Appendix: Summary of Best Papers Selected for the 2020 Edition of the IMIA Yearbook, CRI Section

Paddock S, Abedtash H, Zummo J, Thomas S

Proof-of-concept study: Homomorphically encrypted data can support real-time learning in personalized cancer medicine

Data protection, and in particular complying with the GDPR, when reusing real world data for research is recognised to be challenging. This can be an important barrier to scaling up personalised medicine research, when it may be difficult or impossible to robustly anonymise very fine-grained health and treatment history and biological profiles. This study applied the technique of Homomorphic Encryption (HE) to the problem, which allows the source data to remain encrypted when a research query is executed. The only knowledge needed in advance would be the structure of the database and its semantics such as the terminology systems used. Analysis using HE means that personal data does not need to be disclosed to any person or computational process during query execution. In this proof of concept study, a previously published HE technique was used to encrypt a database containing 5000 simulated patient records modelled on personalized medicine treatment scenarios. The authors sought to detect outlying (and therefore rare) drug responses from the data. Exceptional treatment response queries were performed directly on the encrypted data via HE, using conventionally available computing facilities. They found that the queries were able to return correct responses, permitting their sample research questions about exceptional drug response to be answered. Although this method is more time consuming to perform the queries, the authors argue that this computational time is relatively small within the total time frame of a personalised medicine study. This research was included as a best paper because it demonstrates a valuable method for more data sharing and distributed data querying in cases where the fine granularity of the clinical/molecular data does not permit robust anonymisation.

Suchard MA, Schuemie MJ, Krumholz HM, You SC, Chen R, Pratt N, Reich CG, Duke J, Madigan D, Hripcsak G, Ryan PB

Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis

At present, clinical guidelines for the treatment of hypertension recommend several different classes of drug in patients without a comorbidity or significant risk factor. Thiazide or thiazide-like diuretics, angiotensin-converting enzyme (ACE) inhibitors, angiotensin receptor blockers, dihydropyridine calcium channel blockers, and non-dihydropyridine calcium channel blockers are all candidates for consideration, with no prioritisation amongst this list. Given that the difference in treatment effects between them may be very small, a classical Randomised Controlled Trial would not be feasible to identify if one of these candidates is more effective. This paper reports the results of a large-scale real world data study to identify an effectiveness difference and also to demonstrate the potential value of a federated health data network for big data observational research. The authors undertook a distributed analysis of 4.9 million patient records in retrospective cohorts. These patients were from nine claims and EHR databases in the US, Japan, South Korea and Germany, within the LEGEND-HTN study which is comparing antihypertensive drug treatments. The analysis was undertaken through the Observational Health Data Science and Informatics (OHDSI) distributed data network. This platform mapped all of the data sources to the Observational Medical Outcomes Partnership (OMOP) common data model, which permitted uniform federated (distributed) analysis queries to be executed. Fifty-five health outcomes were studied. Thiazide-like diuretics had better effectiveness than ACE inhibitors to reduce the incidence of acute myocardial infarction, hospitalisation for heart failure, and stroke risk. This research was included as a best paper because it reports the largest scale study on this topic, which was only possible because of the federated database network, coupled with a robust study methodology. It demonstrates the evidence-generation value of federated health data networks.

Yu Y, Ruddy KJ, Hong N, Tsuji S, Wen A, Shah ND, Jiang G

ADEpedia-on-OHDSI: A next generation pharmacovigilance signal detection platform using the OHDSI common data model

Underreporting of adverse drug events (ADEs) is a key challenge in drug safety surveillance. Although a valuable resource for pharmacovigilance, the US Food and Drug Administration (FDA)'s Adverse Event Reporting System (FAERS) only capture adverse drug reactions (ADRs) spontaneously reported by healthcare professionals, patients, and pharmaceutical manufacturers. Longitudinal observational databases like Electronic Health Records (EHRs) and transactional claims can be used as additional data sources for pharmacovigilance to address gaps in coverage and increase population heterogeneity. In order to integrate the data from those two types of sources - spontaneous reporting system (SRS) and EHRs – with different data models and vocabularies, Yu *et al.,* considered the use of the OMOP common data model (CDM). This model is not only increasingly adopted by data research networks leveraging EHRs data but has also been intensively used to identify and assess associations between medical interventions and health-related outcomes in many pharmacovigilance and pharmacoepidemiology studies. The authors converted into the OMOP format the last version of the FEARS data base (including 4,619,362 adverse event cases reported between 2012 and 2017). A dedicated tool has been developed to extract, transform, and load (ETL) the FEARS data into the OMOP data base (version 5). An important part of the ETL process was dedicated to terminology mappings of the drug names to RxNorm

and of the adverse events, indications, and outcomes to SNOMED CT. The structure mapping between FEARS tables and OMOP CDM required multiple rounds of discussion involving two experts with medical informatics background. The evaluation of the work was two-fold. The authors first validated the mappings and the conversion process and conducted a replication study in order to evaluate the impact of the conversion and information loss on signal detection.

This paper was selected as a best paper because it demonstrates with a robust methodology not only the feasibility of converting a SRS data base into the OMOP format but also the accuracy of the resulting framework called ADEpedia-on-OHDSI and its capability to improve signal detection through standardization. Furthermore, this work paves the way for seamless integration of SRS with EHRs or other RWD enabling better signal detection and further discovery about adverse events such as causes, confounders, or possible corrective actions.