradigms will be integrated into the framework in the future to achieve higher applicability.

5. Conclusions

In this paper, a framework for drug evaluation based on EHR data is proposed and a prototype system is developed. Based on a data model customized for the drug evaluation process, a unified pipeline can be applied to different scenarios of post-marketing drug evaluation. The system can help users to conduct drug evaluation researches more easily and automatically generate evaluation result reports.

Acknowledgement

This work was supported by the National Key Research and Development Program of China (No. 2022YFC2504605) and Key Research Project of Zhejiang Lab (No. 2022ND0AC01).

References

[1] A. Lavertu, B. Vora, K.M. Giacomini, R. Altman, and S. Rensi, A New Era in Pharmacovigilance: Toward Real‐World Data and Digital Monitoring, *Clinical Pharmacology & Therapeutics* **109** (2021), 1197-1202, doi: https://doi.org/10.1002/cpt.2172.

[2] R. Liu, S. Rizzo, S. Whipple, N. Pal, A.L. Pineda, M. Lu, B. Arnieri, Y. Lu, W. Capra, R. Copping, and J. Zou, Evaluating eligibility criteria of oncology trials using real-world data and AI, *Nature* **592** (2021), 629-633, doi: https://doi.org/10.1038/s41586-021-03430-5.

[3] B. Schurman, The framework for FDA's real-world evidence program, (2019),

[4] R.J. Desai, M.E. Matheny, K. Johnson, K. Marsolo, L.H. Curtis, J.C. Nelson, P.J. Heagerty, J. Maro, J. Brown, S. Toh, M. Nguyen, R. Ball, G. Dal Pan, S.V. Wang, J.J. Gagne, and S. Schneeweiss, Broadening the reach of the FDA Sentinel system: A roadmap for integrating electronic health record data in a causal analysis framework, *npj Digital Medicine* **4** (2021), 170, doi: https://doi.org/10.1038/s41746-021-00542-0.

[5] V.K. Patadia, P. Coloma, M.J. Schuemie, R. Herings, R. Gini, G. Mazzaglia, G. Picelli, C. Fornari, L. Pedersen, J. van der Lei, M. Sturkenboom, and G. Trifirò, Using real-world healthcare data for pharmacovigilance signal detection – the experience of the EU-ADR project, *Expert Review of Clinical Pharmacology* **8** (2015), 95-102, doi: https://doi.org/10.1586/17512433.2015.992878.

[6] N. Pratt, M. Andersen, U. Bergman, N.K. Choi, T. Gerhard, C. Huang, M. Kimura, T. Kimura, K. Kubota, and E.C.C. Lai, Multi‐country rapid adverse drug event assessment: the Asian Pharmacoepidemiology Network (AsPEN) antipsychotic and acute hyperglycaemia study, *Pharmacoepidemiology and Drug Safety* **22** (2013), 915-924, doi: https://doi.org/10.1002/pds.3440.

[7] J.S. Brown, J.C. Maro, M. Nguyen, and R. Ball, Using and improving distributed data networks to generate actionable evidence: the case of real-world outcomes in the Food and Drug Administration's Sentinel system, *Journal of the American Medical Informatics Association* **27** (2020), 793-797, doi: https://doi.org/10.1586/17512433.2015.992878.

[8] I. Reinecke, M. Zoch, C. Reich, M. Sedlmayr, and F. Bathelt, The Usage of OHDSI OMOP - A Scoping Review, *Stud Health Technol Inform* **283** (2021), 95-103, doi: 10.3233/shti210546.