Learning from Administrative Health Registries

Jonathan Rebane, Isak Karlsson, Lars Asker, Henrik Boström, and Panagiotis Papapetrou

Dept. of Computer and Systems Sciences, Stockholm University, Sweden
{rebane,isak-kar,asker,henrik.bostrom,panagiotis}@dsv.su.se

Abstract. Over the last decades the healthcare domain has seen a tremendous increase and interest in methods for making inference about patient care using large quantities of medical data. Such data is often stored in electronic health records and administrative health registries. As these data sources have grown increasingly complex, with millions of patients represented by thousands of attributes, static or time evolving, finding relevant and accurate patterns that can be used for predictive or descriptive modelling is impractical for human experts. In this paper, we concentrate our review on Swedish Administrative Health Registries (AHRs) and Electronic Health Records (EHRs) and provide an overview of recent and ongoing work in the area with focus on adverse drug events (ADEs) and heart failure.

1 Introduction

Swedish Administrative Health Registries (AHRs) and Electronic Health Records (EHRs) provide a valuable source of information for a patient's medical history. They typically include billing codes of diagnoses (e.g., ICD10), laboratory tests, pharmaceutical information (e.g., drug prescriptions), and clinical notes (e.g., short texts written by healthcare practitioners). Such data sources can be exploited for developing robust predictive models for solving challenging tasks within the domain of healthcare, such as detecting adverse events (AEs), as well as for understanding different variations in treatment of heart failure.

In contrast to spontaneous reports, which usually contain only a limited snapshot of the circumstances surrounding a suspected ADE for a specific individual or particular treatments for heart failure, EHRs and AHRs and provide medical practitioners and clinical pharmacologists with a much richer description of the medical history of not only individual patients but also of large groups of patients sharing a similar medical conditions or with high similarity in their recorded clinical history. Such rich and complex data sources can be effectively and efficiently processed, studied, and analyzed through the usage of advanced machine learning techniques, both in a supervised and an unsupervised manner.

In this paper, we discuss the current state-of-the-art and present recent work on Swedish AHRs and EHRs. We provide an overview of recent and ongoing work in the area with focus on adverse drug events (ADEs) and heart failure. Our main contributions include:

- we present our recent work on detecting and understanding ADEs, with focus on predictive modeling techniques, methods for mining disproportional patterns, and finally we present the ADE explorer, a tool for studying ADEs from EHRs;
- we present ongoing work on understanding the effectiveness of treatment of heart failure from AHRs, with emphasis on understanding the reasons behind the large variation of the usage rate of basic treatment by 50 hospitals in Stockholm County;
- we discuss directions for future research on our current projects involving EHRs and AHRs.

2 Detecting Adverse Drug Events

Adverse drug events (ADEs), commonly defined as undesired harms caused by the intake of medications [12], account for an increasing amount of hospitalizations and deaths worldwide [3, 7]. Adverse events are both a serious health concern, estimated to be the seventh most common cause of death in Sweden [21], and a significant burden on the health care system [15]. Although a benefit–risk analysis of newly developed drugs is already conducted during clinical trials, post-marketing detection and surveillance are often performed to detect unanticipated events. For instance, clinical trials are normally performed with a limited sample of patients, who are followed for a limited period of time. As a result, not all serious adverse events can be detected prior to market deployment, which results in drugs being withdrawn from the market due to serious adverse reactions not detected during clinical trials.

The activities related to the detection, signaling, and assessment of adverse drug events is referred to as pharmacovigilance or post-marketing drug surveillance. During post-marketing surveillance, a vast array of automatic approaches for detecting potential safety hazards of drugs have been investigated, cf. [1, 13], using various data sources, the most prominent of which is a disproportionality analysis of spontaneous individual case reports [19]. One of the main obstacles, however, with current systems for collecting and analyzing data regarding adverse drug events is the fact that serious ADEs are heavily under-reported, while known ADEs are over-reported, by both clinicians, in the case of EHRs, and by patients, in the case of individual case reports [5]. Complementary and alternative sources have thus been investigated, such as online communities [10], EHRs and other administrative health registries. The major benefit of EHRs and health registries is that they typically contain longitudinal observational data of large samples of patients, including demographic information, medical history, drug consumption with exposure time and dose information, and clinical measurements, including lab results and drug concentrations. To improve the reporting rate, systems have thus been investigated to automatically detect ADEs from electronic health records, which avoids several of the limitations present in case reports cf. [9].

Next, we discuss work that has been carried out to (a) mine disproportionate (i.e., unexpected) patterns and (b) construct predictive models that are used

to identify patients with potential adverse drug reactions. We also describe the prototype of a system where health practitioners can explore and test hypotheses with respect to adverse reactions from drugs.

2.1 Mining disproportionate patterns

In recent work[2], we have explored ways of improving ADE detection by combining sequential pattern mining with disproportionality analysis. In particular, we investigated the use of sequential pattern mining for finding frequent drug sequences, which then form the basis for the disproportionality analysis, i.e., instead of looking for unexpected drug-diagnosis pairs, the novel method will find unexpected pairs of drug sequences and diagnoses. Since the proposed method is better suited to handle drug interactions, it is expected to handle cases where a sequential administration of interacting drugs is responsible for a certain ADE. An empirical investigation of the novel method has been performed using a subset of the Stockholm EPR corpus [4]. The data used in this study consists of all diagnoses and medications for 3189 patients that have received at least one heart related diagnosis during the period 2008 - 2010.[1]

The empirical investigation showed that the proposed method indeed could discover some patterns with sufficient support and that they occur much more frequently for the patient groups with the diagnoses of interest than what is expected in general. Using frequency-based sequential mining alone would not highlight the discovered patterns as they would be ranked far behind patterns that appear more frequently in the whole patient group, i.e., independently of whether the diagnosis is present or not. On the other hand, traditional disproportionality analysis would not allow us to find candidate interactions, since that type of analysis is based on one diagnosis and one drug at a time.

The patterns discovered by the proposed method must however be treated as possible hypotheses or candidates for adverse drug interaction, rather than actual causes of the disease, since there may be many natural explanations for why a certain combination of drugs occur more frequently for patients with the diagnosis of interest than for patients in general, some of which have been pointed out for the findings concerning drugs for cardiovascular patients. Hence, any findings need to be further carefully analyzed to allow for finding true adverse drug interactions.

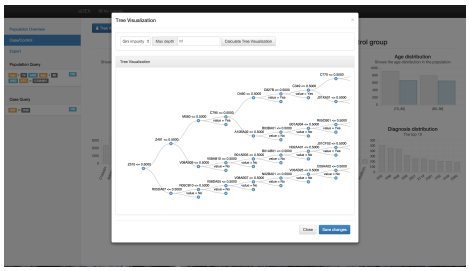
2.2 Predictive modelling of ADEs

Knowledge extraction from EHRs is a rather new research area, since EHRs were, until recently, not only relatively rare but also not easily accessible for researchers due to their sensitive nature. However, they have become more abundant and accessible for research during recent years in several countries, such as USA and Sweden. In one example study, 667,000 inpatient and outpatient EHRs

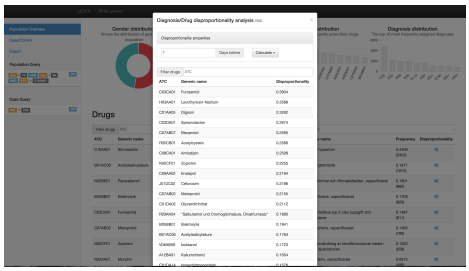
[1] This research has been approved by the Regional Ethical Review Board in Stockholm, permission number 2012/834-31/5.

were analyzed to discover new relations between drug intake and reactions. The prediction of AEs using EHRs is an ongoing research endeavor, in which most efforts to date have focused on using either structured[22] or unstructured data [6], separately. Preliminary results show that predictive performance is substantially improved by combining heterogeneous types of clinical data sources [6]. A comprehensive overview of the research area can be found in [11].

Recent work on detecting adverse drug events has shown that a bag-of-words model that exploits the set of drug prescriptions and diagnoses for a large set of patients can give promising results in terms of recall and AUC. Nonetheless, recent studies have shown that there is still plenty of room for improvement by exploiting the temporal properties of the data sources in EHRs [20]. For example, the early diagnosis or prediction of an AE can be highly correlated with a positive prognosis and timely treatment [11]. Hence, the temporal and causal features inherent in EHRs, e.g., in the form of time-evolving variables or cause-effect relations, should be modeled and exploited to the largest extent possible.



(a) Decision tree model constructed from the case and control group.



(b) Disproportionality analysis of the events in the case group as compared to the population.

Fig. 1: Exploring patients between ages 74 and 90 that have and have not been diagnosed with *essential hypertension* using (a) decision trees and (b) disproportionality analysis

2.3 Adverse drug event explorer (ADEX)

The Adverse Drug Event eXplorer (ADEX) is an exploratory prototype system for investigating and testing hypotheses regarding ADEs. The system allows medical practitioners to define, using a complex rule system, a population and a case group (e.g., patients that have experienced an adverse event and those who have not) and then explore these patients based on disproportionate events, importance of attributes, or rules. For instance, if one is interested in exploring patients older than 75 years that have and have not suffered from heart diseases, then one could define the population as `Age > 75` and the case group as `ICD`

`= I109` (see Fig. 1). These groups can then be analyzed using a plethora of different methods, which in ADEX includes decision trees, random forests and disproportionality analysis.

Although ADEX allows for practitioners to investigate and explore medically relevant hypotheses, these hypotheses are still rather limited. In particular, the tool does note allow one to explore medical trajectories of patients or to incorporate temporality in neither rules, decision trees nor pattern importances. Moreover, to fully investigate the impact of treatments the rule language used to describe case and control groups should be refined to allow one to express rules temporal rules, e.g., `ICD = I109 before ATC = C01AA01`.

3 Effectiveness of Treatment in Heart Failure Patients

The Swedish National Board of Health and Welfare has issued national guidelines for cardiac care [18], including recommendations for the treatment of patients with heart failure. These include descriptions of preferred medications based on diagnosis, severity of symptoms and factors related to the individual patient. In the recommendations, it is stated that the basic treatment of patients with heart failure should be renin angiotensin aldosterone system (RAS) inhibitors combined with beta blockers. In order to evaluate the compliance to the recommendations, the National Board of Health and Welfare has also established target levels for various indicators describing the desired portion of certain patient groups that should be eligible for specific treatments [16]. One such target is that at least 65% of the HF patients should receive basic treatment consisting of a combination of RAS-inhibitors and beta blockers. Other indicators are also defined for the use of mineralcorticoid receptor antagonists (MRA) and CRT pacemakers. In a recent evaluation of the compliance to the guidelines [17], it was noted that for heart failure, only three of the 22 county councils in Sweden reach the target levels regarding basic medication of heart failure. For the Stockholm region in total, only 57% of patients with heart failure were medicated according to the national recommendations. However, deviations from the basic treatment can often be motivated, which explains why the target level is not set to 100%. For example, the basic treatment is known to have a good effect on patients with heart failure with reduced ejection fraction (HFrEF), but there is less evidence that is has the same effect on patients with heart failure with preserved ejection fraction (HFpEF). International studies and the Swedish quality register RiksSvikt indicate that between 50 – 80 % of the patients have HFrEF [17].

In order to better understand and characterize patients that receive the basic treatment for heart failure, and what distinguishes them from patients that according to the formal requirements should, but have not, received the basic treatment, an analysis of administrative records collected in the Stockholm City Council during 2010-2016 has been made [8]. In that work, we presented results from applying frequent pattern mining on data from heart failure patients receiving basic treatment, and where these patterns are ranked according to de-

viations from the expected, using heart failure patients that have not received basic treatment as the control. These deviations were quantified by using disproportionality analysis [14] on the frequencies of the discovered patterns in the test group against the control group.

3.1 Itemset mining formalization

More concretely, we define a set of possible item labels Σ, which correspond to diagnoses or prescription codes. An itemset of size k, also called a *k-itemset*, is defined as a set $I = \{l_1, \ldots, l_k\}$ of k labels, with $l_i \in \Sigma, \forall j \in [1, k]$. A *transaction* $\mathcal{T} = \{I_1, \ldots, I_N\}$ is a set of itemsets. The size of the transaction is the number of itemsets that it contains. We say that a transaction $\mathcal{T}$ *contains* an itemset I and denote it as $I \sqsubseteq \mathcal{T}$, if there is at least one occurrence of the itemset in the transaction. Given a set of M transactions $\mathcal{D} = \{\mathcal{T}_1, \ldots, \mathcal{T}_M\}$, the *frequency* of an itemset I in $\mathcal{D}$ is equal to the fraction of transactions that contain the itemset, i.e.,

$$freq(I, \mathcal{D}) = \frac{|I \sqsubseteq \mathcal{D}|}{M}$$

Hence, the objective of frequent itemset mining is to identify the set of frequent itemsets $\mathcal{F}$ in $\mathcal{D}$, given a support threshold min_sup, where for each $F_i \in \mathcal{F}$ it holds that $freq(F_i, \mathcal{D}) \geq min\_sup$. Finally, considering an additional collection of transactions $\mathcal{C}$, acting as a *control group* to $\mathcal{D}$, it should hold that the two sets are independent, i.e., $\mathcal{C} \cap \mathcal{D} = \emptyset$. Using these sets, we define the *degree of itemset disproportionality* of an itemset I in $\mathcal{D}$ against $\mathcal{C}$ is defined as follows:

$$Idisp(I, \mathcal{D}, \mathcal{C}) = \frac{freq(I, \mathcal{D})}{freq(I, \mathcal{C})} \ . \tag{1}$$

3.2 Preliminary results

The experiments carried out so far[2] have used data extracted from a regional healthcare data warehouse GVR/VAL, which contains diagnoses (ICD-10), drugs (ATC), and other data related to consultations in primary and secondary care for more than 2 million inhabitants of the greater Stockholm area. All information is anonymous in order to preserve patient integrity. The population selected for this study consists of all individuals in the GVR/VAL data warehouse that during the time period 2010-2016 have been hospitalized for at least one night and been diagnosed with at least one heart failure related diagnosis; ICD-10: I50, I11.0, I42 (excluding I42.1 and I42.2), and I43. In total the data used in this study relates to 70,474 unique heart failure patients (9,446,322 events defined as assigned diagnosis codes). In total these patients are described by approximately 7,013 distinct diagnose codes. For this study, all diagnoses were represented by the first three characters of the ICD-10 code. In the experiments the patients are

[2] This research has been approved by the Regional Ethical Review Board in Stockholm (Dnr. 2016/479-31/5)

divided into two groups based on whether or not the patient has been prescribed the primary basic treatment for heart failure as defined by the national guidelines from Socialstyrelsen [18].

The results of the analysis highlight groups of patients that are more likely to receive basic treatment. By ranking the frequent itemsets according to their relative frequency between the groups, we get a better understanding of what types of patients are present in the respective groups. The most disproportional frequent itemsets where approximately 7 to 10 times more frequent among patients that received primary basic treatment than among those that did not. These itemsets consisted of combinations of 4 or 5 of the following 7 diagnoses: **E11** (type 2 diabetes mellitus), **I10** (essential (primary) hypertension), **I25** (chronic ischemic heart disease), **I48** (atrial fibrillation and flutter), **I50** (heart failure), **Z92** (personal history of medical treatment) and **Z95** (presence of cardiac and vascular implants and grafts) (Table 1). For more details the reader may refer to the original paper by Karlsson et al. [8].

Table 1: The top ranked itemsets, showing combinations of diagnoses that are more likely to indicate basic treatment. The itemsets are ranked based on their disproportionality score which shows how much more likely it is that a patient with this combination of diagnoses will receive basic treatment compared to not receiving it. A bullet indicates the presence of an item in the itemset.

Disp.	**E11**	**I10**	**I25**	**I48**	**I50**	**Z92**	**Z95**
9.675			•	•	•	•	•
9.410		•	•	•	•	•	
9.254			•		•	•	•
9.221		•	•		•	•	
8.220		•	•		•		•
7.945			•	•		•	•
7.707		•	•	•		•	
7.638	•	•		•	•	•	
7.376		•	•	•			•
7.356		•			•	•	•

4 Future outlook and challenges

Future endeavours into this domain will consist of a variety of challenges that are both data and clinically centered. Firstly, due to the nature clinical data reporting being inconsistent, the data sets in focus must be processed to account for: various missing values, measurements being recorded at once during the day, and various other types of errors which impact quality. Although current directions in healthcare involve moving towards more consistent and structured data sets, this is a prospect of the future, and data sets at hand must undergo

well-thought-out pre-processing procedures such as interpolation, extrapolation, replacement and deletion.

Secondly, a key challenge involves providing models that perform well enough on metrics such as AUC and recall, such that they are clinically reliable to the perception of clinicians. Indeed, the use of predictive models in clinical practice imposes a loss of autonomy on a clinician which can be perceived as a treat if the results of such models do not demonstrate a direct benefit and often deviate from a clinician's expectations. To combat a perceived threat to autonomy, more interpretable models can be chosen, such as decision trees, that are more descriptive in explaining predictions in real world logic. However, such descriptive models may not preform as well as black box models such as random forests which possess poor interpretability. This trade-off between medical interpretability and performance thus presents a challenge which can perhaps best be resolved through questioning medical professionals and experimenting with the compliance of systems in real world settings. On another note, such predictive models must demonstrate high performance in a clinical domain due the the high cost involved related to patient outcomes. Although false negatives may result in the overlooking of certain ADEs, such oversights are inevitable. In regards to legal and ethical accountability, final prescriptive decisions must be left to the discretion of medical experts to correct for such errors.

Finally, it is an ongoing challenge to deal with the ever changing health care system, which consistently encounters a high velocity of novel relevant data such as medications and treatments. Prescriptive models for use in clinical practice should ideally be constantly updated and built on real-time data as a means of early detection for a variety of medical situations such as ADEs. Although such real-time systems provide a clear means towards revolutionizing health care, various challenges are faced such as possessing data sets at a necessary quality, and with ensuring that updated models maintain the required validity for use in clinical practice.

5 Conclusions

We have presented an overview of recent and ongoing research on learning from EHRs and AHRs. Our review focused on Swedish health registries and concentrated around ADE understanding and detection, as well as modeling and understanding the effectiveness of treatment for heart failure patients. More concretely, we presented our recent work on detecting and understanding ADEs, with focus on predictive modeling techniques, methods for mining disproportional patterns, and finally we present the ADE explorer, a tool for studying ADEs from EHRs. in addition, we presented ongoing work on understanding the effectiveness of treatment of heart failure from AHRs, with emphasis on understanding the reasons behind the large variation of the usage rate of basic treatment by 50 hospitals in Stockholm County. Finally, we discussed directions for future research on our current projects involving EHRs and AHRs.

Acknowledgments

This work was partly supported by the VR-2016-03372 Swedish Research Council Starting Grant and by the Stockholm County Council.

References

1. Almenoff, J., Pattishall, E., Gibbs, T., DuMouchel, W., Evans, S., Yuen, N.: Novel statistical tools for monitoring the safety of marketed drugs. Clinical Pharmacology & Therapeutics 82(2), 157–166 (2007)
2. Asker, L., Boström, H., Karlsson, I., Papapetrou, P., Zhao, J.: Mining candidates for adverse drug interactions in electronic patient records. In: Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments, PETRA 2014, Island of Rhodes, Greece, May 27 - 30, 2014. pp. 22:1–22:4 (2014), http://doi.acm.org/10.1145/2674396.2674420
3. Beijer, H., De Blaey, C.: Hospitalisations caused by adverse drug reactions (ADR): a meta-analysis of observational studies. Pharmacy World and Science 24(2), 46–54 (2002)
4. Dalianis, H., Hassel, M., Henriksson, A., Skeppstedt, M.: Stockholm epr corpus: A clinical database used to improve health care. In: Proceedings of the Fourth Swedish Language Technology Conference (2009)
5. Hazell, L., Shakir, S.A.: Under-reporting of adverse drug reactions. Drug Safety 29(5), 385–396 (2006)
6. Henriksson, A., Zhao, J., Boström, H., Dalianis, H.: Modeling electronic health records in ensembles of semantic spaces for adverse drug event detection. In: 2015 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2015, Washington, DC, USA, November 9-12, 2015. pp. 343–350 (2015), https://doi.org/10.1109/BIBM.2015.7359705
7. Howard, R., Avery, A., Slavenburg, S., Royal, S., Pipe, G., Lucassen, P., Pirmohamed, M.: Which drugs cause preventable admissions to hospital? A systematic review. British journal of clinical pharmacology 63(2), 136–147 (2007)
8. Karlsson, I., Papapetrou, P., Asker, L., Boström, H., Persson, H.E.: Mining disproportional itemsets for characterizing groups of heart failure patients from administrative health records. In: Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments, PETRA 2017, Island of Rhodes, Greece, June 21-23, 2017. pp. 394–398 (2017), http://doi.acm.org/10.1145/3056540.3076177
9. Karlsson, I., Zhao, J., Asker, L., Boström, H.: Predicting adverse drug events by analyzing electronic patient records. In: Proceedings of the Conference on Artificial Intelligence in Medicine in Europe. pp. 125–129. Springer (2013)
10. Liu, X., Chen, H.: Azdrugminer: an information extraction system for mining patient-reported adverse drug events in online patient forums. In: Proceedings of the International Conference on Smart Health. pp. 134–150. Springer (2013)
11. Meystre, S.M., Savova, G.K., Kipper-Schuler, K.C., Hurdle, J.F.: Extracting information from textual documents in the electronic health record: a review of recent research. Yearbook of medical informatics pp. 128–44 (Jan 2008), http://www.ncbi.nlm.nih.gov/pubmed/18660887
12. Nebeker, J.R., Barach, P., Samore, M.H.: Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. Annals of internal medicine 140(10), 795–801 (2004)

13. Pariente, A., Gregoire, F., Fourrier-Reglat, A., Haramburu, F., Moore, N.: Impact of safety alerts on measures of disproportionality in spontaneous reporting databases the notoriety bias. Drug safety 30(10), 891–898 (2007)
14. van Puijenbroek, E.P., Bate, A., Leufkens, H.G., Lindquist, M., Orre, R., Egberts, A.C.: A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiology and drug safety 11(1), 3–10 (2002)
15. Schneeweiss, S., Hasford, J., Göttler, M., Hoffmann, A., Riethling, A.K., Avorn, J.: Admissions caused by adverse drug events to internal medicine and emergency departments in hospitals: a longitudinal population-based study. European journal of clinical pharmacology 58(4), 285–291 (2002)
16. Socialstyrelsen: Nationella riktlinjer – Målnivåer, Hjärtsjuk-vård (2015), artikelnummer 2015-10-3
17. Socialstyrelsen: Nationella riktlinjer – Utvärdering, Hjärtsjukvård (2015), `http://www.socialstyrelsen.se`
18. Socialstyrelsen: Nationella riktlinjer för hjärtsjukvård (2015), `http://www.socialstyrelsen.se`
19. Suzuki, A., Andrade, R.J., Bjornsson, E., Lucena, M.I., Lee, W.M., Yuen, N.A., Hunt, C.M., Freston, J.W.: Drugs associated with hepatotoxicity and their reporting frequency of liver adverse events in VigiBase^{TM}. Drug safety 33(6), 503–522 (2010)
20. Velupillai, S., Skeppstedt, M., Kvist, M., Mowery, D.L., Chapman, B.E., Dalianis, H., Chapman, W.W.: Cue-based assertion classification for swedish clinical text - developing a lexicon for pycontextswe. Artificial Intelligence in Medicine 61(3), 137–144 (2014), `https://doi.org/10.1016/j.artmed.2014.01.001`
21. Wester, K., Jönsson, A.K., Spigset, O., Druid, H., Hägg, S.: Incidence of fatal adverse drug reactions: a population based study. British journal of clinical pharmacology 65(4), 573–579 (2008)
22. Zhao, J., Henriksson, A., Asker, L., Boström, H.: Detecting adverse drug events with multiple representations of clinical measurements. In: 2014 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2014, Belfast, United Kingdom, November 2-5, 2014. pp. 536–543 (2014), `https://doi.org/10.1109/BIBM.2014.6999216`