Article published online: 2023-12-26

Appendix: Summary of Best Papers Selected for the IMIA Yearbook 2023, CRI Section

Ahuja Y, Zou Y, Verm, A, Buckeridge D, Li Y

MixEHR-Guided: A guided multi-modal topic modeling approach for large-scale automatic phenotyping using the electronic health record

J Biomed Inform 2022 Oct;134:104190. doi: 10.1016/j.jbi.2022.104190

It can be challenging to identify a disease cohort within an EHR repository, especially when precision is required to accurately identify patients with a particular disease variant, biomarker specificity or other precisely defined criteria. EHRs may contain a number of working diagnoses that have been assumed during a diagnostic pathway and may contain a number of working diagnoses that have been assumed during a diagnostic pathway and through disease evolution. Electronic phenotyping is a process of defining a set of EHR data item values that, when found together, are highly suggestive of a particular disease, or conversely may establish its absence. This yearbook chapter has included electronic phenotyping algorithms in previous years, but this methodology by Ahuja *et al.* has been included this year because it offers an advance on previous methods. To maximise accuracy, the authors utilise highly dimensional EHR data, and have mapped these to a portfolio of reference phenotypes in order to enhance the efficiency over the classical Latent Dirichlet Allocation (LDA) approach. The authors have modelled around 1,500 phenotype topic maps, and have demonstrated a high performance of matching 1.3 million patients in Québec, Canada to these phenotype topic maps, which can be performed simultaneously. This methodology may therefore advance the ability to construct virtual cohorts and perform precisely targeted big data analyses on heterogeneous health data.

Gruendner J, Deppenwiese N, Folz M, Köhler T, Kroll B, Prokosch HU, Rosenau L, Rühle M, Scheidl MA, Schüttler C, Sedlmayr B, Twrdik A, Kiel A, Majeed RW

The Architecture of a Feasibility Query Portal for Distributed COVID-19 Fast Healthcare Interoperability Resources (FHIR) Patient Data Repositories: Design and Implementation Study

JMIR Med Inform 2022 May 25;10(5):e36709. doi: 10.2196/36709

In this paper, Gruendner and colleagues report the design, implementation and evaluation of a platform with a user query design tool to author clinical trial eligibility criteria as electronic health record queries, and to execute those across a network of hospital electronic health record systems. The reuse of EHRs for clinical trial feasibility is not new, but methods act now have required the export of EHR data into a separate clinical data warehouse, often utilising an OMOP or i2b2 architecture. The authors here have utilised HL7 FHIR and a FHIR-specific standard query formalism. The research incorporates the logical sequence of a review of the literature regarding existing tools and methods, a requirements analysis, an architecture design and implementation, and validation using synthetic data distributed across a number of sites in Germany that are part of the Medical Informatics Initiative. Although utilising a standard query representation, the author and environment offers a non-technical friendly interface for constructing the eligibility queries, guided by an ontology. The advantage of this approach is that it can be executed on FHIR servers, which are growing in popularity as a repository architecture for electronic health record information. It is therefore possible through this methodology for a healthcare organisation such as the hospital to leverage the technologies it already has in place and, quite importantly, the skills and expertise of staff it is more likely to already have in the house, to create an EHR endpoint for these distributed queries.

Peters U, Turner B, Alvarez D, Murray M, Sharma A, Mohan S, Patel S

Considerations for Embedding Inclusive Research Principles in the Design and Execution of Clinical Trials

Ther Innov Regul Sci 2023 Mar;57(2):186-95. doi: 10.1007/s43441-022-00464-3

This paper is, rather unusually, being included is the best paper although it is a review paper. The authors present a very logical and well researched analysis of the disparities and biases in the populations that are recruited into clinical trials, leading to a lack of representativeness of the population from a number of demographic perspectives. They highlight in particular race, ethnicity, socio-economic factors, underserved communities and disparities in the approach to reimbursing trial participants for out-of-pocket expenses directly incurred through the process of trial participation. The authors present evidence of these categories of disparity, and argue for the risk that clinical trial findings will therefore not be broadly applicable to the populations intended to be treated. The paper also goes into mitigations: approaches that can be taken to broader equity and inclusion, to improve the representation within trials of true population diversity. This paper has been included in the yearbook because it presents a clear and convincing case and call to action for all of those involved in clinical research to take measures proactively to ensure clinical trial accessibility and inclusion is equitable and that study populations are genuinely representative.

Zenker S, Strech D, Ihrig K, Jahns R, Müller G, Schickhardt C, Schmidt G, Speer R, Winkler E, von Kielmansegg SG, Drepper J

Data protection-compliant broad consent for secondary use of health care data and human biosamples for (bio)medical research: Towards a new German national standard

J Biomed Inform 2022 Jul;131:104096. doi: 10.1016/j.jbi.2022.104096.

This paper addresses a challenge that every European country, and the EU as a whole, struggles with when seeking to find an acceptable approach to the reuse of health data for research. It is often difficult to robustly anonymise health data. Linkage may be required to join records between multiple care providers such as a hospital and a GP, and longitudinally to update health record extracts over time. The presence of linkage identifiers means the data is pseudonymised, and normally regarded as personal data under the European General Data Protection Regulation. Some health data types, some uses such as AI development, and the case of rare diseases with small patient numbers or make it difficult to be rigorous in anonymisation. In such cases informed consent is the usual legal basis for processing health data for research. The difficulty is that the GDPR normally requires that consent is fully informed and specific. Health data ecosystems, in contrast, accumulate health data at scale in order to serve multiple future research purposes that cannot be predicted at the time of obtaining consent. The holy grail solution is to seek broad consent from patients to categories of data use, such as categories of research, but so far it has proved challenging for broad consent to be accepted by data protection legislators. The German Medical Informatics programme has undertaken an extensive process of multi-stakeholder consultation, including patient representatives, on how broad consent might be worded and governed, such that it could be acceptable to them and to decision-makers. The consultation process included all 52 German ethics committees and all 18 German Federal and state data protection authorities. It has now been recognised authoritatively as an acceptable process for obtaining broad consent for research using health data. This paper, which was the top ranked in the peer review process, reports on the reasons for pursuing this and the methodology that was adopted in order to obtain the essential endorsements. There are links in the paper to the actual broad consent wording, in English and German. This initiative is the first across Europe to have created a legal and acceptable basis for broad consent and offers a pathway that other countries could now pursue.