Exploring Complexity in Health: An Interdisciplinary Systems Approach
A. Hoerbst et al. (Eds.)

doi:10.3233/978-1-61499-678-1-247

Making Medication Data Meaningful: Illustrated with Hypertension

Richard WILLIAMS[a,b,1], Benjamin BROWN[a], Niels PEEK[a,b] and Iain BUCHAN[a,b]
[a]*MRC Health eResearch Centre*
[b]*NIHR Greater Manchester Primary Care Patient Safety Translational Research Centre, University of Manchester, Manchester, UK*

Abstract. We demonstrate, with application to hypertension management, an algorithm for reconstructing therapeutic decisions from electronic primary care medication prescribing records. These decisions concern the initiation, termination and alteration of therapy, and have further utility in: monitoring patient adherence to medication; care pathway analysis including process mining; advanced phenotype construction; audit and feedback; and in measuring care quality.

Keywords. Pharmacoepidemiology; phenotype extraction; drug prescriptions; electronic prescribing; care pathway.

1. Introduction

Electronic health records (EHRs) typically contain coded data about prescriptions, which, in the UK, are readily available for research from anonymised collections of primary care records [1, 2]. These data usually describe orders to dispense medication, but do not explicitly record the therapeutic decisions of when treatment is commenced, changed or terminated. These events, which are more clinically meaningful than individual prescriptions, can be used in a variety of analyses and tools: monitoring patient adherence to medication; care pathway analysis including process mining; next-generation phenotyping[3]; realistically-complex quality indicators; and advanced audit. However the raw EHR data must first be processed with consideration given to: the clinical codes; the drug family and active ingredients; and the dose amount, frequency and duration of all repeat prescriptions in a patient's history. This pre-processing is often done as part of research but is usually simplified or tailored to the analysis [4–8].

A common approach for inferring drug usage is to count the number of prescriptions within the study period and to only include patients who exceed a certain threshold. Sometimes this is a single prescription [4], but more often this is two or more prescriptions [5–7], presumably to ensure a certain degree of continued usage and to exclude one off prescriptions. Another method for determining continued drug usage is to look for prescriptions in adjacent time windows [8]. These approaches are valid if the dosage or presence of a drug is considered as a single covariate, however, to contextualize prescribing to the clinical decisions made on care pathways it is necessary to convert the EHR data into decision events such as when therapy is commenced, changed or terminated.

[1] Corresponding Author: richard.williams2@manchester.ac.uk

Similar research has been performed in extracting medication information from free text in discharge summaries [9, 10], from physician notes [11] and from drug purchase databases [12]. However, our approach relies on well coded prescription events, such as those available from the UK primary care system where all prescribing is electronic.

Tanskanen et al. [12, 13] construct drug use periods defined by a start and end date with an algorithm similar to the one described here. Our method goes further by detecting not only when therapy is commenced and terminated, but also when the dosage is changed leading to the extraction of a greater number of clinically meaningful events on a patient's individual care pathway.

Here we focus on medications prescribed for hypertension.

2. Method

We used an anonymized extract of primary care data from 53 general practices in Salford, UK (population 234k) from the Salford Integrated Record (SIR). SIR collects primary care data for research purposes, with consent on an opt-out basis, from the general practices in Salford. The data consist of Read codes (version 2) and EHR vendor-specific codes. SIR contains coded data for all prescribing in primary care for its population for at least the past fifteen years.

All prescriptions of drugs recommended by the National Institute for Health and Care Excellence (NICE) for the treatment of hypertension [14] in the UK were extracted. These are angiotensin-converting enzyme inhibitors (ACEIs); angiotensin-II receptor blockers (ARBs); alpha-adrenoceptor blockers (α-blockers); beta-adrenoceptor blockers (β-blockers); calcium channel blockers (CCBs); thiazides and related diuretics; spironolactone and other diuretics. We extracted the clinical code, the number of tablets prescribed and the patient instructions (e.g. "Take 2 once a day").

We created a mapping between each drug code, the active ingredient(s) and the tablet dose (mg). We found 653 Read codes and 199 vendor-specific codes, covering 178 brand names and 70 generic names of antihypertensive medications. This mapping information is publically available from the UK Health and Social Care Information Centre (HSCIC) for Read codes, however for the vendor-specific codes this was done manually based on the text description associated with each code.

Text mining, using regular expressions, on 216,101 distinct textual patient instructions yielded the number of tablets taken per day. We then iteratively developed an algorithm to take the amount, frequency, duration and type of medication, together with the prescription date to extrapolate meaningful events as shown in Table 1 and Table 2. The iterations continued until the proportion of unclassified events achieved an acceptably low level. The algorithm was then validated by two authors (RW, a software engineer, and BB, a clinician) who independently reviewed the records of a random sample of 100 patients to determine if the correct sequence of events had been extracted, and if not, recorded the discrepancy. Cohen's κ showed fair inter-rater agreement (0.45). Disagreements were resolved through discussion.

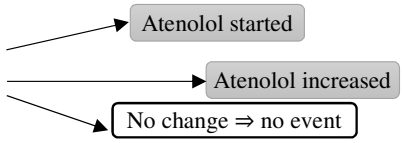
The full algorithm including: code lists, mapping files, and regular expressions for converting patient instructions to tablets per day are all available online at https://github.com/rw251/research-events-medication-htn.

Table 1. The extracted medication events and the reasoning for each one

Medication event	Reasoning
Start	Active ingredient first prescribed
Restart	Active ingredient previously prescribed but last event produced by the algorithm was a stop
Stop	Active ingredient previously prescribed but time has elapsed without a repeat prescription
Dose increase	Dose per day = (tablets per day) * (mg per tablet) increases
Dose decrease	Dose per day = (tablets per day) * (mg per tablet) decreases

Table 2. Example of a conversion from EHR to meaningful clinical events

Date	EHR text	Instruction	Tabs
2015-02-15	Atenolol 25mg capsules	1 in the am	28
2015-03-16	Atenolol 25mg capsules	Take 2 daily	28
2015-04-01	Atenolol 50mg capsules	One each day	28



3. Results

The algorithm was developed over six iterations involving the detailed examination of 179 patient records. A total of 10,311,973 prescriptions were extracted for 81,096 patients (demographic information in Table 3) over the period 7 July 1977 to 12 December 2014. The breakdown for each family of drugs is shown in Table 4. The algorithm produced sequences with a combined total of 850,028 events (28% starts, 34% stops, 15% restarts, 16% increases, 8% decreases and 0.02% unclassified).

Table 3. Patient characteristics for the 81,096 patients extracted from the dataset. Values are n (%) unless otherwise specified.

Demographic	Value
Age	
Mean (SD)	62.43 (18.69)
Sex	
Female	43419 (53.5%)
Male	37646 (46.4%)
Unknown	31 (0.04%)
Ethnicity	
White	46712 (57.6%)
Other	4264 (5.26%)
Unknown	30120 (37.1%)
Deprivation (Index of Multiple Deprivation [15] quintiles)	
1st (Most deprived)	37850 (46.7%)
2nd	19349 (23.9%)
3rd	12262 (15.1%)
4th	6851 (8.45%)
5th (Least deprived)	3813 (4.70%)
Unknown	971 (1.20%)

Table 4. The number of antihypertensive drugs prescribed

Drug family	Distinct drug types per family	Number of prescriptions
Angiotensin-converting enzyme inhibitors	10	2,295,190
Angiotensin-II receptor blockers	7	841,217
Beta blockers	15	2,074,013
Calcium-channel blockers	11	2,021,822
Alpha blockers	8	543,226
Thiazide-like diuretic	11	1,278,005
Other diuretics	8	1,258,500

During validation the algorithm achieved a PPV of 92% (95% CI 85%-96%). Of the 100 records reviewed only eight had incorrect sequences and these were still partially correct. Four were missing a single stop event, two were false increases due to an erroneous prescription, one had an extra stop event, and one had a decrease for a switch from 2.5mg indapamide to 1.5mg modified release; these are actually clinically identical.

4. Discussion

We have developed a method for transforming unstructured and semi-structured prescription data into clinically meaningful therapeutic decisions on a care pathway from EHR data. From these events it is then easy to determine: when a patient is taking a medication; when there are adherence issues; when an intervention is or isn't made by a physician; when guidelines are being followed correctly; and if treatment is having the desired effect. The events can also be used as part of a phenotype extraction process where the presence of a particular medication is indicative of a specific condition or diagnosis. Furthermore the algorithm has the potential to be improved by addressing the occasions where an incorrect sequence was produced.

The method described here addresses a specific aspect of the conversion of data from a primary care database into a form ready for further research and analysis. It can be viewed as a single module within a wider framework, where other modules might address problems such as constructing clinical code lists, imputing missing data or inferring diagnoses from other relevant information. This modular paradigm has many advantages over the monolithic approach: individual modules can be reused for multiple purposes; module development is more manageable and can be distributed across different research or analytic centres; and modules can be easily swapped for alternate or improved versions.

Future Work: We plan to incorporate these medication events into care pathways analyses to discover how pathway variation affects patient outcomes in managing long-term conditions. *Strengths*: The algorithm being developed by a software engineer and a clinician; both having extensive experience of primary care prescribing and primary care datasets; the iterative development of the algorithm; and the high PPV. *Limitations*: Development and validation were performed by the same authors; the algorithm is pharmacologically ignorant; we do not have information as to whether prescriptions were collected or taken. However, the act of prescribing is close to the clinician-centered decision context of care pathways.

Conclusion: Extracting research-quality information from routinely collected datasets is hard, time consuming, and prone to errors and bias. Accelerating research in these areas requires reproducible methods and tools that are open to scrutiny and easily reused and extended – all code for this project is available on github.com.

Acknowledgements

Funded by the National Institute for Health Research Greater Manchester Primary Care Patient Safety Translational Research Centre (NIHR GM PSTRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

[1] Walley T, Mantgani A., "The UK General Practice Research Database.," *Lancet* (London, England), vol. **350**, no. 9084, pp. 1097–9, Oct. 1997.

[2] Akbarov A, Kontopantelis E, Sperrin M, et al., "Primary Care Medication Safety Surveillance with Integrated Primary and Secondary Care Electronic Health Records: A Cross-Sectional Study.," *Drug Saf.,* vol. **38**, no. 7, pp. 671–82, 2015.

[3] Hripcsak G, Albers DJ., "Next-generation phenotyping of electronic health records," *J. Am. Med. Informatics Assoc*., pp. 117–121, 2012.

[4] Macedo AF, Douglas I, Smeeth L, Forbes H, Ebrahim S., "Statins and the risk of type 2 diabetes mellitus: cohort study using the UK clinical practice research datalink.," *BMC Cardiovasc. Disord*., vol. **14**, p. 85, Jan. 2014.

[5] McGlynn KA, Hagberg K, Chen J, et al., "Statin use and risk of primary liver cancer in the clinical practice research datalink.*," J. Natl. Cancer Inst.,* vol. **107**, no. 4, p. djv009–, Apr. 2015.

[6] George J, Majeed W, Mackenzie IS, Macdonald TM, Wei L., "Association between cardiovascular events and sodium-containing effervescent, dispersible, and soluble drugs: nested case-control study.," *BMJ*, vol. **347**, no. nov26_3, p. f6954, Jan. 2013.

[7] Hagberg KW, Sahasrabuddhe V V, McGlynn KA, Jick SS., "Does Angiotensin-Converting Enzyme Inhibitor and β-Blocker Use Reduce the Risk of Primary Liver Cancer? A Case-Control Study Using the UK Clinical Practice Research Datalink," Pharmacother. J. *Hum. Pharmacol. Drug Ther.,* vol. **36**, no. 2, pp. 187–195, 2016.

[8] Avery AJ, Rodgers S, Cantrill JA, et al., "A pharmacist-led information technology intervention for medication errors (PINCER): A multicentre, cluster randomised, controlled trial and cost-eff ectiveness analysis," *Lancet*, vol. **379**, no. 9823, pp. 1310–1319, 2012.

[9] Gold S, Elhadad N, Zhu X, Cimino JJ, Hripcsak G., "Extracting structured medication event information from discharge summaries*.," AMIA Annu. Symp. Proc*., pp. 237–41, Jan. 2008.

[10] Uzuner O, Solti I, Cadag E., "Extracting medication information from clinical text.," *J. Am. Med. Inform. Assoc.,* vol. **17**, no. 5, pp. 514–8, Jan. 2010.

[11] Turchin A, Kolatkar NS, Grant RW, Makhni EC, Pendergrass ML, Einbinder JS., "Using regular expressions to abstract blood pressure and treatment intensification information from the text of physician notes.," *J. Am. Med. Inform. Assoc.,* vol. **13**, no. 6, pp. 691–5, Jan. 2006.

[12] Tanskanen A, Taipale H, Koponen M, et al., "From prescription drug purchases to drug use periods – a second generation method (PRE2DUP*)," BMC Med. Inform. Decis. Mak*., vol. **15**, no. 1, pp. 1–13, 2015.

[13] Tanskanen A, Taipale H, Koponen M, et al., "From prescriptions to drug use periods - things to notice.," *BMC Res. Notes*, vol. **7**, no. 1, p. 796, 2014.

[14] Nice., "Hypertension - Clinical management of primary hypertension in adults," *NICE Clin. Guidel.* **127**, no. April 2007, 2011.

[15] DCLG., "The English Indices of Deprivation 2010," *Neighb. Stat. Release*, pp. 1–20, 2011.