

## Appendix: Summary of Best Papers Selected for the 2021 Edition of the IMIA Yearbook, CRI Section

Bell SK, Delbanco T, Elmore JG, Fitzgerald PS, Fossa A, Harcourt K, Leveille SG, Payne TH, Stametz RA, Walker J, DesRoches CM

### Frequency and types of patient-reported errors in electronic health record ambulatory care notes

JAMA Netw Open 2020 Jun 1;3(6):e205867

The accuracy of electronic health record data matters more than ever, especially due to the proliferation of clinical decision support, workflow systems and learning health systems. Research, usually in general practice, has established that patients can discover errors and advise on the correction that is needed. This paper reports on a major US study involving 22,889 patients who have had access to their hospital or community practice EHRs and read them. They were invited to review their online health data and to respond by means of a questionnaire. (A larger pool of patients were invited, and 21.7% responded.) Of those, around 21% identified an error in their records and 40% of those regarded those errors as serious (one in ten indicated “very serious”). These errors were mostly related to errors in documented diagnoses (incorrect or missing), also in medication, allergies and procedures. The authors note that “Older and sicker patients were twice as likely to report a serious error compared with younger and healthier patients, indicating important safety and quality implications”. These errors were rated by patients themselves, which may not be the same as clinician judgement of importance (which is not discussed). The respondents reported mixed reactions from healthcare professionals to these errors, sometimes they were corrected promptly and another times repeated requests still resulted in no rectification of the error. The authors recognise the low response rate and discuss this. There will inevitably be unknown selection biases in terms of health status and social diversity. The significance of this paper is to highlight

the relatively high frequency with which health records contain errors, and also the value of patients in helping to correct them.

Estiri H, Strasser ZH, Murphy SN

### High-throughput phenotyping with temporal sequences

J Am Med Inform Assoc 2021 Mar 18;28(4):772–81

Personalized medicine research and machine learning require the characterization of patients into stratified subpopulations, accurately enough to allow differentiation of disease course, treatment effectiveness and outcomes. However, clinical documentation is intended primarily to serve continuity of care decision-making and medico-legal record-keeping, and scientific research purposes are not usually the priority. It can therefore be difficult to infer facts like the date when a health condition first arose in a patient or to link a causal series of events such as treatment changes and complications. Most advanced phenotyping initiatives developed for cohort building do not utilize the temporal dimension of the disease progression or treatment outcomes. Estiri *et al.* aimed at utilizing a vector sequential representation of the EHR data combined with a novel sequential pattern mining (SPM) algorithm to characterize temporal relationships between EHR data instances – in particular diagnoses, complications and medications – and to significantly improve the performance of a high throughput feature selection algorithm to predict phenotypes. A representation mining algorithm was first developed to construct from EHR diagnosis, and medication records five classes of feature sets: a baseline representation for computational phenotyping (aggregated vector representation (AVR), temporal sequential representations using two different algorithms (SPM and tSPM) and combined classes (AVR+SPM, AVR+tSPM). A computational phenotyping algorithm was then trained on the five feature sets extracted from the data from the Mass General Brigham Biobank to predict 10 phenotypes and evaluated against the gold standard labels from validated disease cohorts. The results show improved performance across all 10 phenotypes compared to

existing classifiers published in the literature. This paper demonstrates that sequencing the diagnoses and medications results in rich feature sets having the capability to enhance the performance of downstream phenotyping algorithms. This new method enables new insights in disease trajectories and help to improve the accuracy of future machine learning and the delivery of personalised medicine.

Geva A, Stedman JP, Manzi SF, Lin C, Savova GK, Avillach P, Mandl KD

### Adverse drug event presentation and tracking (ADEPT): semiautomated, high throughput pharmacovigilance using real-world data

JAMIA Open 2020 Oct;3(3):413–21

Pharmacovigilance based on real world data is challenging because clinicians rarely document adverse drug events (ADEs) in a structured form, might not document a symptom as being due to a medication item in free text, and might not recognise a symptom as being attributable to a medicine. Extracting this information in a fully computable form by natural language processing (NLP) is not usually accurate enough. In this research, Geva *et al.* report on an optimised workflow for computer assisted annotation of the indicators of ADE in EHR data. The authors have developed a methodology and user-facing tool (ADEPT – source code available) to offer experts a visualisation of candidate occurrences of an adverse drug event within clinical narratives, with annotation tools to facilitate rapid decision making on each presented candidate match. Extracted concepts were mapped to the Unified Medical Language System Concept Unique Identifier (UMLS CUI), and colour coded by CUI class to assist with visual interpretation. The authors’ method incorporated two independent reviewers and an adjudication user for cases of non-agreement. ADEPT was validated by searching for occurrences of seizure as an ADE (and not as a co-morbidity or unrelated event) while taking sildenafil in 416 patients. Using NLP, 72 candidate mentions were identified, and screened by the expert reviewers who were on average able to arrive at a decision in

less than four minutes per patient, although only nine seconds per document per reviewer before adjudication. This is substantially less time than existing methods and would appear to be flexibly extensible to other drugs and possible ADEs. This research has been selected as a best paper because it offers an efficient advance on methods to detect ADEs by computer-assisting expert reviewers with annotated candidate mentions in clinical documents.

**Zhang Z, Yan C, Mesa DA, Sun J, Malin BA**

**Ensuring electronic medical record simulation through better training, modeling, and evaluation**

**J Am Med Inform Assoc 2020 Jan 1;27(1):99–108**

There is rapidly expanding interest for re-using health data for the spectrum of learning health system purposes including quality improvement, public health screening and in-

terventions and various kinds of research and innovation. Valid inferences from clinical data require relatively precise fine-grained information, which poses challenges for data protection, since most of these re-uses are not occurring on the basis of informed consent but on the basis that the data are considered de-identified. There have been decades of research into anonymization methods, and ways of establishing whether a dataset is sufficiently de-identified that individuals cannot be recognized within it. These methods require a difficult balance to be found between scientific utility and data protection. Various approaches have been developed to mitigate risk, including record simulation via a well-established method for generating synthetic EHR generative adversarial networks (GANs). These have the ability to generate realistic synthetic data from real records but with the loss of certain statistical properties of the real data. The objective of Zhang *et al.* was to enhance the learning model of GANs for generating

diagnoses and procedure codes and evaluate the resulting pipeline on real EHRs, based on new evaluation measures. The new GAN generator developed in this work is able to learn from smaller training data sets and with greater capability to incorporate low-prevalence concepts, utilising Wasserstein divergence. The method includes cycles that compare the generated data with real EHR data to verify their similarity whilst verifying the preservation of privacy. Two evaluation measures were designed to compare the utility and privacy of the new and existing GANs for generating categorical data, using a large billing code data set of 1 million real EHRs at Vanderbilt University Medical Center. The proposed model outperformed the state-of-the-art approaches with significant improvement without sacrificing the privacy provided by such models. This best paper shows that EHR data simulation through GANs can be substantially improved. The limitation of the method is to generate only categorical data and in a static manner.