
Research and Applications

Comparison of the cohort selection performance of Australian Medicines Terminology to Anatomical Therapeutic Chemical mappings

Guan N Guo ,^{1,2} Jitendra Jonnagaddala ,^{1,2} Sanjay Farshid ,¹ Vojtech Huser,³ Christian Reich,⁴ and Siaw-Teng Liaw^{1,2}

¹School of Public Health and Community Medicine, University of New South Wales, Sydney, Australia, ²WHO Collaborating Centre for eHealth, University of New South Wales, Sydney, Australia, ³Lister Hill National Centre for Biomedical Communications, National Library of Medicine National Institutes of Health, Bethesda, Maryland, USA, and ⁴IQVIA, Cambridge, Massachusetts, USA

Corresponding Author: Jitendra Jonnagaddala, School of Public Health and Community Medicine, University of New South Wales, Sydney, Australia (z3339253@unsw.edu.au)

Received 11 January 2019; Revised 10 June 2019; Editorial Decision 16 July 2019; Accepted 22 July 2019

ABSTRACT

Objective: Electronic health records are increasingly utilized for observational and clinical research. Identification of cohorts using electronic health records is an important step in this process. Previous studies largely focused on the methods of cohort selection, but there is little evidence on the impact of underlying vocabularies and mappings using vocabularies used for cohort selection. We aim to compare the cohort selection performance using Australian Medicines Terminology to Anatomical Therapeutic Chemical (ATC) mappings from 2 different sources. These mappings were taken from the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) and the Pharmaceutical Benefits Scheme (PBS) schedule.

Materials and Methods: We retrieved patients from the electronic Practice Based Research Network data repository using 3 ATC classification groups (A10, N02A, N06A). The retrieved patients were further verified manually and pooled to form a reference standard which was used to assess the accuracy of mappings using precision, recall, and F measure metrics.

Results: The OMOP-CDM mappings identified 2.6%, 15.2%, and 24.4% more drugs than the PBS mappings in the A10, N02A and N06A groups respectively. Despite this, the PBS mappings generally performed the same in cohort selection as OMOP-CDM mappings except for the N02A Opioids group, where a significantly greater number of patients were retrieved. Both mappings exhibited variable recall, but perfect precision, with all drugs found to be correctly identified.

Conclusion: We found that 1 of the 3 ATC groups had a significant difference and this affected cohort selection performance. Our findings highlighted that underlying terminology mappings can greatly impact cohort selection accuracy. Clinical researchers should carefully evaluate vocabulary mapping sources including methodologies used to develop those mappings.

Key words: Australian Medicines Terminology, anatomical therapeutic chemical classification, electronic health records, cohort selection, clinical trials

INTRODUCTION

Routinely collected data from electronic health records (EHR) are increasingly being utilized for clinical decision support systems and observational research.^{1–3} Such data allow large-scale observational research to complement clinical research.^{4,5} The data also enable researchers to identify specific patient cohorts for observational, genomic,⁶ and clinical research.^{7–12} However, using electronic health records (EHR) data for cohort selection can be time-consuming and challenging depending on the complexity of the selection criteria, data quality, and mapping accuracy.^{13–15}

Classification terminologies, such as the Anatomical Therapeutic Chemical (ATC) classification, can efficiently be used to identify patients for clinical research,¹⁶ monitor antibiotics and antibiotic resistance worldwide,¹⁷ and identify drug-related problems in poly-pharmacy studies.¹⁸ The use of the ATC in data repositories requires the local drug concepts from various vocabularies to be mapped to the classification classes of ATC. Creating mappings between them allows the classification to be used consistently throughout and across multiple data repositories. Health informaticians around the world have been actively mapping between various vocabularies¹⁹ and creating rich resources, such as the Unified Medical Language System (UMLS) metathesaurus.²⁰

The ATC is also used in cohort selection and electronic phenotyping, usually for clinical trials.^{21–25} Electronic phenotyping identifies patients with specific characteristics by querying EHR systems and EHR-based data repositories using specific algorithms based on rules and may include machine learning. The 5-level hierarchical structure of ATC allows selection of cohorts with a degree of control in the granularity. Huber et al²⁵ demonstrated that the ATC can be used to identify patients with chronic conditions and measure their disease status.

Previous studies largely focused on approaches and models of cohort selection,¹¹ such as the use of machine learning²⁶ and natural language processing.²⁷ However, there is little research that shows how the underlying terminology or classification system affects the accuracy and effectiveness of cohort selection. Using the ATC for cohort selection relies on accurate mappings from source drug terminologies to ATC.²⁸ Reich et al¹⁴ demonstrated that differences in vocabulary mapping can significantly impact the outcome of cohort selection. There is also substantial variation between different classification systems depending on the context and clinical issues being investigated such as opioid exposure.²⁹ It is therefore important to investigate the limitations and accuracy of mappings between terminologies and classifications.^{30–32} Fung et al³² compared the performance of 2 mapping methods, semantic and lexical, used to map Systematized Nomenclature of Medicine (SNOMED) concepts to International Classification of Diseases, ninth revision, clinical modification (ICD9CM) concepts. They found that the 2 methods had their own strengths and weaknesses, and combining the methods achieved higher performance than either 1 alone. Furthermore, Hripcsak et al¹³ found that using knowledge engineering can significantly reduce the error rate in cohort selection. Knowledge engineering refers to application of computational rules to imitate expert knowledge. Specifically, authors devised automatic and semiautomatic methods based on expert knowledge to map concepts between 2 different vocabularies. It is important that mappings have a high precision (positive predictive value) rather than higher sensitivity,³³ as precision can greatly impact cohort selection and, ultimately, the outcomes of clinical trials.³⁴

In this study, we investigated the effect of mappings on the accuracy of cohort selection. We took a similar approach to the previous studies and compared the performance of the Australian Medicines

Terminology (AMT) to ATC mappings originating from 2 different sources: The Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) and the Pharmaceutical Benefits Scheme (PBS) schedule mapping by the Australian government.

BACKGROUND

Clinical terminologies

A clinical terminology is a collection of terms which describes a set of concepts and the relationships among them. In this study we have used the terms vocabulary and terminology interchangeably. Terminologies can be divided into 3 main types³⁵: Interface Terminology is the clinical language used for a particular domain; Reference Terminology is a context-free description of concepts and a common reference point from multiple different terminology systems; and Aggregating Terminology provides a systematic arrangement of hierarchical and disjoint classes based on common characteristics—often referred to as classifications.³⁶ One example of reference terminology is the Systematized Nomenclature of Medicine Clinical Terminology (SNOMED CT) which covers a wide range of domains to comprehensively represent routine health care. SNOMED CT is also used as a clinical terminology for digital health systems such as EHRs.^{37,38} Aggregating terminologies are classification systems such as ATC³⁶ and International Classification of Diseases (ICD),³⁹ both of which categorize information into groups based on their properties. Classification terminologies do not aim to describe all possible concepts, but rather they allow the easy storage, retrieval, and analysis of sets of concepts for research and decision-making.⁴⁰ These are effective tools for identifying and aggregating information for both public health and clinical research. Hence, it is vital to have accurate and consistent mappings.

Australian Medicines Terminology

The AMT is the Australian national standard for the identification and naming of medicines in Australia.⁴¹ It is managed by the Australian Digital Health Agency and is structured to have 7 classes: Medicinal Product (MP), Medicinal Product Pack (MPP), Medicinal Product Unit of Use, Trade Product (TP), Trade Product Pack (TPP), Trade Product Unit of Use, Containerized Trade Product Pack. Initially developed as a stand-alone medicines terminology modelled after SNOMED CT,⁴² the AMT is now part of the SNOMED CT-AU in the form of an extension. As a result, there is significant duplication between MP and Substance concepts in AMT and SNOMED CT (Supplementary Appendix A). For example, an MP from AMT *bupropion*, 21798011000036100 has the SNOMED CT International equivalent of *Bupropion-containing product*, 96199001. Most of the AMT substances are identical to SNOMED substances, with exceptions including vaccines and food supplements where AMT terms are more specific.

Anatomical Therapeutic Chemical classification

The ATC classification was developed to standardize research to improve the quality use of medicines; it is maintained by the World Health Organization Collaborating Centre (WHOCC) for Drug Statistics Methodology.¹⁶ The WHO ATC was introduced in 1976 and is updated annually to include new ingredients approved for use in the previous year. The classification has a 5-level structure, dividing active ingredients into groups according to the organ or system they act on at the first level; therapeutic subgroup at the second level; chemical (therapeutic, pharmacological) properties at the third

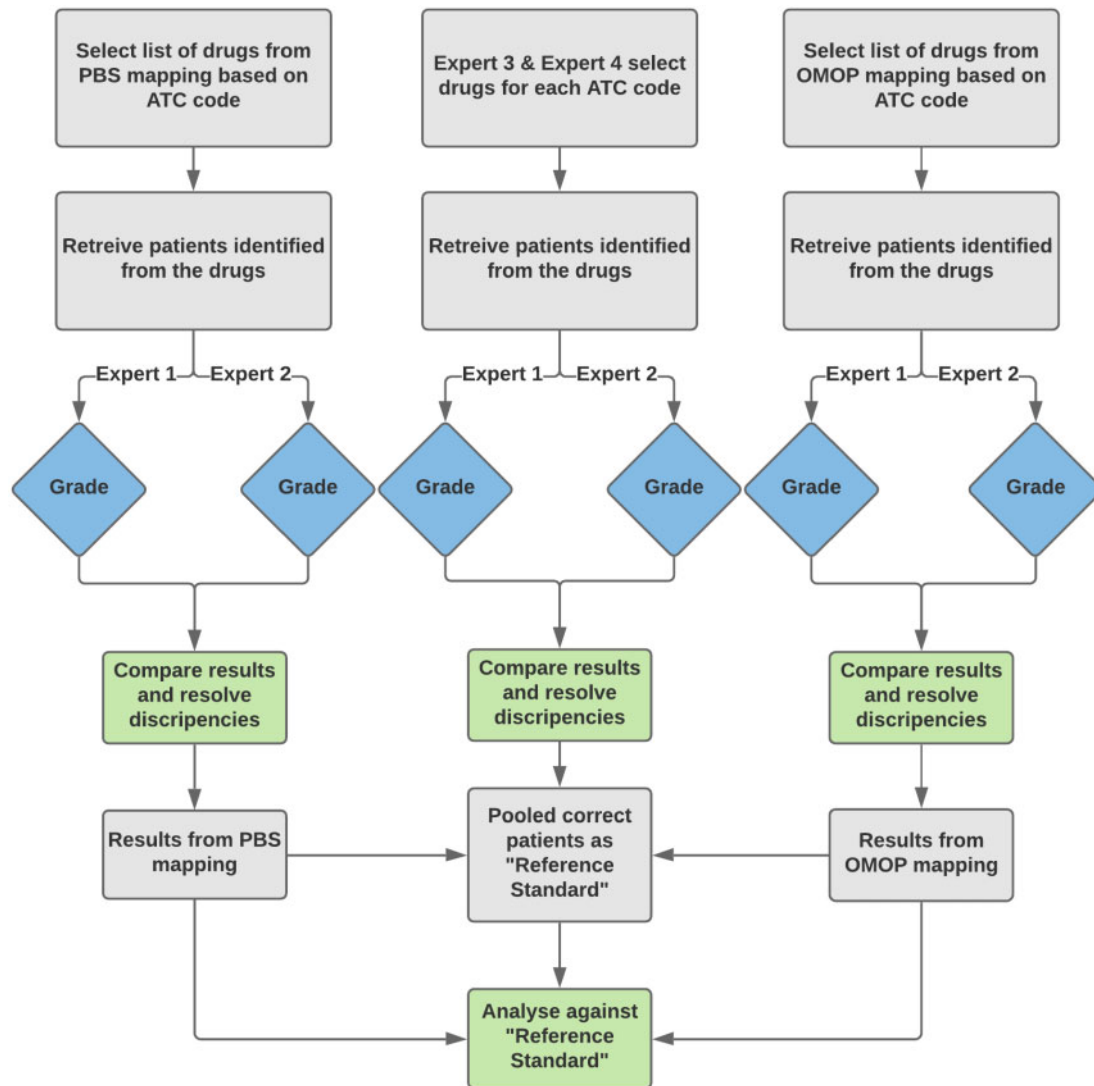


Figure 1. Overview of the study design.

and fourth levels; then chemical substance or combinations at the fifth level. Usually only 1 ATC code is assigned to ingredients of similar strength, route of administration, or indication. However, ingredients with different strength or route of administration may be assigned more than 1 ATC code.⁴³ For example, *dapsone* is categorized under 2 different categories: *D10AX (Other anti-acne preparation for topical use)* and *J04BA (Drugs for treatment of lepra)* to reflect its varying strength and routes of administration. In case of an ingredient of the same strength and route with 2 or more indications, only 1 ATC code is generally allocated by the WHO. For example, *amitriptyline* is commonly indicated for treating depression, neuropathic pain, and migraines. However, it only has 1 ATC code (*N06AA09*) reflecting its indication for depression.

MATERIALS AND METHODS

An overview of our methods is shown in Figure 1. Three ATC classification groups at the second and third levels were chosen for this study: *A10 Drugs used in Diabetes*, *N02A Opioids*, and *N06A Antidepressants*. We first retrieved the AMT concepts based on their

ATC code using OMOP-CDM, PBS, and a set of expert generated mappings. We then identified the patients that have taken these drugs. The patients were reviewed manually by another 2 experts to verify the correctness of the patients retrieved. To create the reference standard, we adapted a method described by Hripcsak and Wilcox⁴⁴ in which experts serve multiple roles. Similar methods have been used in studies evaluating the use of natural language processing for detection of clinical conditions from clinical reports, where there is no existing reference standard with which to compare the evaluation results.⁴⁵ We created the reference standard by pooling the verified patients from the 3 mappings and removing duplicates, leaving only the correctly identified patients of each ATC class. The reference standard was then used to compare the performances of the OMOP and PBS mappings. We have shared the code developed as part of this study publicly.⁴⁶

AMT-ATC mappings from PBS

The PBS is a prescription drug subsidy program in Australia. The PBS schedule lists all the medicines available to patients at government subsidized prices. The PBS listing contains AMT identifiers

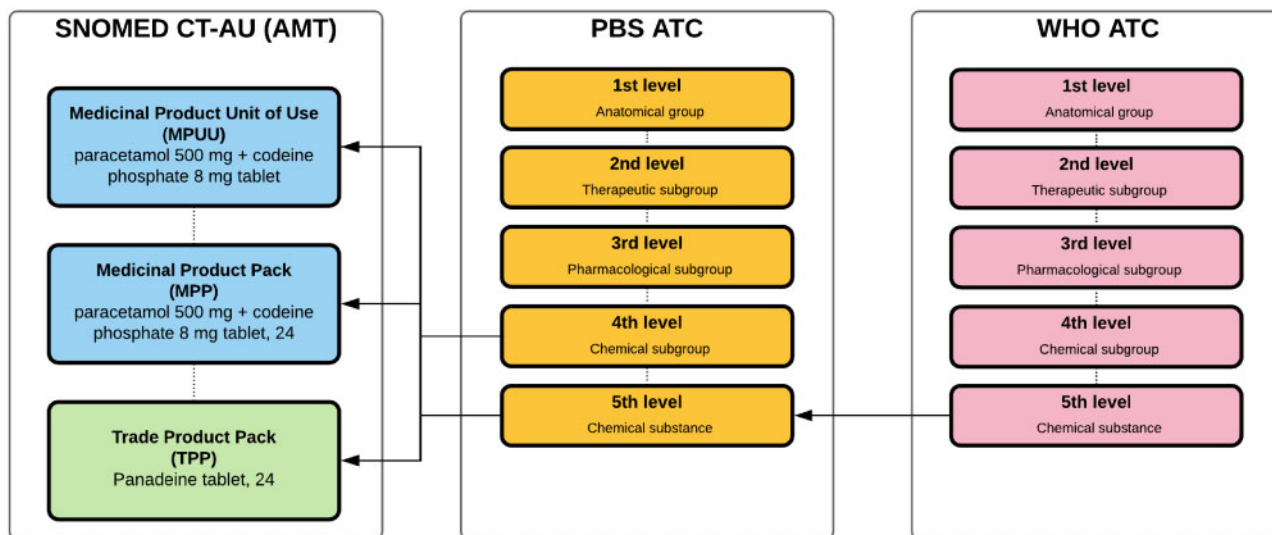


Figure 2. PBS mapping of ATC code to AMT concepts. The arrows indicate where the ATC code assignment originated from and where they were assigned to. fifth-level ATC codes assigned to the AMT concepts were carried over from the WHO ATC, while fourth-level codes were assigned by PBS.

and descriptions of Medicinal Product Unit of Use and Trade Product Unit of Use for chemotherapy infusible items; and MP, MPP, and TPP for all other items. The schedule is updated at the end of each month and new drugs are incorporated into the AMT the following month. The WHO ATC is updated annually in January with the PBS updated to incorporate these new or amended ATC codes by April of each year.

Appropriate ATC codes are assigned to PBS listings with most of the fifth-level codes carried over from WHO ATC. On occasions where a fifth-level code is not yet available, an appropriate fourth-level code (or, more rarely, a third-level code) will be assigned by the PBS (Figure 2). Information from various sources including the WHOCC website (a code may have been proposed but not yet ratified), the Therapeutic Goods Administration approved product information and indications for use, and the indications for which the drug is to be PBS-listed are used in this process. As a result, ATC codes included as part of PBS data may be different from those assigned by the WHOCC. Discrepancies can occur when a single drug has multiple indications. An example of this is dapsone, whose ATC code (*J04BA02*) assigned by the WHO for the oral preparation relates to its use in the treatment of leprosy. The department was made aware that medical practitioners were misinterpreting the heading for dapsone in the PBS schedule to mean that it was only listed for this condition. As a result, a second ATC code, *D11AX05*, was assigned to reflect its additional indication.

PBS mappings used in this study were from the PBS publication released July 1, 2018.⁴⁷ The *amt_20180701.txt* and *drug_20180701.txt* files containing the drug information were used to source the PBS mappings. In the source file, each row represents a single drug concept, with the *amt_20180701.txt* containing codes at the MP, MPP, and TPP levels, and *drug_20180701.txt* containing the ATC classification of each drug. The 2 files were joined by matching PBS code, MP (active ingredient), and MPP to construct a single table with AMT to ATC mappings. For example, the output consisted of drug name *dapagliflozin 10 mg tablet, 28*, ATC code *A10BK01*, PBS code *10011X* extracted from *drug_20180701.txt*; and AMT MP *91851000036102*, MPP *89811000036101*, and TPP *89801000036103* from the *amt_20180701.txt*.

AMT-ATC mappings from OMOP-CDM

OMOP-CDM aims to solve the problem of disparity between systems when sharing clinical data.⁴⁸ It allows replicable systematic analyses across databases by harmonizing data from different sources into a standardized common data model and vocabulary. The drug vocabulary structure is based on RxNorm, which also forms the core of the drug content of OMOP-CDM. RxNorm comprehensively describes the drug market in the United States of America. It may not contain products available in other countries. Drug concepts that do not exist in RxNorm are added as RxNorm Extension in OMOP-CDM, using an OMOP-CDM generated vocabulary with the same structure and properties as RxNorm (Figure 3). Since August 2013, ATC has been a source vocabulary in RxNorm,⁴⁹ meaning RxNorm ingredients would have existing mappings with fifth-level ATC equivalents.

OMOP-CDM used the UMLS Metathesaurus to source mappings for a range of terminologies, including ATC and RxNorm. However, the Metathesaurus does not include the SNOMED CT-AU extension and AMT vocabularies. The AMT mapping to ATC was therefore completed by Observational Health Data Sciences and Informatics (OHDSI) program. AMT was added into the OMOP-CDM vocabulary on September 6, 2017.⁵⁰ New drug vocabularies such as AMT were added with an OHDSI-developed script which breaks down each new drug concept into its attributes and matches them against existing concepts. New concepts in RxNorm extension are created for those that do not match. OHDSI is experimenting with knowledge-engineering vendors to import and map new vocabularies using available sources with automated mapping and also manually translate the concepts. A vocabulary working group works through these new concepts and makes necessary changes.

This study used OMOP-CDM v5.3 and vocabulary version v5.0 18-JAN-18. We first retrieved all mappings for the AMT to ATC from the OMOP-CDM vocabularies, obtaining a total of 56 889 AMT concepts with their corresponding ATC codes. We refer to these mappings as “OMOP-CDM mappings.” Examples of the final output of the mapping retrieved is shown in [Supplementary Appendix A](#). An example OMOP-CDM concept for *Dapagliflozin 10 Mg Tablet* is available publicly.⁵¹

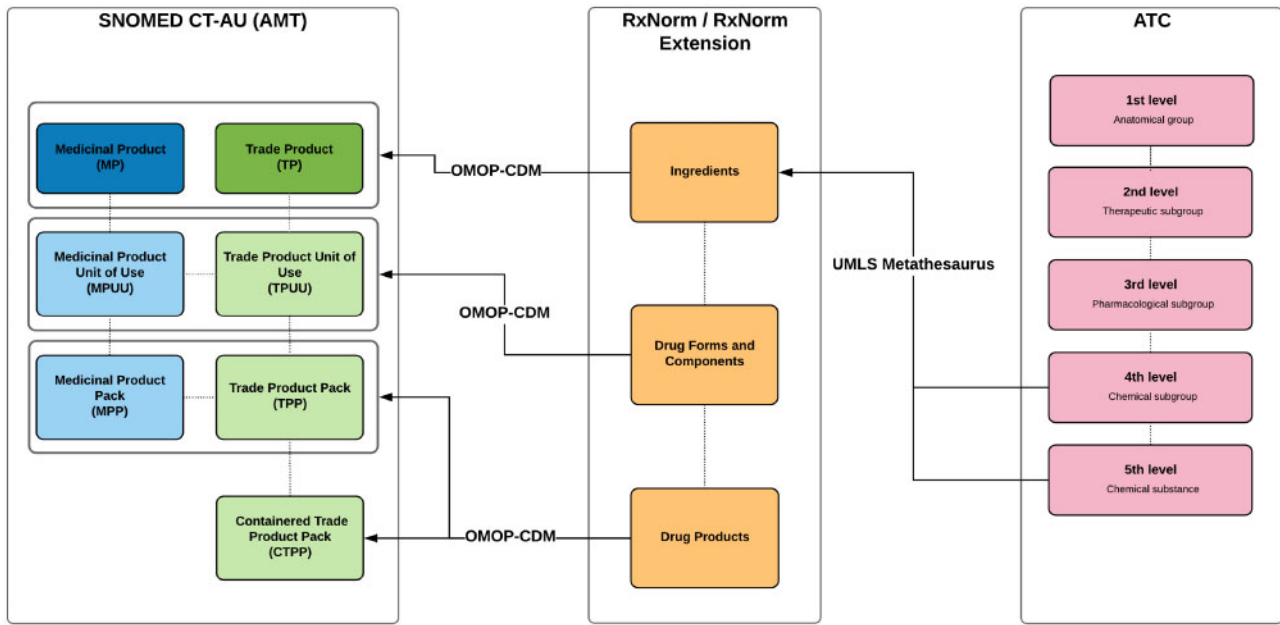


Figure 3. OMOP-CDM mapping of ATC code to AMT concepts. The OMOP-CDM mappings between ATC and AMT concepts were based on RxNorm/RxNorm extension. The ATC codes were mapped to the ingredients level of RxNorm. By mapping the components of RxNorm/RxNorm extension to the AMT equivalent concepts, the internal hierarchical structure of RxNorm/RxNorm extension is able to assign ATC codes to its descendent concepts and pass on the ATC codes to the AMT concepts it maps to.

Cohort selection

Cohort selection was performed on the ePBRN data repository at the University of New South Wales in Sydney, Australia.⁵² The data is extracted every 6 months from 18 general practices and community and hospital services, including outpatient clinics, in health neighborhoods in South Western Sydney. The data are pseudonymized, extracted, and linked in a secure manner^{53,54} and includes patient demographics, medications, conditions, and visits to general practices and hospitals converted into OMOP-CDM.⁵⁵ The linked data are used for analysis and investigation for various health research purposes. The ePBRN data repository has been used in various studies providing an opportunity to address any data quality issues encountered during the execution of studies.^{52,56} We selected patients who have attended the practice at least 3 or more times in the past 2 years from the ePBRN (November 2017 data extract). This was based on the definition of an “active patient” set out by the Royal Australian College of General Practitioners.⁵⁷ We identified patients with eligible medications and implemented rules and filters to verify the quality of data.

Evaluation of mapping accuracy

Evaluation of the drugs identified from the 2 mappings was based on whether the patients retrieved were correctly identified for taking a drug of a given group (Figure 1). The focus was whether the drug used to identify the patient has been used in Australia for the purpose of the given ATC group. For example, when a patient is identified by *Mirtazapine*, which is primarily used to treat *depression* but was instead prescribed for the treatment of *anxiety*, the patient was still considered to be a valid identified patient. In such cases, we explored whether a more appropriate ATC code existed to which the same drug could have been mapped. This process was performed by 2 medical experts with each expert reviewing all the patients. The experts were provided with a summary of the patients’ medical

history, including prescription history, diagnosis, and pathology test results to aid them with the review process. Discrepancies in the review were resolved by consensus.

Evaluation metrics

Precision, recall, and F measure calculated against the reference standard were used to compare the 2 mappings.^{58,59} These measures are capable of quantifying inter-rater agreement when a reference standard is available. We calculated precision as the percentage of mapping results that matches the reference standard (equivalent to positive predictive value); recall as the percentage of the reference standard that matches the mapping results (equivalent to sensitivity); and F measure is the harmonic mean between recall and precision.⁵⁹ To calculate these metrics, true positives (TP, patients correctly identified by the mapping and verified by experts), false positives (FP, patients incorrectly identified by the mapping), false negatives (FN, patients not identified by the mapping but by the reference standard), true negatives (TN, patients not identified by either the mapping or the reference standard) were generated. Examples of each measure is provided in [Supplementary Appendix A](#). To assess the statistical significance, we used bootstrap sampling of 10 000 iterations to estimate the 95% confidence intervals of the F measure.⁶⁰

RESULTS

Number of drugs retrieved

Using the 2 mapping sources we first identified drugs for each ATC group in the ePBRN data set. The differences in the number of drugs are not good indicators of mapping performance as mappings to different AMT groups may cause duplicates of the same drug. Nevertheless, the differences between the mappings were within a reasonable range of less than 25%. We also found that PBS

Table 1. Number of AMT drug product concepts identified by the ATC groups with the 2 mappings

ATC Group	OMOP-CDM Mappings	PBS Mappings	Difference
<i>A10 (Drugs used in diabetes)</i>	467	455	12 (2.6%)
<i>N02A (Opioids)</i>	639	542	97 (15.2%)
<i>N06A (Antidepressants)</i>	669	506	163 (24.4%)

Abbreviations: ATC, Anatomical Therapeutic Chemical; OMOP-CDM, Observational Medical Outcomes Partnership Common Data Model; PBS, Pharmaceutical Benefits Scheme.

Table 2. Number of active patients retrieved

ATC Group		OMOP-CDM Mappings	PBS Mappings	Difference
<i>A10 (Drugs used in diabetes)</i>	All sites	5122	5093	29
	Selected site	82	84	2
<i>N02A (Opioids)</i>	All sites	11 314	20 629	9315
	Selected site	62	141	79
<i>N06A (Antidepressants)</i>	All sites	13 295	12 796	499
	Selected site	115	112	3

Abbreviations: ATC, Anatomical Therapeutic Chemical; OMOP-CDM, Observational Medical Outcomes Partnership Common Data Model; PBS, Pharmaceutical Benefits Scheme.

Table 3. Number of overlapping patients of each mapping and percentage of overlapping patients of each mapping

ATC Group	Patients captured in both OMOP-CDM and PBS	Captured in OMOP-CDM	Captured in PBS
<i>A10 (Drugs used in diabetes)</i>	81	98.8%	96.4%
<i>N02A (Opioids)</i>	55	88.7%	39%
<i>N06A (Antidepressants)</i>	112	97.4%	100%

Abbreviations: ATC, Anatomical Therapeutic Chemical; OMOP-CDM, Observational Medical Outcomes Partnership Common Data Model; PBS, Pharmaceutical Benefits Scheme.

mappings overall identified less drugs than OMOP-CDM mappings (Table 1).

Cohort identified in the ePBRN (November 2017 data extract)

Using the drugs retrieved from both mappings, we identified the active patients taking these drugs in the ePBRN data repository (Table 2). The difference in the number of patients identified in *A10* and *N06A* was 29 and 499, respectively. An additional 9315 patients were identified by the PBS compared to OMOP-CDM mappings for *N02A*, even though PBS had 15.2% fewer identified drugs. The results from a selected general practice site within the ePBRN data repository followed a similar pattern with small differences in ATC groups *A10* and *N06A*. More patients were identified using PBS than OMOP-CDM in *N02A* (Table 2).

Overlap between the 2 mappings

There was a high proportion of overlapping patients retrieved by both mappings. In the *A10* group, 81 were overlapping patients; this covered 98.8% of OMOP-CDM mapping and 96.4% of PBS mapping (Table 3). The *N02A* group had a larger difference between the 2 mappings, with OMOP-CDM and PBS having fewer overlapping patients at 88.7% and 39%, respectively. The OMOP-CDM mapping in the *N06A* group was able to identify all the patients from PBS with an overlap of 97.4%.

Mapping accuracy

Evaluation metrics were calculated for active patients from a single general practice site based on the results generated from the expert review process. The 2 experts had an inter-rater agreement (F measure) of 98.4%. “Drugs being prescribed for off-label use” was the most common reason for discrepancies. The recall values of PBS mappings were close to perfect, with all scoring above 0.887; whereas OMOP-CDM mappings scored similarly in the *A10* and *N06A* groups but exhibited a much lower recall of 0.389 for *N02A*. The precision across the board had a value of 1 as there were no falsely identified drugs by either mapping. The F measure for all groups of PBS mapping had values above 0.94, and OMOP-CDM mappings also performed well in all groups except *N02A* with a value of 0.561. Using bootstrap sampling, we found there were no significant differences between the F measure in ATC *A10*, *N06A* groups (Figure 4). However, the F measure for the PBS mapping was significantly higher than OMOP-CDM for *N02A* ($P < .0001$).

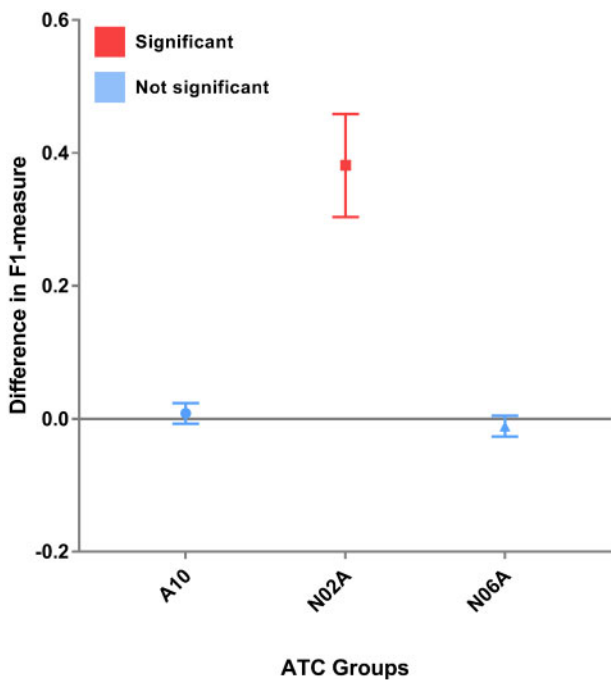
DISCUSSION

This study found that mappings from OMOP-CDM and PBS had their own strengths and weaknesses for cohort selection. The reference standard generated by sourcing from experts and the 2 existing mappings was shown to be an effective method to produce a reference standard where no existing gold standard exists. The experts identified additional drugs which were not picked up by either mapping. OMOP-CDM and PBS were not able to identify all the available patients found in the reference standard. The PBS performed

Table 4. Accuracy of patients retrieved using OMOP-CDM and PBS mappings

ATC Group	OMOP-CDM Mapping			PBS Mapping			Difference in F-measure	95% CI	P-value
	Recall	Precision	F-measure	Recall	Precision	F-measure			
A10 (<i>Drugs used in diabetes</i>)	0.943	1	0.970	0.966	1	0.982	0.012	-0.00693, 0.0241	.234
N02A (<i>Opioids</i>)	0.389	1	0.561	0.887	1	0.940	0.379	0.299, 0.450	<.0001
N06A (<i>Antidepressants</i>)	1	1	1	0.974	1	0.987	0.013	-0.0264, 0.00496	.127

Abbreviations: ATC, Anatomical Therapeutic Chemical; OMOP-CDM, Observational Medical Outcomes Partnership Common Data Model; PBS, Pharmaceutical Benefits Scheme.

**Figure 4.** Difference in F measure across the A10, N02A, and N06A ATC groups.

significantly better than OMOP-CDM for the opioid class drug mappings. All identified patients belonged to the group to which they had been mapped to under the ATC. They therefore all passed expert grading, even when a drug was prescribed for a purpose other than the main condition treated by the ATC group (ie, diabetes mellitus, depression, or pain). This is because we chose to verify the correctness of the drug mappings rather than the validity of using drugs to identify patients with certain conditions. For example, if a patient is taking a drug (*metformin*) primarily used for treating diabetes, but the patient did not have diabetes based on their medical record we still considered the patient as a true case. This decision was justified because of the way cohort selection exercises are performed using algorithms and rules containing multiple variables.¹¹ Accurate identification of phenotype relies on the combination of these variables and how accurate each variable is in identifying values in its own intended domain. Overall, the perfect precision (positive predictive value) of both mappings suggest that either of the mappings would be suitable to use for cohort selection using EHR data.

Further investigation into the group N02A Opioids revealed that 54.7% of patients identified by *Panadeine*- and 2.7% by *methadone*-containing drugs were uniquely mapped by the PBS. The biggest factor causing the difference in this group was due to the

drug 835991000168101 *Panadeine Forte uncoated tablet*, 20. There were approximately 350 000 prescriptions of *Panadeine Forte* and 14 700 unique active patients taking those prescriptions in the ePBRN data repository. *Panadeine* is the trade name for the combination of *paracetamol* and *codeine* with the existing WHO ATC code of N02AJ06. This drug is missing from the N02A class from the OMOP-CDM mapping as it was mapped to *acetaminophen* (*paracetamol*) and not *codeine* (classified as N02A Opioid). This error in the mapping is likely due to the lack of the incorporation of local drug knowledge leading to unmapped concepts in the OMOP-CDM mapping. However, OMOP-CDM was able to pick up drugs such as *rikodeine* and *oxycodone*. This finding is consistent with the knowledge-engineering techniques that were required to produce high performing mappings.¹³

The groups A10 *Drugs used in diabetes* and N06A *Antidepressants* performed similarly between the 2 mappings. However, there were some noteworthy differences in the N06A group. One example of this is *bupropion*, which the WHO ATC code N06AX12 groups as an antidepressant, an indication for which it is used overseas. In Australia, the drug is registered for use and PBS-listed for the treatment of nicotine dependence. In this case the official WHO ATC code was stored as a separate part of the PBS schedule for information only, not for publication. Another ATC code N07BA was assigned to the drug which groups it in the published schedule with other PBS-listed drugs available for treatment of nicotine dependence. For the same reason, it was also assigned to the N07BA group in other geographical regions such as Ireland in the General Medical Services Payments Board prescription database.⁶¹ Although the difference in identified patients was not statistically significant, it is evident that additional knowledge input was utilized by the PBS to reflect drug use in Australia.

Interestingly, when drugs had been prescribed for uses other than their primary purpose, there was not a more appropriate category in the ATC where they could be listed. For example, antidepressants were frequently prescribed to treat neuropathic pain. The only mention of neuropathic pain in the ATC is in the description of N01BX *Other local anaesthetics*. Neuropathic pain is not in the hierarchy or grouped under analgesics. Conversely, when looking at drugs of N02A Opioid class, we discovered that some drugs in this group have also been mapped to a different class with similar chemical properties N07BC *Drugs used in opioid dependence* (eg, *methadone hydrochloride 5 mg/mL oral liquid, 200 mL*, an opioid analgesic with the WHO ATC of N07BC). However, it was assigned to both N07BC and N02A in the PBS mappings to reflect its opioid analgesic properties, whereas OMOP-CDM only assigned N07BC. This raises a problem of reproducibility even when studies use the same ATC group but with different source terminology and mappings. For example, Smith et al⁶² investigated opioid analgesic use during pregnancy using N06A. They manually included *methadone*

and *buprenorphine* even though it was grouped as *N07BC* in their data set. However, a similar study specifically excluded these 2 drugs.⁶³ Additionally, few studies had no mention of drugs included as part of the *N06A* group.^{64,65}

We tested this by identifying the number of patients taking anti-hypertensive drugs in addition to the 3 selected ATC groups. The results are included in [Supplementary Appendix A](#). The ATC group *C02 Antihypertensives* only identified 7 (0.5%) patients from the selected general practice from each mapping—far from the 10.6% prevalence of hypertension in Australia.⁶⁶ We found that most hypertensive patients were also prescribed drugs grouped in *C03 Diuretics* (6.1%) and *C09 agents acting on the renin-angiotensin system* (13.5%). In most of the patients, drugs from more than 1 of the above groups had been prescribed. This suggests that inclusion of all appropriate groups is essential when performing such tasks. While researchers have successfully overcome this issue by carefully selecting and excluding ATC groups,¹⁸ others may misuse the ATC classification system by assuming the named categories include all relevant drugs. We also noticed a difference in the 2 mappings such that an antihypertensive drug *Avapro Hct 300/25 Table* containing both *irbesartan* and *hydrochlorothiazide* was only mapped to *C09DA04 irbesartan and diuretics* by PBS, while OMOP-CDM had 2 additional ATC codes, namely *C03AA03 hydrochlorothiazide* and *C09CA04 irbesartan*.

Limitations

First, we only compared 3 classes of drugs of ATC and the choice of drugs was not random. Translation of our findings to other drug classes is unknown. Second, we have used only data from 1 practice for manual grading. This cohort might not be representative of the population observed in the same health district. Third, we have not addressed the fact that the mappings from OMOP-CDM and PBS are at different stages of maturity. While OMOP-CDM does source from more mature mappings such as UMLS⁶⁷ where possible, there are still a number of newly established mappings within OMOP-CDM that need to develop further. It should be noted that mappings from both sources are regularly updated and ATC mappings of OMOP-CDM have been updated since we conducted this study. Finally, incomplete or incorrect mappings can be caused by factors such as local practice variation. Clinical variations exist within and across health care systems where the clinicians need to provide individualized care according to their local guidelines. In addition, drug repurposing can pose an issue.⁶⁸ The impact of these factors on the performance of electronic phenotyping has not been well investigated. It is also important to note that this study is not a comprehensive and systematic comparison of the 2 mappings, but rather highlights the impact of vocabulary mapping choice on cohort selection, using preselected drug categories, and the need to assess the underlying mappings. In future, we plan to undertake a systematic exploration of these factors across all the ATC categories, as well as compare AMT with other international drug vocabularies.

CONCLUSION

In conclusion, we found both mappings, PBS and OMOP-CDM, were not able to identify all the available patients in our EHR-based data repository. However, both mappings achieved perfect precision, which is crucial for cohort selection in clinical trials. Our findings highlight that underlying terminology mappings can greatly impact the cohort selection process. Clinical researchers should

carefully evaluate the vocabulary mapping sources including methodologies used to develop those mappings. One might need to leverage mappings from various sources and consider incorporating mappings from additional experts for more accurate cohort identification.

FUNDING

This work was supported by UNSW Sydney's Research Infrastructure Scheme (RIS) grant. Vojtech Huser's work was supported by the Intramural Research Program of the National Institutes of Health (NIH)/National Library of Medicine (NLM)/Lister Hill National Center for Biomedical Communications (LHNCBC).

AUTHOR CONTRIBUTIONS

JJ conceived the study. GNG, JJ, and STL developed the study design. GNG, JJ, and STL drafted the article. GNG, JJ, and SF undertook the experiments. STL, VH, and CR provided critical revisions. STL contributed in funding acquisition and supervision.

SUPPLEMENTARY MATERIAL

[Supplementary material](#) is available at *Journal of the American Medical Informatics Association* online.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

- Charles D, Gabriel M, Furukawa MF. Adoption of electronic health record systems among US non-federal acute care hospitals: 2008–2012. *ONC Data Brief* 2013; 9: 1–9.
- Safran C. Using routinely collected data for clinical research. *Stat Med* 1991; 10 (4): 559–64.
- de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. *Fam Pract* 2006; 23 (2): 253–63.
- Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000; 342 (25): 1887–92.
- Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence—what is it and what can it tell us. *N Engl J Med* 2016; 375 (23): 2293–7.
- Gottesman O, Kuivaniemi H, Tromp G, et al. The electronic medical records and genomics (eMERGE) network: past, present, and future. *Genet Med* 2013; 15 (10): 761.
- Banda JM, Halpern Y, Sontag D, et al. Electronic phenotyping with APH-RODITE and the Observational Health Sciences and Informatics (OHDSI) data network. *AMIA Joint Summits on Translational Science proceedings. AMIA Jt Summits Transl Sci* 2017; 2017: 48–57.
- Mo H, Thompson WK, Rasmussen LV, et al. Desiderata for computable representations of electronic health records-driven phenotype algorithms. *J Am Med Inform Assoc* 2015; 22 (6): 1220–30.
- Wei W-Q, Teixeira PL, Mo H, et al. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. *J Am Med Inform Assoc* 2016; 23 (e1): e20–7.
- Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet* 2012; 13 (6): 395.

11. Shivade C, Raghavan P, Fosler-Lussier E, *et al.* A review of approaches to identifying patient phenotype cohorts using electronic health records. *J Am Med Inform Assoc* 2014; 21 (2): 221–30.
12. Liao KP, Cai T, Gainer V, *et al.* Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care Res* 2010; 62 (8): 1120–7.
13. Hripcsak G, Levine ME, Shang N, *et al.* Effect of vocabulary mapping for conditions on phenotype cohorts. *J Am Med Inform Assoc* 2018; 25 (12): 1618–25.
14. Reich C, Ryan PB, Stang PE, *et al.* Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases. *J Biomed Inform* 2012; 45 (4): 689–96.
15. Jonnagaddala J, Dai H-J, Ray P, *et al.* Mining electronic health records to guide and support clinical decision support systems. In: Siaw-Teng Liaw, Andrew Nunn, Tony Sahama, eds. *Improving Health Management through Clinical Decision Support Systems*. Pennsylvania, USA: IGI Global; 2016: 252–69.
16. WHO Collaborating Centre for Drug Statistics Methodology. *Guidelines for ATC Classification and DDD Assignment*. Oslo: WHO; 2005.
17. Natsch S, Hekster YA, Jong R, *et al.* Application of the ATC/DDD methodology to monitor antibiotic drug use. *Eur J Clin Microbiol Infect Dis* 1998; 17 (1): 20–4.
18. Basger BJ, Chen TF, Moles RJ. Application of a prescribing indicators tool to assist in identifying drug-related problems in a cohort of older Australians. *Int J Pharm Pract* 2012; 20 (3): 172–82.
19. Barrows RC Jr, Cimino JJ, Clayton PD. Mapping clinically useful terminology to a controlled medical vocabulary. In: Proceedings of the Annual Symposium on Computer Application in Medical Care. November 5–9, 1994; Washington, DC. Pennsylvania: Hanley & Belfus, Inc.
20. Humphreys BL, Lindberg D. The UMLS project: making the conceptual connection between users and the information they need. *Bull Med Libr Assoc* 1993; 81 (2): 170.
21. Caughey GE, Roughead EE, Vitry AI, *et al.* Comorbidity in the elderly with diabetes: Identification of areas of potential treatment conflicts. *Diabetes Res Clin Pract* 2010; 87 (3): 385–93.
22. Inacio MC, Hansen C, Pratt NL, *et al.* Risk factors for persistent and new chronic opioid use in patients undergoing total hip arthroplasty: a retrospective cohort study. *BMJ Open* 2016; 6 (4): e010664.
23. Vitry AI, Roughead EE, Preiss AK, *et al.* Influence of comorbidities on therapeutic progression of diabetes treatment in Australian veterans: a cohort study. *PLoS One* 2010; 5 (11): e14024.
24. Pahor M, Chrischilles E, Guralnik J, *et al.* Drug data coding and analysis in epidemiologic studies. *Eur J Epidemiol* 1994; 10 (4): 405–11.
25. Huber CA, Szucs TD, Rapold R, *et al.* Identifying patients with chronic conditions using pharmacy data in Switzerland: an updated mapping approach to the classification of medications. *BMC Public Health* 2013; 13 (1): 1030.
26. Peissig PL, Costa VS, Caldwell MD, *et al.* Relational machine learning for electronic health record-driven phenotyping. *J Biomed Inform* 2014; 52: 260–70.
27. Yu S, Ma Y, Gronsbell J, *et al.* Enabling phenotypic big data with Phenorm. *J Am Med Inform Assoc* 2018; 25 (1): 54–60.
28. Saitwal H, Qing D, Jones S, *et al.* Cross-terminology mapping challenges: a demonstration using medication terminological systems. *J Biomed Inform* 2012; 45 (4): 613–25.
29. Defalco FJ, Ryan PB, Soledad Cepeda M. Applying standardized drug terminologies to observational healthcare databases: a case study on opioid exposure. *Health Serv Outcomes Res Method* 2013; 13 (1): 58–67.
30. Choi Y, Jung C, Chae Y, *et al.* Comparison of validity of mapping between drug indications and ICD-10. *Methods Inf Med* 2014; 53 (3): 195–201.
31. Park H-T, Lu D-F, Konicek D, *et al.* Nursing interventions classification in systematized nomenclature of medicine clinical terms: a cross-mapping validation. *Comput Inform Nurs* 2007; 25 (4): 198–208.
32. Fung KW, Bodenreider O, Aronson AR, *et al.* Combining lexical and semantic methods of inter-terminology mapping using the UMLS. *Stud Health Technol Inform* 2007; 129 (Pt 1): 605.
33. Steinberg DM, Fine J, Chappell R. Sample size for positive and negative predictive value in diagnostic research using case-control designs. *Biostatistics* 2008; 10 (1): 94–105.
34. Burt T, Button KS, Thom HHZ, *et al.* The burden of the “false-negatives” in clinical development: analyses of current and alternative scenarios and corrective measures. *Clin Transl Sci* 2017; 10 (6): 470–9.
35. Schulz S, Rodrigues JM, Rector A, *et al.* Interface terminologies, reference terminologies and aggregation terminologies: a strategy for better integration. *Stud Health Technol Inform* 2017; 245: 940–4.
36. Chute CG. Clinical classification and terminology: some history and current observations. *J Am Med Inform Assoc* 2000; 7 (3): 298–303.
37. Bowman SE. Coordination of SNOMED-CT and ICD-10: Getting the Most Out of Electronic Health Record Systems. *J AHIMA* 2005; 76 (7): 60–1.
38. Chiang MF, Hwang JC, Alexander CY, *et al.* Reliability of SNOMED-CT coding by three physicians using two terminology browsers. *AMIA Annu Symp Proc* 2006; 2006: 131–5; Washington, DC.
39. World Health Organization. Classification of diseases (ICD). 2018. <https://www.who.int/classifications/icd/en/>. Accessed December 23, 2018.
40. Jonnagaddala J, Hu F. Automatic coding of death certificates to ICD-10 terminology. In: Linda Cappellato, Nicola Ferro, Lorraine Goeuriot, Thomas Mandl, eds. *CLEF (Working Notes)*. Dublin, Ireland: Sun SITE Central Europe; 2017.
41. National Clinical Terminology Service. Australian medicines terminology fact sheet. 2016. <https://www.digitalhealth.gov.au/using-the-my-health-record-system/digital-health-training-resources/guides/australian-medicines-terminology-amt-fact-sheet>. Accessed November 14, 2018.
42. International Health Terminology Standards Development Organisation. SNOMED CT July 2018 International Edition—SNOMED International Release notes. 2018. <https://confluence.ihtsdotools.org/display/RMT/SNOMED+CT+July+2018+International+Edition+-+SNOMED+International+Release+notes>. Accessed December 11, 2018.
43. Bodenreider O, Rodriguez LM. Analyzing U.S. prescription lists with RxNorm and the ATC/DDD Index. *AMIA Annu Symp Proc* 2014; 2014: 297–306.
44. Hripcsak G, Wilcox A. Reference standards, judges, and comparison subjects: roles for experts in evaluating system performance. *J Am Med Inform Assoc* 2002; 9 (1): 1–15.
45. Hripcsak G, Friedman C, Alderson PO, *et al.* Unlocking clinical data from narrative reports: a study of natural language processing. *Ann Intern Med* 1995; 122 (9): 681–8.
46. ePBRN. ePBRN/AMT-Study-JAMIA. 2019. <https://github.com/ePBRN/AMT-Study-JAMIA>. Accessed June 5, 2019.
47. Pharmaceutical Benefits Scheme. PBS publications archive. 2018. <http://www.pbs.gov.au/info/publication/schedule/archive>. Accessed December 11, 2018.
48. Hripcsak G, Duke JD, Shah NH, *et al.* Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015; 216: 574.
49. US National Library of Medicine. RxNorm release notes. 2013. https://www.nlm.nih.gov/research/umls/rxnorm/docs/2013/rxnorm_release-notes_full_08052013.html. Accessed December 11, 2018.
50. Observational Health Data Sciences and Informatics. Release notes 2017-09-06. 2017. <https://github.com/OHDSI/OMOP-Standardized-Vocabularies/blob/master/Release%20Notes%202017-09-06.txt>. Accessed November 10, 2018.
51. Observational Health Data Sciences and Informatics. Athena—OHDSI Vocabularies Repository. 2015. <http://athena.ohdsi.org/search-terms/terms/43351380>. Accessed December 10, 2018.
52. Taggart J, Liaw ST, Yu H. Structured data quality reports to improve EHR data quality. *Int J Med Inform* 2015; 84 (12): 1094–8.
53. Boyle D, Rafael N. BioGrid Australia and GRHANITE™: privacy-protecting subject matching. *Stud Health Technol Inform* 2011; 168: 24–34.
54. Liaw S-T, Boyle D. Secure data linkage and information sharing with GRHANITE. In: Heather Grain, ed. HISA Health Informatics Conference

- (HIC) 2008. Melbourne: Health Informatics Society of Australia; August 31–September 2 2008, pp. 159–165.
55. Farshid S, Jonnagaddala J, Guo GN, Wu M, Liaw, S.-T. Harmonising primary care data using international standard vocabularies for observational research; 2018. <https://zenodo.org/record/1401691#.XUF9ngzaUn>. Accessed December 10, 2018.
56. Kohler F, Liaw S-T, Pennock R, *et al*. Integrated health care—a population health approach in South Western Sydney. *Int J Integr Care* 2014; 14 (9).
57. Royal Australian College of General Practitioners. *RACGP Standards for General Practices*. 5th ed. East Melbourne, Australia: RACGP. 2017.
58. Meystre S, Haug PJ. Natural language processing to extract medical problems from electronic clinical documents: performance evaluation. *J Biomed Inform* 2006; 39 (6): 589–99.
59. Hripcsak G, Rothschild AS. Agreement, the F-measure, and reliability in information retrieval. *J Am Med Inform Assoc* 2005; 12 (3): 296–8.
60. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *International Joint Conference on Artificial Intelligence*, 1995.
61. Tilson L, Bennett K, Barry M. Prescribing trends for nicotine replacement therapy in primary care. *Irish Med J* 2004; 97 (9): 270–73.
62. Smith MV, Costello D, Yonkers KA. Clinical correlates of prescription opioid analgesic use in pregnancy. *Matern Child Health J* 2015; 19 (3): 548–56.
63. Handal M, Engeland A, Rønning M, *et al*. Use of prescribed opioid analgesics and co-medication with benzodiazepines in women before, during, and after pregnancy: a population-based cohort study. *Eur J Clin Pharmacol* 2011; 67 (9): 953.
64. Dale O, Borchgrevink PC, Fredheim OMS, *et al*. Prevalence of use of non-prescription analgesics in the Norwegian HUNT3 population: impact of gender, age, exercise and prescription of opioids. *BMC Public Health* 2015; 15 (1): 461.
65. Hasselström J, Liu-Palmgren J, Rasjö-Wrååk G. Prevalence of pain in general practice. *Eur J Pain* 2002; 6 (5): 375–85.
66. Australian Bureau of Statistics. 4364.0.55.001—*National Health Survey: First Results, 2014–15*. 2015. Canberra, Australia: Australian Bureau of Statistics.
67. Dhombres F, Charlet J. As ontologies reach maturity, artificial intelligence starts being fully efficient: findings from the section on knowledge representation and management for the yearbook 2018. *Yearb Med Inform* 2018; 27 (01): 140–5.
68. Hamunen K, Laitinen-Parkkonen P, Paakkari P, *et al*. What do different databases tell about the use of opioids in seven European countries in 2002? *Eur J Pain* 2008; 12 (6): 705–15.