Identifying and reducing inappropriate use of medications using Electronic Health Records

Hojjat Salmasian

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2015

ABSTRACT

Identifying and reducing inappropriate use of medication using electronic health records

Hojjat Salmasian

Inappropriate use of medications (IUM) is a global problem that can lead to unnecessary harm to the patients and unnecessary costs across the health care system. Identifying and reducing IUM has been a long-lasting challenge and currently, no systematic and automated solution exists to address it. IUM can be manually identified by experts using medication appropriateness criteria (MAC).

In this research I first conducted a review of approaches used to identify IUM and reduce IUM. Next, I developed a conceptual model for representing the MAC, and then developed a tool and a workflow for translating the MAC into structured form. Because indications are an important component of the MAC, I conducted a critical appraisal of existing knowledge sources that can be used to that end, namely the medication-indication knowledge-bases. Finally, I demonstrated how these structured MAC can be used to identify patients who are potentially subject to IUM and evaluated the accuracy of this approach.

This research identifies the knowledge gaps and technological challenges in identifying and reducing IUM and addresses some of these gaps through the creation of a representation for MAC, a repository of structured MAC, and a set of tools that can assist in evaluating the impact of interventions aimed to reduce IUM or assess its downstream effects. This research also discusses the limitations of existing methods for executing computable decision support rules and proposes solutions needed to enhance these methods so they can support implementation of the MAC.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This thesis and the projects described herein would not come to be without the continuous and unconditional support that I received from my advisor, Carol Friedman, who provided me with a nurturing environment to follow my ambitions and fulfil my dream of becoming an independent scholar. Carol builds an atmosphere in which one constantly strive to become a better person, professionally and otherwise, and for that I am deeply grateful.

To my thesis committee: Herbert Chase, Gilad Kuperman, David Vawdrey and David Bates, I would like to take this opportunity to express my gratitude for your endless support and suggestions throughout my project. I am sincerely grateful for everything I learned from you and feel both humbled and honored for that opportunity.

I would also like to acknowledge, with utmost gratitude and sincerity, my parents for their endless love, support and encouragement. They have devoted their lives in teaching me how to be determined in life and to be proud and humble at the same time. I don't imagine a life without their love and blessing.

My thanks to my many friends and colleagues at the Department of Biomedical Informatics, who have assisted me in the course of this work, especially Adler Perotte, Rimma Pivovarov and Nicole Weiskopf.

I would like to dedicate this thesis to my brilliant and exceptionally loving and supportive wife, Mersedeh, who gives meaning to my life.

Chapter 1: Introduction and significance

Inappropriate use of medications (IUM) is a serious issue of global concern. It has been reported in studies conducted in different countries such as Germany, United Kingdom (UK), Italy, Lebanon and the United States (US).[1–5] IUM not only leads to a waste of healthcare resources, but also potentially harms the patients due to inadvertent side effects.[6–8]

IUM is part of the general problem commonly known as unnecessary care or overuse. Unnecessary care is a well-known problem; it is estimated that in one in every three dollars spent on healthcare in the US is excess cost, and that approximately one-fourth of this excess cost is due to unnecessary care.[9] However, the scope of IUM is not well understood. Previous studies have shown that several groups of medications are frequently subject to inappropriate use including antibiotics, antidepressants, antipsychotics, bronchodilators, cyclooxygenase-2 (COX-2) inhibitors, non-steroidal anti-inflammatory drugs (NSAIDs), proton pump inhibitors (PPIs), and statins.[10–14] Numerous studies have aimed to develop methods for identifying and reducing IUM, with the focus of their intervention spanning from healthcare professionals and patients, to financial, organizational, and regulatory approaches. However, many of these approaches rely on manual interventions and are difficult to automate. The problem of IUM is still considered understudied.[15]

Identifying IUM depends on extracting certain information about the patient, the medication, and other treatments, and comparing them with guidelines that defines appropriate use; these reference standards are usually called "appropriateness criteria".[16] Traditionally, manual processes are used to extract the required information and compare them against the appropriateness criteria. However, the advent of informatics and the increased adoption of

electronic health record (EHR) systems provide an exceptional opportunity to automate these processes.

Reducing IUM often involves some form of feedback to the providers. EHR systems can facilitate these types of interventions, for example through computerized decision support (CDS) via the computerized provider order entry (CPOE) component of the EHR. Currently, studies related to automated methods for identifying and reducing IUM through automated methods are scarce.

In this thesis, I aimed to develop a framework for automated identification and reduction of IUM. This included developing methods for defining and quantifying IUM, formally representing them, and deploying them into a computable form.

Root Causes of IUM

The most common reasons for overtreatment at the level of provider include peer pressure, outdated knowledge (old habits), lack of expertise, education in tertiary centers and defensive medicine.[17,18] Providers are also more likely to overuse diagnostic and therapeutic resources when they are more accessible to them.[19] Additionally, lack of price transparency hinders both providers' and patients' perception of the true cost of care.[20] Patients also play a significant role in the problem of medication overuse. Patient demand is considered a significant drive for unnecessary care.[17,18] Previous studies have shown that patients equate more testing and treatment with better care,[21] and providers believe patients are more likely to switch providers if they do not receive the care they want.[18]

At the policy level, the fee-for-service payment model has been blamed for incentivizing providers to perform more unnecessary procedures, thereby contributing to the overuse

problem.[22] Although this may not directly drive medication overuse (typically providers are not paid more if they prescribe more), it may still nurture a culture of overuse. Shapiro *et al* name various financial approaches to counteract this cultural effect, such as expenditure caps and global budgeting for hospitals.[22] The effect of payment model on medication overuse is understudied, and previous studies with a broad focus on all types of overuse (including medications, diagnostics, etc.) have equivocal conclusions: some authors believe reducing overuse should be achieved through changing policies governing healthcare expenditure,[23] while others argue the reduction of overuse is a prerequisite for – and not a result of – changes in policy,[24] and that payment reform alone may not lead to significant improvements in healthcare expenditure.[25] A recent systematic review of literature also identified studies that showed the rate of use of antibiotics was higher in managed care settings compared to the fee-for-service sector.[26] Finally, policies regarding direct-to-consumer advertising of prescription drugs are also considered to have an influence on overuse of prescription medications.[18] However, medication overuse is not exclusively controlled by policies and regulations. As an example, the rate at which proton pump inhibitors (PPIs) are overused is similar in the US, UK, Italy and Lebanon [2–5] while these countries have different payment models and regulations regarding direct-to-consumer advertisements.

Appropriateness Criteria

In the medical domain, the most common approach used for identifying IUM is via the application of "medication appropriateness criteria". Appropriateness criteria are not unique to medications, and similar criteria also exist for assessing the appropriate use of other medical services, such as radiology tests,[27] and other diagnostic procedures.[28]

Appropriateness criteria are developed using expert consensus and through review of literature.[29] These criteria are only available in free-text form, and, as shown later in this thesis, they commonly contain phrases that are context-dependent or vague, thereby allowing for differences in the interpretation of the criteria, which can lead to inconsistent application of the criteria. These factors limit the application of appropriateness criteria for automated identification of IUM. Additionally, even when used manually, the criteria can only identify those uses of medication that are *potentially* inappropriate; therefore, authors frequently used the phrase *potentially inappropriate medications* (PIMs) to describe IUM.

Outline of Aims

This thesis consists of four aims, as depicted in **Figure 1**. First, knowledge gaps are identified (Aim I, Chapter 2). Next, the most prominent knowledge gap is addressed (in Aim II, Chapter 3) by creating a structured representation for medication appropriateness criteria (MAC) and a workflow is created and evaluated for translating narrative MAC into computable form (Aim III, Chapter 4). Last, structured MAC are deployed using EHR data, and challenges in scaling this process using existing tools for automated implementation of guidelines and decision support systems are discussed (Aim IV, Chapter 5).

In the next pages, the specific aims are described in more detail. Studies that address each aim are described in Chapters 2 to 5.

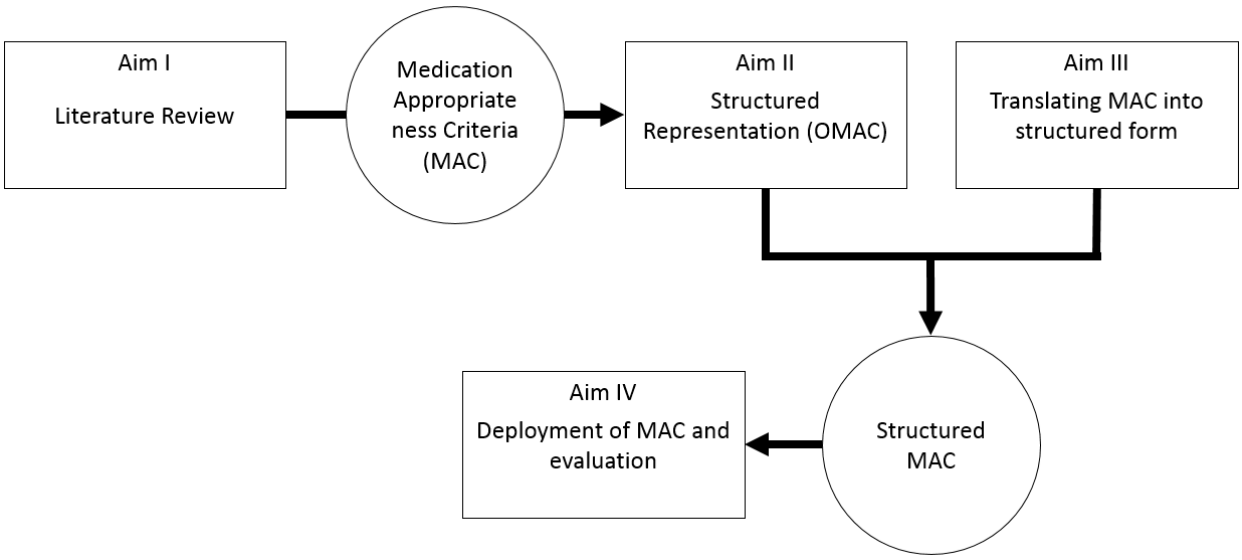


Figure 1 – Overview of the aims.

Aim I: Identify the methods used for detecting and reducing IUM

Objective: Identify the most commonly used methods for detecting IUM and the interventions that aim to reduce IUM.

Research Questions:

- What categories of IUM exist?
- What methods are used to identify IUM?
- What interventions are used to reduce IUM?
- What are the current challenges in developing automated solutions for detecting and reducing IUM?

Methods:

The primary methodology of this aim was a systematic review of the literature. The reviewer focused on peer-reviewed articles describing original research that included (a) identifying IUM

and (b) interventions aimed to reduce IUM. A two stage process was used to identify articles. First, a search was performed in PubMed using appropriate keywords and relevant MeSH headings. This search produced a large number of results, which were then narrowed down to relevant articles by manual review. Other systematic reviews in this field were held out, so that their categorization of methods would not affect our results. Next, the methods used for identification and/or reduction of IUM were extracted from each original article. Finally, these methods were aggregated using thematic analysis through grounded theory, and the final classification was compared with other pertinent systematic reviews that were found in the original query, to ensure the completeness of the classification.

Primary Findings:

Three major categories of IUM were described in the literature: inappropriate by indication, inappropriate by population, and inappropriate by route/strength/frequency. Almost all studies used some form of "appropriateness criteria" for identifying IUM. These criteria could be implicit (i.e. a clinician would be asked to provide their personal judgment on appropriateness of the use of certain medication for a certain patient) or explicit (i.e. objective criteria were provided using which decision on appropriateness was made. The most commonly used form of implicit criteria is the Medication Appropriateness Index (MAI).[30] Explicit criteria vary for different medications, and for some groups of medications multiple appropriateness criteria were found which at times were not completely congruent. The most well-known explicit criteria are those that target IUM in the elderly. They include the Beer's criteria,[31] and the screening tool of older person's prescriptions (STOPP).[32]

Thematic analysis showed that there are five types of interventions that are commonly used to reduce IUM. These include feedback to providers (which include academic detailing), provider

profiling, enforcing appropriate use guidelines, indication-based prescribing, and education approaches.

The findings from literature review indicate that because the majority of the methods used for identifying IUM either partially or entirely rely on manual processes, automation of this process is not feasible unless objective appropriateness criteria are encoded into structured form. In contrast, some of the methods used for reducing IUM, e.g. indication-based prescribing, feedback to providers, and enforcement of guidelines have automated counterparts that are either currently in use in the modern EHR systems or at least theoretically feasible. The results of implementing these approaches for reducing IUM in an automated fashion, however, have been variable and not always successful. This could be in part due to the inaccuracy and lack of standardization of existing automated methods for identifying IUM.

Aim II: Developing a conceptual model for medication appropriateness criteria (MAC)

Objective: Develop a conceptual model for structured representation of medication appropriateness criteria (MAC).

Hypothesis: A conceptual model can be developed that can represent the majority of existing MAC in a computable form.

Research Questions:

- What are the essential concepts that comprise the MAC?
- What proportion of concepts that appear in the MAC can be represented using standardized objective medical concepts?

- What proportion of concepts that appear in the MAC is not explicitly defined (NED)?
- Can an anonymous, iterative approach assist the experts with reaching consensus on the breakdown of concepts identified in the MAC?

Methods:

Aim II relied upon a qualitative approach in order to develop an understanding of what concepts comprise the MAC, as well as an assessment of existing knowledge-bases (KBs) that are used to represent the appropriate indications for medications.

The analysis of existing KBs was again based on a systematic review of literature, and a critical appraisal of the KBs. PubMed was searched using suitable keywords to identify all studies in which a KB for medication indications was developed. The inclusion criteria was restricted to those articles that were describing a freely available resource. The methods used to develop each KB was qualitatively analyzed and thematically grouped. Using a survey, a group of domain experts (clinicians and pharmacists) were asked to identify examples of "challenging" indications, and the results were combined to create an assessment tool that was used to evaluate the KBs. Additionally, the coverage of, and overlap between various KBs was visualized using descriptive statistics and graphical visualizations, respectively.

In another study, a randomly selected series of 40 MAC were analyzed manually by this author to identify the elements that comprise these MAC. These elements were then classified into the minimum sufficient number of categories to include them all, and relationships between these classes were defined. The entire conceptual model was then represented as a frame-based model, using Protégé, and was called Objective Medication Appropriateness Criteria (OMAC). This process is also described schematically in **Figure 2**. Next, an independent group of experts

evaluated a held-out set of MAC and identified the elements in a similar fashion. This data was used to evaluate OMAC for its completeness and accuracy. Then the conceptual model was qualitatively evaluated to assess whether all the elements identified by independent experts were included in the representation.

Primary Findings:

Seven publicly available medication-indication KBs were found that contained structured information regarding medications and their indications. There was a large amount of non-overlap between existing medication-indication KBs. Furthermore, these KBs were limited because of lack of normalization of indication concepts into a structured form, lack of specification for preventive versus therapeutic indications, lack of information regarding appropriate dose forms and dosages of medication, lack of information regarding contraindications or the primary choice of therapy, lack of information regarding co-medications, or issues of granularity in the definition of indications.

We also demonstrated that even though experts may have had internal disagreements in how they

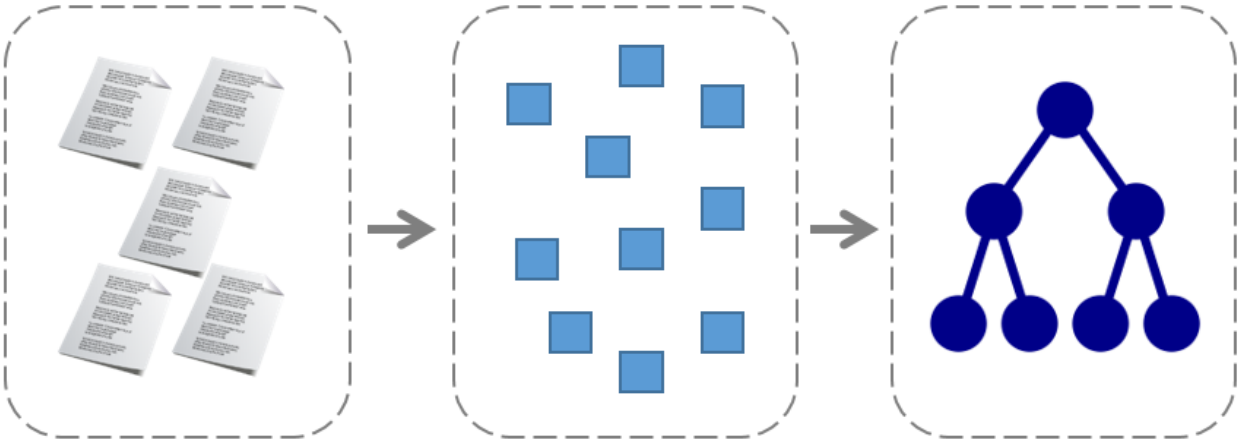


Figure 2 – Developing a formal representation for medical appropriateness criteria.

identified the concepts comprising the MAC, these disagreements could be resolved using an anonymous, iterative consensus-building process, and in our experience only two iterations were enough to reach consensus using this approach.

OMAC, provided a conceptual framework that allowed representing all of the elements of MAC, as long as these concepts were objectively defined. Several instances of MAC were found where an ambiguous or context-dependent term was used; this signifies the need for a "translation" stage in which the MAC are first reviewed by the experts and all the necessary disambiguations are performed. The latter step leads to Aim III below.

Aim III: Translating MAC into computable form

Objective: Developing and demonstrating a workflow for translating MAC into computable form, including a process for substituting vague or context-dependent terms with explicit definitions.

Hypothesis: Annotation of MAC into computable form can be facilitated using pre-annotation of the MAC through an automated process.

Hypothesis: It is possible to develop a tool for domain experts to elicit objective and computable definitions for NED concepts in the MAC.

Research Questions:

- What proportion of concepts that appear in MAC can be represented using existing biomedical terminologies?
- Can experts substitute all NED concepts that appear in MAC with explicit definitions based on concepts in existing terminologies?

- Can experts reach agreement in substituting NED concepts with explicit definitions using an anonymous, iterative consensus forming process?

Methods:

We aimed to create a workflow that would allow translating a large body of MAC into computable form, and demonstrate its usability in a lab study. This was achieved in multiple steps. First, we developed a workflow for anonymous, collaborate disambiguation of the MAC by experts. The workflow was based on the Delphi approach,[33] and was originally implemented in form of an electronic survey tool. We used a randomly selected sample of existing MAC which spanned different medications, and asked a group of experts to disambiguate the concepts to the point that every concepts was a well-defined (objective) entity, such as a disease, a medication, a clinical metrics, a number, etc. We then qualitatively analyzed this process to identify the amount of time necessary for the experts to reach consensus, the areas in which consensus was easier or harder to achieve, and the proportion of concepts in the MAC that needed disambiguation. Using these findings, we then designed a prototype tool that aimed to facilitate the process of translating MAC into computable form, with the help of an interactive web-based tool. This tool, called MAC Annotator, uses on-the-fly natural language processing (NLP) capabilities of an online biomedical text annotation service, namely the NCBO Annotator, and facilitates the process of normalizing the concepts to those found in existing biomedical terminologies. Finally, we assessed the usability of this prototype by asking a group of experts to utilize it to disambiguate another held-out set of MAC and assessed various metrics such as inter-annotator agreement, and time needed to complete each task.

Primary Findings:

We identified that many concepts found in MAC are already in an objective form and do not require disambiguation. However, MAC commonly contain context-dependent terms (such as "monotherapy", "uncomplicated" or "long-term") which require the use of external knowledge to be translated into a set of objective concepts that could be represented in OMAC. Additionally, certain concepts in the MAC which are supersets or subsets of concepts in existing terminologies may not be directly found in any terminology, indicating that issues of granularity with existing biomedical terminologies may prevent accurate representation of some of the MAC in structured form. Using the MAC Annotator, we noticed a high degree of agreement between annotators; this was also observed in the preceding study which used a survey tool, and we also showed that using only two iterations the experts could reach consensus about disambiguating the vague or context-dependent concepts. The average time spent for annotating the MAC, including all automated annotations that were suggested by the tool, was only a few minutes.

Aim IV: Deploying structured MAC using EHR data

Objective: Demonstrate the feasibility of implementing structured MAC by applying it to EHR data, and identify the associated challenges.

Hypothesis: Once converted into structured form, the MAC can be applied to EHR data to identify patients who are potentially subject to IUM, with accuracy comparable to manual review.

Research Questions:

- Is it possible to identify IUM using structured MAC and data from the EHR with an accuracy comparable to the manual review of medical records and manual application of the MAC?

- What challenges exist in deploying the structured MAC and applying it to clinical data that is stored in the EHR?
- Can the implementation of MAC be scaled using existing methods for automated guidelines or clinical decision support rules?

Methods:

In this aim, we first conducted a proof-of-concept study in which we demonstrated that it is possible to identify IUM in an automated way using structured MAC, in comparison to manual application of the same criteria through manual chart review by the experts. Next, we posited that the MAC can be viewed as a special form of clinical guidelines or decision support rules. Therefore, we assessed whether the deployment of MAC can be scaled using existing methodologies that are used for automated implementation of clinical guidelines and decision support rules. Specifically, we assessed whether MAC can be readily represented as guidelines using the Guideline Interchange Format (GLIF), and whether it can be readily represented using the free tools that are developed by the OpenCDS project. This involved a review of the internal structure of GLIF and OpenCDS along with their execution engines, and a comparison of their data models with what is necessary for structured representation of MAC. The latter studies were descriptive only, and the analyses were qualitative.

Primary Findings:

The proof-of-concept study showed that it is possible to identify IUM, at least for one specific test case, with a very high sensitivity and specificity, when using manual review as the gold standard. The manual review process was time-intensive, while the automated solution was feasible within a fraction of a second. However, this study also showed two challenges in

automated implementation of MAC. First, a notable amount of information that is needed to implement the MAC comes from unstructured data. In our case, because we used a state-of-the-art NLP solution, this data was available to the automated solution, but this may not always be the case; for instance, social factors or other clinical factors that are not captured by NLP or are not documented at all may not be apparent to the automated solution, resulting in false positives or negatives. Second, manual reviewers occasionally used rules outside the original MAC to determine appropriateness. In other words, manual reviewers had a tendency to use "implicit" criteria in addition to explicit criteria to determine the appropriateness of medications, and because these implicit criteria were not available to the automated system, the results were not ideal.

Finally, our analysis of GLIF and OpenCDS showed that both of these models use high-level approaches for representing the guideline and CDS rules. None of them had built-in support at the level of granularity that is needed to represent the MAC in structured form, indicating the need for at least two additional layers of abstraction before MAC can be implemented using either GLIF or OpenCDS. One layer of abstraction would be necessary to specify the necessary elements of MAC in a structured form; this may involve specifying a data model, developing validation methods to ensure the compliance of input data with the expected format, etc. A second layer of abstraction would also be needed to translate the clinical data form EHR into the specific format that each of these tools can accept as input. The latter may involve mapping between terminologies, use of natural language processing, use of phenotyping algorithms, and so on, and is outside the scope of this thesis.

Significance and Contributions

Although the issue of IUM has been studied for decades and a few automated methods for reducing IUM have been developed, research on automated methods for identifying and reducing IUM remains scarce. Additionally, existing appropriateness criteria are usually difficult to automate. Finally, electronic methods that have been used to mitigate IUM are mainly interruptive and hard to generalize, and their scope is limited to specific types of IUM such as drug-drug interaction. The significance of this thesis is that it identifies two main sources of information that are necessary for computable applications targeting IUM (namely, medication appropriateness criteria and the medication-indication knowledge-bases), provides a formal representation for MAC, provides a critical appraisal tool for assessing medication-indication KBs, enables dissemination of a large number of computable criteria for identifying IUM, and should lead to improved medication safety and health care via incorporation into an EHR system.

The biggest contribution of this investigation is the development of a computable framework for representing and implementing well-defined appropriateness criteria that can be incorporated into a clinical decision support system. In this thesis, the framework will be applied only to medication appropriateness criteria but the framework is general and could be applicable to other appropriateness criteria as well. Specific contributions include: (1) a conceptual model for defining explicit appropriateness criteria for medications, (2) a tool that facilitates translation of narrative MAC into structured form with the help of experts, (3) computable appropriateness criteria for different medications that are frequently overused, and (4) a tool within the EHR system that provides decision support about appropriateness of medications and facilitates pharmacist-mediated detailing. This investigation will also provide scientific evidence about the usability and efficacy of this tool in a controlled environment.

Limitations

The scope of the current work is only limited to those medications for which MAC exists. While the approaches are designed to be generalizable, other forms of MAC that are developed in the future may require modifications in the representation, and may affect the process for the *translation* and *localization* steps, that are unforeseeable at the moment.

Another limitation of the current work is that although it provides a framework for converting the MAC into structured form, it does not provide a repository of executable MAC. This can be addressed in future work, where the MAC is translated into executable cohort definitions that are encoded using a standard clinical database format such as that developed by the Observational Health Data Science and Informatics (OHDSI).[34]

Chapter 2: Methods used to identify and reduce IUM

Background

The first logical step in addressing the problem of IUM using automated solutions is to identify the existing methods that are used, whether manually or in automated form, to identify and reduce IUM. Our aim is to avoid reinventing existing solutions and to identify the knowledge gaps that need to be addressed.

In this chapter, the current state of knowledge about identifying and reducing IUM is described using a rapid review approach. This chapter describes the current state of the art and concludes with recognizing the existing knowledge gaps in this domain. This serves as the cornerstone of the subsequent chapters, justifying their significance and explaining the choice of methods and material at the high level.

Study 1: Methods used to identify and reduce IUM

Introduction

As stated in Chapter 1, IUM is a global problem leading to unnecessary cost and harm to the patients and the health care system as a whole. With the wide adoption of EHR systems and as electronic health care data is becoming available, it is desirable to address the problem of IUM using automated solutions. The first step in developing automated solutions for identifying and reducing IUM using EHR data is to review the domain literature to identify previous work on these methods, including both manual and automated efforts that have been previously conducted. In this review, we identified and summarized the relevant published studies. By the end of this chapter, the reader should be familiarized with various types of approaches that have been used to identify and reduce IUM.

This review summarizes existing approaches used to identify and reduce medication overuse, and intends to provide directions about the potential role of informatics in tackling this problem.

Methods

We conducted a rapid review of the literature during December 2012 and January 2013. A rapid review is defined as a "streamlined approach to synthesizing evidence in a timely manner".[35] Rapid reviews exercise slightly different approaches compared to traditional systematic reviews (such as limiting search strategies or record screening processes) to produce overviews of evidence in shorter time frames.

PubMed was searched through the end of 2013 to identify existing studies in the literature that discuss methods used for identifying or reducing IUM. The following search strategy was used: "Inappropriate Prescribing"[Mesh] OR overuse?[tiab] OR overtreat*[tiab] OR overutiliz*[tiab]

OR overutilis*[tiab] OR "utili?ation review"[tiab] OR "utilization review"[MeSH Terms]. We limited the scope of this review to original studies for which an English abstract and full text were available.

We reviewed the titles and abstracts of the retrieved results, identified potentially relevant studies and obtained their full texts. Subsequently, relevant studies were identified by reviewing the full texts, and a six-phase thematic analysis approach was used to summarize these studies.[36] The first phase of the thematic analysis focused on becoming familiar with the data and organizing the results into summaries. In the second phase, each article was assigned one or more codes based on its emerging theme; examples of these themes included 'medication appropriateness index', 'academic detailing', 'physician auditing' and 'physician profiling'. In the third phase, subsequent queries were executed on PubMed using new keywords learned from the previous phase to identify any additional studies about the same themes and identify the overlaps between the themes. In the fourth phase, the themes were reviewed and those that were very closely related were combined (e.g. 'physician auditing' was combined into 'physician profiling'). In the last two phases, the remaining themes were labeled and a definitions were provided.

All review articles and editorials were excluded from the thematic analysis, to ensure that our thematic analysis would not be influenced by that of other researchers. After the thematic analysis was completed, a comparison was made between themes emerging from the current study and those mentioned in other review articles in this field. This allowed evaluating the completeness of this review, and also allowed for exploring changes in the trending topics over time.

Results

Our search in PubMed generated 10718 citations, which were screened thoroughly by the first author to identify potentially relevant studies. Out of 1564 studies that were marked as possibly relevant, 181 original studies were included in the final review. After thematic analysis, two classifications were found for methods used in identifying IUM, and six major themes emerged among methods used to reduce IUM, as described below.

Identifying IUM

Identifying IUM is rarely a trivial task. In its simplest form, it involves ascertaining that a medication was utilized, identifying the reason for its utilization, and appraising whether the reason is appropriate. Most medications have more than one appropriate use (indication). For instance, proton pump inhibitors (PPIs) have numerous possible appropriate uses such as 'stress ulcer prevention', treatment of 'gastrointestinal bleeding', or as therapy for 'Barrett syndrome'.[37] In addition, identifying IUM is more difficult when factors such as disease severity or the clinical context play a role in the appropriateness of therapy. For instance, ascertaining a diagnosis of "depression" may not be enough to ensure appropriate use of antidepressants, as their benefit is questionable in those with only mild or moderate depression.[38] Similarly, other contextual information such as age, comorbidities, coexisting symptoms, frequency and dose of medication, route of administration, etc. can also contribute to the identification of IUM. In fact, patient age is very commonly listed as a criterion in tools focused on inappropriateness and several guidelines have been developed that specifically focus on the appropriate use of medication in the elderly.[32,39,40]

To assess the appropriateness of the drug, please answer the following questions and circle the applicable score:				
1. Is there an indication for the drug? Comments:	1 Indicated	2	3 Not Indicated	9 DK†
2. Is the medication effective for the condition? Comments:	1 Effective	2	3 Ineffective	9 DK
3. Is the dosage correct? Comments:	1 Correct	2	3 Incorrect	9 DK
4. Are the directions correct? Comments:	1 Correct	2	3 Incorrect	9 DK
5. Are the directions practical? Comments:	1 Practical	2	3 Impractical	9 DK
6. Are there clinically significant drug–drug interactions? Comments:	1 Insignificant	2	3 Significant	9 DK
7. Are there clinically significant drug–disease/condition interactions? Comments:	1 Insignificant	2	3 Significant	9 DK
8. Is there unnecessary duplication with other drug(s)? Comments:	1 Necessary	2	3 Unnecessary	9 DK
9. Is the duration of therapy acceptable? Comments:	1 Acceptable	2	3 Unacceptable	9 DK
10. Is this drug the least expensive alternative compared to others of equal utility? Comments:	1 Least expensive	2	3 Most expensive	9 DK

*Complete instructions in the use of the scale are available upon request.
†Don't know.

Figure 3 – The Medication Appropriateness Index (MAI).

Methods used for identifying IUM may either focus on "inappropriateness" of the medication use – i.e. they would specify conditions in which the medication should be avoided, e.g. the Beer's criteria[31] – or on "appropriateness" of the medication use – i.e. listing all conditions in which its use is justified, e.g. the criteria by Ahrens et al on the use of PPIs[41] – or would include items about both appropriate and inappropriate use of medications, e.g. the works of Oborne et al.[40,42]

Our thematic analysis showed that the major approaches in identifying IUM can be categorized with respect to two characteristics: focus, and formalism. Focus is based on the number of medications and types of populations that are targeted. For instance, Beers criteria consists of several items each focusing on conditions in which certain medications should be avoided in the elderly (inappropriate use), [39] while Ahrens *et al* provide a list of indications in which the use of PPIs is appropriate, therefore identifying overuse of PPIs by exclusion.[41] The former focuses on a specific patient population and includes a large list of medications, while the latter doesn't focus on any specific population but includes only one class of medications.

Formalism can be described as a spectrum ranging from implicit to explicit. Implicit criteria are frequently used for manual chart review where a subjective judgment call is made by an expert about the appropriateness of the medication use. The most well-known example of implicit criteria is the Medication Appropriateness Index (MAI), a tool that was developed to help the clinicians in identifying IUM.[30] MAI consists of 10 questions, each focusing on a different aspect of the appropriateness of medications (e.g. indication, effectiveness, dose, route, etc. as shown in **Figure 3**). Experts respond to each question based on their own judgment, and the tool does not specify which indications, doses, routes, etc. are acceptable or unacceptable for each medication.

Explicit criteria more objectively specify the conditions that need to be met before the decision about appropriateness can be made. Each explicit criterion can contain specifications for the medication, the indications or contraindications, and other contextual information that affect the decision on appropriateness of the medication. Collections of explicit criteria for identifying IUM have been developed and verified by different groups of researchers, and examples include the Assessing Care Of the Vulnerable Elder (ACOVE) project, the Beer's criteria, the Screening Tool of Older Person's Prescriptions (STOPP), and the algorithms developed by C. Alice Oborne

and her colleauges.[32,39,42,43] A listing of all explicit criteria for medication use that were found in this review is provided in **Appendix 1**.

Previous studies have suggested that although implicit criteria are simpler to develop, their results are harder to reproduce, and may suffer from lower validity or reliability. Explicit criteria provide higher validity and reproducibility, but they are difficult to develop and are not readily available for many cases.[44,45]

Reducing IUM

Methods that have been employed to reduce IUM span from focusing on healthcare professionals and patients, to financial, organizational, and regulatory changes. After the thematic analysis reached saturation, six major categories were identified. **Figure 4** shows how these themes align with the classifications found in previously published review articles in this field. It is notable that the use of informatics methods is described as a separate method in the newer review articles. In the following subsections, we describe each theme separately and provide a qualitative summary of the studies associated with that theme.

Feedback to Providers

The promise of this approach is to discourage IUM by providing feedback to the prescribing providers about the prescriptions that they make. Two main forms of feedback are generally used: *passive feedback* occurs when the feedback is not at the point of care (e.g. mailed feedback) or general feedback is provided about IUM (i.e. feedback is not specific to a particular prescription or medication), while *active feedback* occurs when feedback is specific and takes place at the time of care (e.g. face-to-face or through a phone call).

	Current Study	Arnold '05	Sketris '09	Georgiou '11	Clyne '12
Feedback to Providers	✓	✓	✓	✓	✓
Provider Profiling	✓	✗	✓	✗	✗
Indication-based Prescribing	✓	✗	✗	✓	✗
Educating Providers	✓	✓	✓	✗	✗
Appropriate Use Guidelines	✓	✓	✓	✓	✗
Informatics Solutions	✓	✗	✗	✓	✓

Figure 4 – Methods used to reduce IUM.

While passive feedback was effective in reducing overuse in some studies[46–50] in other studies such impact was either not observed or was only short-term.[51–55] In contrast, studies that used active feedback almost always resulted in improvements of outcomes.[56–66] It should be noted, however, that many of these studies also used some of the subsequently described approaches (e.g. appropriate use guideline, etc.) to generate the feedback, so their effect may be a results of the bundled approach rather than just the use of active feedback.[56–58,61,62,67]

Provider Profiling

Also called *physician profiling* or *physician audit*, this method is focused on developing (usually graphical) profiles of the utilization of a certain resource by a certain provider, and comparing each provider with a predefined norm. The promise of this approach is that providers that are utilizing the resource more than the expected norm may be likely to change their behavior given this knowledge. These *profiles* are typically used as part of feedback to the providers.[18,67,68] The effect of profiling has also been studied in combination with other methods such as provider education[69,70] or appropriate use guidelines[71,72] as well. While the majority of these studies reported a positive impact, one of the earliest studies signified that the effectiveness of this approach highly depends on obtained the providers' "buy in".[73]

This approach is associated with two challenges. The first challenge is to define the norm with which providers are compared. This norm may differ among various disciplines (e.g. the norm for prescribing contraceptives by a cardiologist differs format that of a family physician), as well as within disciplines (e.g. the norm for prescribing contraceptives in a family physician focused on outpatient care differs from that of a similar physician focused on acute care). In most cases, such norms are not previously defined, and deriving them from past data is subject to various types of bias. Most of the existing studies define this norm using process outcomes such as total number of prescription[69,70,74] or cost of care.[17] This results in a second challenge: provider profiling promotes reduction in the *overall* utilization of a resource, and does not necessarily focus on *inappropriate* utilization alone. Therefore, this approach may reduce the quality of care if appropriate uses are inadvertently discouraged.

Appropriate Use Guidelines

Appropriate use guidelines are the most organized form of explicit criteria for identifying IUM. Several studies have utilized these guidelines, in combination with other approaches, to enforce appropriate use of medications. While earlier studies suggested that enforcing guidelines alone may not provide enough incentives for providers to change their behavior,[75] subsequent studies that combined guidelines with various methods of feedback [56,57,59,60,64] and provider profiling [22,71,72] unanimously showed a positive impact.

The biggest challenge for this approach is that evidence-based appropriate use guidelines (and explicit criteria, as a whole) are only available for certain medication and patient populations, therefore the results of existing studies may not be generalizable. Additionally, these studies relied on manual implementation of the guidelines which is resource-intensive, and the cost-benefit tradeoff was never studied.

Indication-based Prescribing

The promise of this approach is that by requiring the providers to include an appropriate reason (indication) for each medication order at the time of prescription, unnecessary prescriptions will be discouraged and reduced. This method, which is usually implemented within the order form or the computerized provider order entry (CPOE) system, has been associated with mixed results. While some studies show this approach can capture the 'indications' with high accuracy and therefore reduce IUM,[76,77] others have shown lower accuracy in capturing indications and no impact on IUM.[78–81]

Although this approach may be less difficult to automate in CPOE systems, there are major challenges in its implementation. First, using mandatory indication fields can transform the problem into one of 'over-diagnosis'; in other words, providers are more likely to complete the indication field inaccurately, to bypass the obtrusive mechanism that prevents ordering without a suitable indication. This is reflected by the low accuracy of the indications provided by prescribers when compared to a manual chart review.[80] Using unobtrusive methods of collecting indications at the time of prescription avoids the previous problem but may hinder the possibility of accurately identifying IUM, especially when the indication data is not available in a computable form. Off-label use of medications introduces major challenges in using indication-based prescribing as well. Off-label use often includes using a drug for a condition that is not among its approved indications, and while this can constitute appropriate use (based on the trade-off of harms and benefits for the particular patient), using a hard-coded list of indications may lead to identifying off-label use as a form of inappropriate use (or conversely, considering a large list of off-label indications as acceptable indications may lead to low sensitivity in identifying

IUM). Due to these challenges and the scarcity of robust controlled studies, the effect of indication-based prescribing (particularly using unobtrusive approaches) is not known.

Provider Education

Numerous studies have evaluated the effect of educating providers on reducing IUM. Multiple mediums have been used for education (e.g. mail-in fliers, lectures during rounds, seminars, etc.) and some studies also evaluated multi-faceted approaches, for example by combining education with feedback regarding providers' past prescriptions, profiling, and enforcement of guidelines. Many studies used the term *academic detailing*, to describe the educational outreach for the purpose of improving decision making regarding appropriate use of medications.[82]

Overall, the majority of these studies reported a reduction in IUM after provider eduation.[13,58,61,69,70,83–88] One study also showed that adding provider education (in the form of either a seminar or academic detailing sessions) to other interventions, namely audit and feedback, resulted in a significantly larger and more sustainable reduction in IUM.[85] Nevertheless, there are very few studies that compare different education approaches or make recommendations about the choice of education medium.

Informatics Solutions

Automated solutions have been tried to varying degrees in studies aimed at reducing IUM. Some of the aforementioned approaches have been implemented in CPOE systems (examples indication-based prescribing, and provider feedback in the form of electronic alerts), and other methods rely on the data obtained from EHR systems (e.g. provider profiling using EHR log data).[89–98] Assuming that drug-drug interactions (DDI) can also be considered a predisposing factor for IUM (specifically, making a medication "inappropriate by population"), decision

support systems focused on DDI can also be considered as informatics solutions for a specific subset of IUM. Literature in DDI checking by clinical decision support is relatively rich, but many studies have shown that the majority of DDI decision support that is delivered as electronic alerts is overridden by the clinicians, the override rate can be as high as 90%,[99,100] and the majority of these overrides were found to be appropriate.[101] Various reasons for the high override rate has been proposed (including high false positive rate of the alerts, issues with delivery of the information, etc.) and recommendations have been made on how to improve DDI alerts and mitigate these issues.[102] It should be noted that the majority of existing literature does not consider DDI a form of IUM.

We did not identify any study that compared automated approaches with non-automated alternatives. On the other hand, electronic feedback to the providers, for example through mandatory indication fields or alerts, can potentially lead to alert fatigue.[103] Additionally, relying on EHR data for automated identification of IUM can introduce challenges with the availability and accuracy of the data, a challenge that is well-known in clinical informatics.[104,105]

Discussion

There are many factors that contribute to the problem of IUM, and as a result, identifying and reducing overuse requires multiple initiatives. This review indicates the lack of a generic method for defining, identifying and measuring IUM, and scarcity of knowledge on how to model IUM in a computable form.

We identified two general types of approaches to identifying IUM. The implicit approach cannot be automated because it relies on the judgement of an expert. Explicit solutions have the potential for automation, however structured versions of these criteria are not available to the

public. This review signified that further research is needed to develop a central repository of the explicit criteria, both in narrative and in structured form. Ideally, all components of the MAC should be described using standardized nomenclature.

We identified six major themes in methods used to reduce IUM. The themes identified in this review align well with those identified by previous published reviews on related topics. Use of informatics solutions is a relatively emerging theme and has been noted in two of the more recent reviews in this domain as well. While each of the six methods used to reduce IUM have been associated with varying levels of success, overall the evidence suggests that bundled approaches involving more than one methods are associated with higher levels of success.

For some of the methods described above, robust, controlled studies that would directly evaluate the outcomes or compare different approaches are scarce or do not exist. Consequently, it is difficult to draw generalizable conclusions about the factors that lead to success or failure of each approach. Moreover, the studied outcome in many of the existing studies constitutes only process measures such as rate of prescription or cost of care, and studies targeting patient outcomes and quality of care are needed.

Limitations

This study is not without limitations. We used a rapid review methodology, i.e. the search scope was limited to PubMed and excluded studies that were not in English; therefore, we acknowledge the possibility of publication and language bias. However, the results of our thematic analysis are in alignment with previous published reviews, some of which are systematic reviews; this suggests that the potential for those biases was minimal.

Appendix 1

In this appendix, a listing of all explicit medication appropriateness criteria that were found in the review of literature is provided.

Citation	Number of criteria	Medication	Population
Ahrens *et al* [41]	21	PPIs	All
Beers Criteria [106]	57	Many	Elderly
Bashford *et al* [81]	9	PPIs	All
Batty *et al* [107]	1	Benzodiazepines	All
Batuwitage *et al* [108]	6	PPIs	All
Bez *et al* [109]	23	PPIs	All
Choudhry *et al* [110]	13	PPIs	All
Craig *et al* [2]	1	PPIs	All
Eid *et al* [5]	19	PPIs	All
Elliott *et al* [59]	1	Antithrombotic Drugs	All
Issa *et al* [4]	8	PPIs	All
Khalili *et al* [86]	14	PPIs	All
Larson *et al* [111]	17	Vancomycin	All
Oborne *et al* [107]	1	ACE Inhibitors	Elderly
Oborne *et al* [112]	1	Neuroleptics	Elderly
Oborne *et al* [40]	1	Amiodarone	All
Oborne *et al* [40]	1	Aspirin	All
Oborne *et al* [40]	1	Steroids	All
Parente *et al* [3]	6	PPIs	All
Piallans *et al* [113]	14	PPIs	All
Williams *et al* [114]	2	NSAIDs, ESAs	All
START [115]	22	Many	Elderly
STOPP [116]	64	Many	Elderly

Acronyms:

ACE = Angiotensin Converting Enzyme

ESAs = Erythropoiesis Stimulating Agents

NSAIDs = Non-steroidal Anti-inflammatory Drugs

PPIs = Proton Pump Inhibitors

START = Screening Tool to Alert doctors to the Right Treatment

STOPP = Screening Tool for Older Person's Prescriptions

Conclusions

Root causes of overuse are diverse, therefore diverse approaches have also been used to identify and reduce overuse. Studies combining multiple approaches are more frequently associated with successful outcomes. Nevertheless, more rigorous studies are needed to evaluate the impact of existing approaches, and a more objective representation of appropriate use is necessary.

Chapter 3: Developing a conceptual model for appropriateness criteria

Background

Inappropriate medication use can be broadly described by the following categories: (a) inappropriate for a *specific clinical indication*, or used in the absence of any appropriate indication, (b) inappropriate for clinical indication in a *specific population*, or (c) excessive *duration or frequency*, or excessively aggressive *route of administration*, given expected clinical benefit.[16]

The first category describes the situation where the indication provided for the medication is not supported by scientific evidence; an example is the use of antibiotics for acute viral bronchitis.[1] The second category occurs when a medication that is appropriate for one population is used for another population (particularly a low-risk or high-risk population) in which case its harms outweigh its benefits; an example is the use of antidepressants in patients with only mild levels of depressive disorders.[38,117] The last category describes the use of a medication in larger doses, for longer periods, or through more aggressive routes than necessary; examples include the use of intravenous medication when an oral counterpart can be used and has the same effectiveness,[2] or the use of antibiotics for longer periods than necessary.[118] The first category is commonly titled "medication overuse", while the last category is frequently referred to as "medication misuse".

It should be noted that these separate categories do not exist in isolation. For example, many patients who are subject to the overuse of PPIs, start receiving PPIs in the presence of an appropriate indication, but continue to receive the medication after that indication has been

resolved.[41] While the problem of inappropriate use of PPIs is generally labeled as "overuse" because of the use of PPIs in the absence of an indication,[5,10,119] it can also be described as excessive duration of the treatment. The most well-known definition for "overuse", which is provided by the Institute of Medicine, avoids this problem by combining the three categories and defining overuse as any situation when "a health care service is provided under circumstances in which its potential for harm exceeds the possible benefit".[120]

Estimating the harm and benefit of prescribing a medication for an individual patient is a fundamental part of the practice of medicine. Spinewine *et al* emphasize that multiple values may need to be incorporated in this decision, including "scientific value, technical rationalism and the general good".[45] These abstract values are hard to quantify and typically require customization, rendering the decision about appropriateness complex. Evidence-based guidelines can assist the clinician in making an informed decision, but they are usually based on population averages, and need to be individualized for each patient.[121] However, it is possible to identify the patients who are potentially at risk for IUM using the explicit criteria that define these at risk populations. Although these criteria only serve as guidelines and certain exceptions may exist, the ability to identify these at risk populations in an automated way is a crucial step in tackling the problem of IUM.

One approach for identifying IUM is to identify patients who do not have any indications to receive a given medication. While we have shown that this approach can accurately identify overuse at least in the case of proton pump inhibitors,[37] this approach is only scalable if complete knowledge-bases (KBs) for medication indications are available. The first study in this chapter evaluates the existing publication KBs for medication indications, while the second study

discusses forms of IUM which may not be captured only using medication-indication KBs and provides a conceptual framework for them.

A conceptual model for appropriateness criteria should specify all the elements that are required to explain each of the different categories of IUM and the different concepts that are needed to quantify the appropriateness of medications. These concepts include but are not limited to: disease characteristics (which define the indications and their severity, as well as contraindications), patient characteristics (such as age and gender), concurrent medications (which provide the possibility for drug-drug interactions), characteristics of the treatment (such as its appropriate indications, and the definition of the intended population) and the dose, route and duration of medication use.

Many of the published studies define and measure appropriateness of medications using implicit criteria, where the appropriateness is subjectively assessed by the expert reviewers.[122,123] This impedes the possibility of designing an automated method for identifying IUM. Even in studies where explicit criteria are used to define appropriateness, these criteria usually do not include every detail about the criteria, as described above.[5,28,60,124] In their review of the quantifiable metrics for measuring IUM, Chan *et al* conclude that the future of research on IUM relies on translating appropriateness criteria into measurable metrics.[16] This further signifies the necessity of an objective representation for explicit appropriateness criteria.

Study 2: Medication-indication knowledge-bases: a review and critical appraisal

Introduction

With the wide adoption of electronic health records (EHR) there has been tremendous focus on developing automated solutions focused on the appropriate use of medications. Computerized clinical decision support (CDS) has been successfully used to prevent medication errors and adverse effects.[125] Designing CDS solutions to promote appropriate medication use is faced by two challenges: to determine the reason for medication prescription, and to represent this information. Medications are prescribed to treat or prevent different signs, symptoms, diseases or conditions, which are collectively known as *indications*.[126] While many indications are relatively simple to represent (e.g. *insulin* is indicated in patients with *type I diabetes mellitus*), some are more complex (e.g. *bismuth subsalicylate* is indicated in the treatment of *Helicobacter pylori infection* only when combined with antibiotics). At least one indication is listed on the drug label for each medication (*on-label* indications), but many medications are also used to treat or prevent conditions that are not explicitly listed on the drug label (*off-label* indications). It is estimated that 21% of medication prescriptions are for off-label use[127,128] and this rate is estimated to be greater in the pediatric population.[129,130]

Linking medications to their indications is essential to providing effective care. It has been shown that treatment outcome and healthcare quality may improve once such links are made, either manually or electronically.[131,132] Medication-indication information is a necessary part of the information that is needed to determine appropriate use of medications. Medications can be deemed inappropriate for several other reasons such as allergies, drug interactions, side effects,

etc. However, it is essential to establish that a medication is being administered for an appropriate indication, to justify that there is at minimum a need for that drug. This depends on the availability of a comprehensive, structured KB of medications and their indications. Such a KB can be used to support order entry, for example through the use of indication-based prescribing.[133] It can also be used to identify cases of potential overuse of medications.[37]

While public resources containing medication-indication data have been available for a long time (including MedicineNet[134] since 1996, DrugBank[135] since 2006 and DailyMed[136] since 2008), these resources only provide the indications in unstructured form (i.e. free-text). Several vendors have also developed proprietary datasets that include medication-indication data (examples include MediSpan®, Epocrates® and Lexicomp®) but these products are impractical to use for research due to associated costs and licensing restrictions. Consequently, there have been a number of efforts towards developing comprehensive KBs of on-label and off-label indications of medications for public use.

We aimed to identify studies discussing the design and evaluation of medication-indication KBs, determine the similarities and differences of approaches used in these studies, and identify potential knowledge gaps. We accomplished this by conducting a systematic review of literature, and subsequently evaluating whether these KBs support the level of complexity necessary to comprehensively represent medication-indication information.

Methods

Systematic Review

We searched the titles and abstracts of all articles cited by PubMed, PubMed Central, EMBASE and CINAHL through September 30, 2014 using the following search strategy: (medication$ OR

drug$) AND (indication$ OR "off-label") AND (resource$ OR knowledge$). The search strategy used '$' as the wildcard character, and it was modified according to the specifications of each bibliographic database accordingly.

We used the following inclusion and exclusion criteria for study selection: (i) studies without an English abstract were excluded (we did not impose any language restrictions to the full text of the article); (ii) only studies were included that described the design and/or evaluation of a medication-indication KB, terminology or database; (iii) only studies where the resource was publicly available were included.

Titles and abstracts were reviewed independently by two authors (HS and CF), and disagreements were resolved using a third author (HC) as an arbitrator. We then obtained the full texts of all articles that were deemed as potentially relevant, and excluded articles that did not meet the above criteria. For each of the articles that were included, we also reviewed the reference list to identify any other relevant studies.

Once the final list of included studies was populated, we extracted the following information from each article: name of resource discussed, methods used for developing/evaluating the resource, and the results of such evaluations. We grouped all studies that were about the same resource and analyzed them together.

Evaluation of the KBs

We obtained a copy of each resource identified in the previous step, and collected basic descriptive information, including the number and nature of entries, as well as representational characteristics and scope of the knowledge included in the KB. We compared the scope of the KBs in three different ways: based on the number of medications included, number of

indications included, and number of medication-indication pairs included. While some KBs defined the medications at the level of main ingredient (IN), others did at a finer level of granularity (i.e. specific dose forms). To be able to compare these KBs, we normalized them to the same level by mapping every medication concept to the main ingredient using RxNorm's Application Programming Interface (RxNorm API).[137] Details of this process are explained in

Appendix 1. Normalizing the indications was difficult not only because the KBs represented the indications in different forms (including UMLS concepts and free-text) but also because closely related indications which were coded (or worded) slightly differently may or may not need to be combined depending on the application. The choice of how to normalize the indications (e.g. using which natural language processing engine, which terminology, and so on) could introduce significant errors and bias into our analysis. Therefore, to bring all KBs to the same form, we stored all indications in free-text form and only aggregated exact matches. Further, we qualitatively evaluated whether the KB supports representation of complex indication information. In order to accomplish the latter, we asked a group of pharmacists and clinicians to provide us with a list of complex or challenging indications for commonly prescribed medications. We then organized these indications based on the type of complexity and identified eight different types of complexities; four corresponded to characteristics of the medication and four were associated with characteristics of the indication (**Table 1**). Finally, we evaluated whether the KBs could represent these complexities.

Results

Our electronic searches retrieved 3791 documents, of which 968 were duplicates (**Figure 5**). From the remainder, 2459 were excluded after reviewing the titles and abstracts, and 42 were marked as potentially relevant. By reviewing the full texts, eight articles were deemed relevant.[138–145] Citation tracking yielded an additional seven relevant articles.[146–151] After applying the inclusion and exclusion criteria and consolidating multiple articles that were about the same resource, seven public medication-indication KBs were identified and included in this review (**Table 2**).

We proceeded by mapping all medication concepts to RxNorm at the main ingredient level as described above. The results of this process are explained in

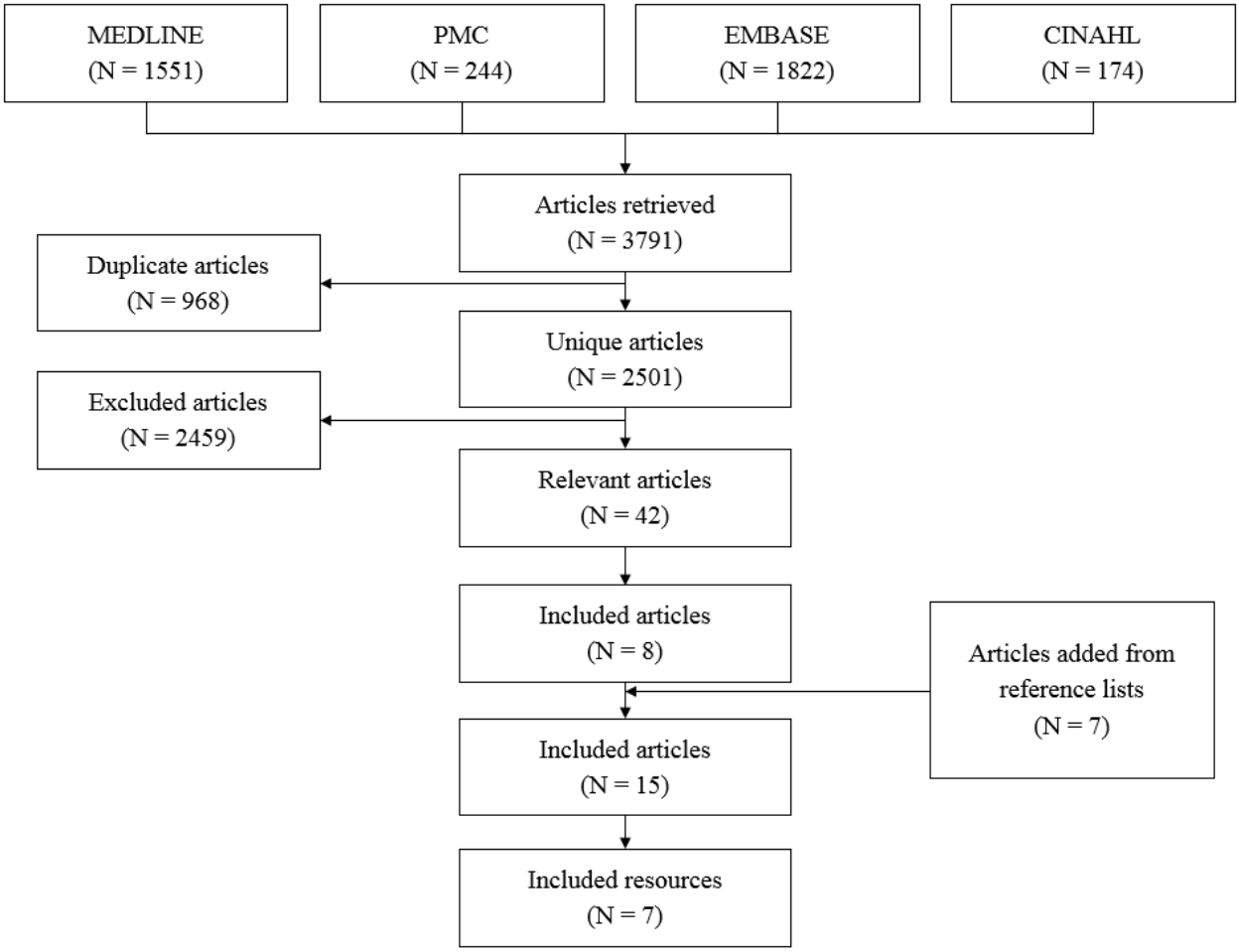


Figure 5 – Study selection process.

Appendix 1. In short, we identified the main ingredient for more than 80% of the unique medications in the KBs; in two KBs this number was 100% while in another two it was more than 98%. Note that our approach only identified one main ingredient for each medication concept, which would be incomplete for combination drugs. Only one resource (SIDER[139]) clearly marked combination drugs; it contained 4,567 unique structured product labels (SPLs) for combination drugs, and these SPLs represented 1,672 unique combination drugs (based on the ingredients). Other resources do include combination drugs, but identifying them requires using external knowledge.

After the mapping, 9,134 unique medications and 7,362 unique indications were found in the union of all KBs, corresponding to 110,104 unique medication-indication pairs (see **Table 3)**. The number of medications, indications, and medication-indication pairs that appeared in all 7 resources was only 82, 32, and 1, respectively (**Figure 6**). This figure shows that the majority of medications, indications and medication-indication pairs are specified only in exactly one KB (i.e. they are unique to the KB in which they are defined).

Figure 7, Figure 8 and **Figure** 9 show the overlap between KBs in terms of unique medications, indications and medication-indication pairs, respectively. The dome-shaped chords start and end in the same sector, and represent the entities that do not appear in any of the subsequent KBs. The relative size of the dome-shaped chords to the sector provides a rough estimate of the amount of information in each KB that is not shared with the subsequent KBs. Bigger domes in the bottom figure indicate that most of the medication-indication pairs are unique to the KBs in which they are defined. Further explanation about how to interpret the chord diagrams is provided in **Appendix 2**, and the overlap between the KBs is numerically summarized in **Table 4**, **Table** 5 and **Table** 6.

	Complexity	Example	Description of the example
Characteristics of the medication	Dose forms	Vancomycin for *C. difficile* infection	Vancomycin is used intravenously for all of its indications, with the exception of recurrent *C. difficile* infection, in which case it is used orally. Therefore, the KB should associate the indication with the correct dose form of the medication.
	Route	Heparin for venous thrombosis	The route of administration of heparin is different when it is used for prevention of venous thrombosis (subcutaneously) versus for treatment of venous thrombosis (intravenously).
	Strength	Heparin for prevention and treatment of venous thrombosis	The dosage of heparin used for prevention of venous thrombosis is different (i.e. lower) that dosage used for treatment of venous thrombosis.
	Duration	Cystitis versus recurrent cystitis	The appropriate duration of antibiotic therapy for initial episode of cystitis (3 days) is different from the appropriate duration of therapy for treating recurrent cystitis (5-7 days).
Characteristics of the indication	Primary choice	First line treatment for essential hypertension	Thiazide diuretics, calcium channel blockers, and angiotensin converting enzyme inhibitors are the preferred medication as the initial therapy for essential hypertension. Other medications that lower blood pressure (such as direct vasodilators or alpha agonists) are not indicated as the first line therapy.
	Comorbidities	Heart failure and asthma	The indicated therapy for heart failure in patients with asthma is different than those without asthma, because non-selective beta-blockers may cause bronchospasm and aggravate asthma.
	Prevention vs. treatment	Aspirin for prevention of cardiovascular risk	Aspirin is indicated in adults with risk factors for cardiovascular disease (CVD). The actual indication is preventative, and aspirin is prescribed in the context of other diseases that are risk factors for CVD, such as hypertension.
	Co-medication	Bismuth for eradication of *Helicobacter pylori*	Bismuth compounds are only used for eradication of *H. pylori* in conjunction with other medications, including two or more antibiotics. Bismuth is not indicated as monotherapy for *H. pylori* eradication

Table 1 – Complex or challenging indications used to evaluate the medication-indication KBs.

	Scope		Curation [a]	Normalization		Sources				
	On-label	Off-label		Medications	Indications	Medline	SPL [e]	Crowd	Multiple	EHR
NDF-RT[138]	Yes	Yes	SA	UMLS [b]	UMLS [b]	■				
SIDER[152]	Yes	No	A	SPL ID	COSTART		■			
McCoy *et al.*[140]	Yes	Yes	SA	None [c]	None [c]			■		
Fung *et al.*[141]	Yes	No	A	RxNorm	UMLS		■			
MEDI[142]	Yes	Yes	A	RxNorm	ICD9				■	
Jung *et al.*[143]	No	Yes	A	None [d]	None [d]					■
LabeledIn[145]	Yes	No	SA	UMLS	UMLS		■			

Table 2 – Description of the medication-indication KBs that are freely available to the public.

[a] Curation method is categorized as automated (A) or semi-automated (SA)

[b] Although NDF-RT itself uses its internal concept identifiers for the medications and indications, the version available through UMLS normalizes these concepts to UMLS concepts.

[c] McCoy *et al.* used a numeric identifier for the medications and indications, but did not explain which coding system was used.

[d] Although Jung *et al.* normalized the concepts to UMLS during the curation of their resource, the final resource only contains the free-text labels for medications and indications.

[e] SPL: structured product labels.

	Medications		Indications	Medication-Indication Pairs	Indications per Medication		
	Unique	Main Ingredient			Range	Mean (SD) [a]	Median
NDF-RT[138]	9,579	1,625	1,010	55,704	1 – 221	5.82 (7.74)	4
SIDER[152]	18,334	1,307	2,194	165,920	1 – 129	8.93 (12.45)	19
McCoy *et al.*[140]	2,537	2,230	1,580	11,166	1 – 122	4.40 (7.08)	6
1,181	2,104	1,181	2,885	19,473	1 – 130	9.26 (12.56)	6
MEDI[142]	2,897	2,162	4,352	65,535	1 – 314	22.02 (24.52)	3
Jung *et al.*[143]	214	212	173	407	1 – 8	1.90 (1.31)	1
LabeledIn[145]	2,376	251	626	19,599	1 – 144	8.29 (1.76)	3

Table 3 – Number of unique concepts in medication-indication KBs.

[a] SD: standard deviation

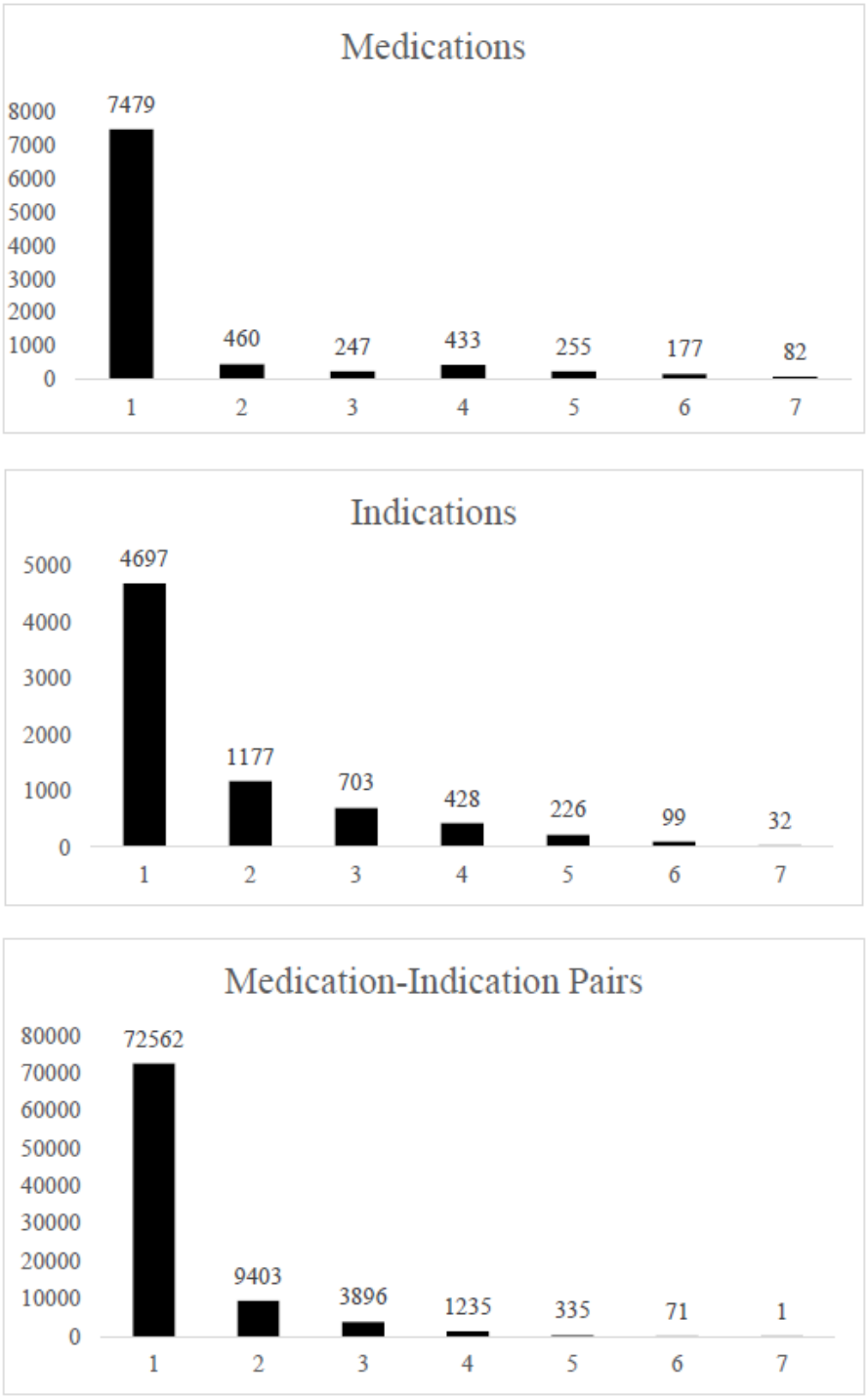


Figure 6 – Descriptive statistics about medication-indication knowledgebases.

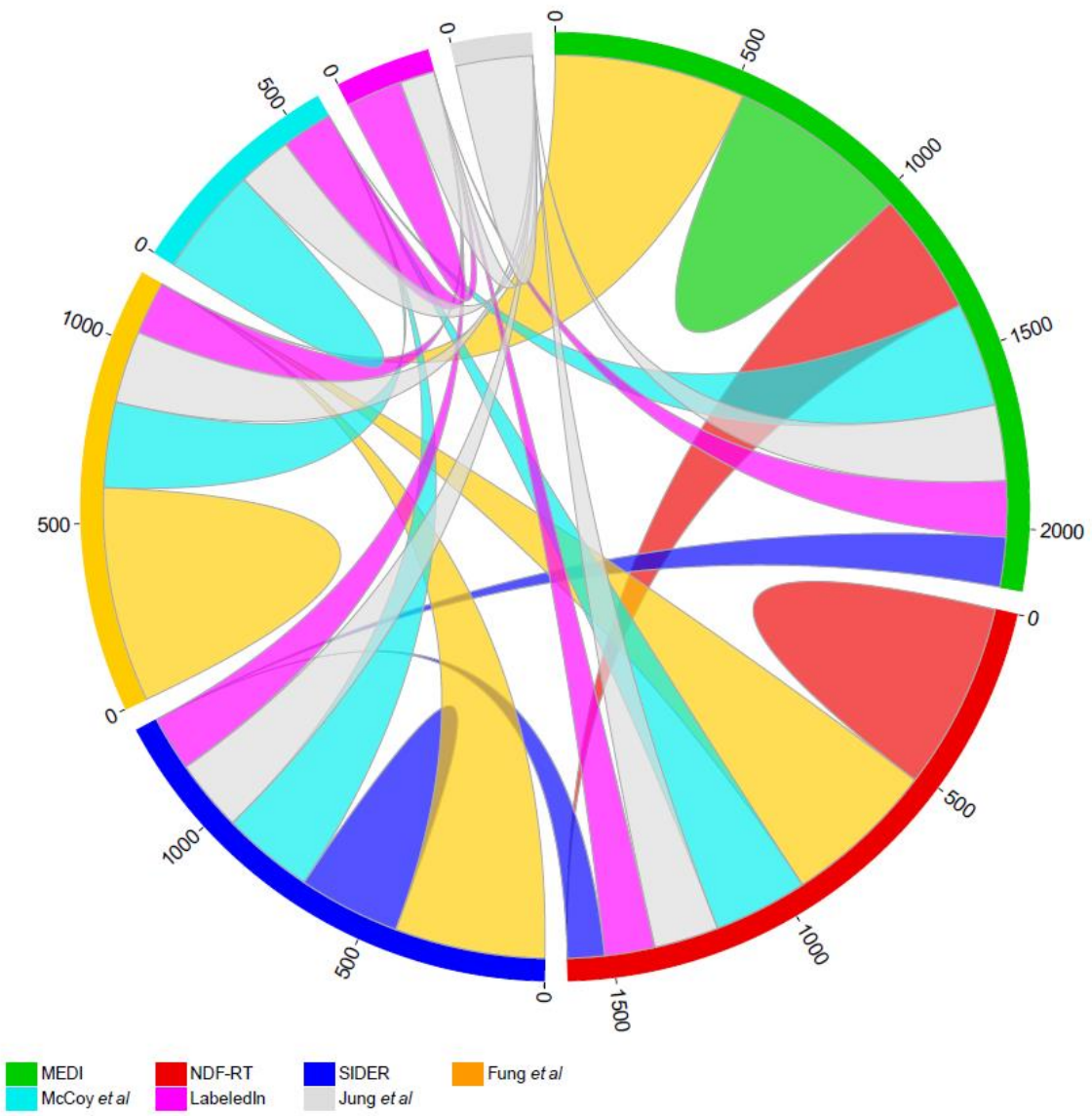


Figure 7 – Overlap between medication-indication KBs (drugs).

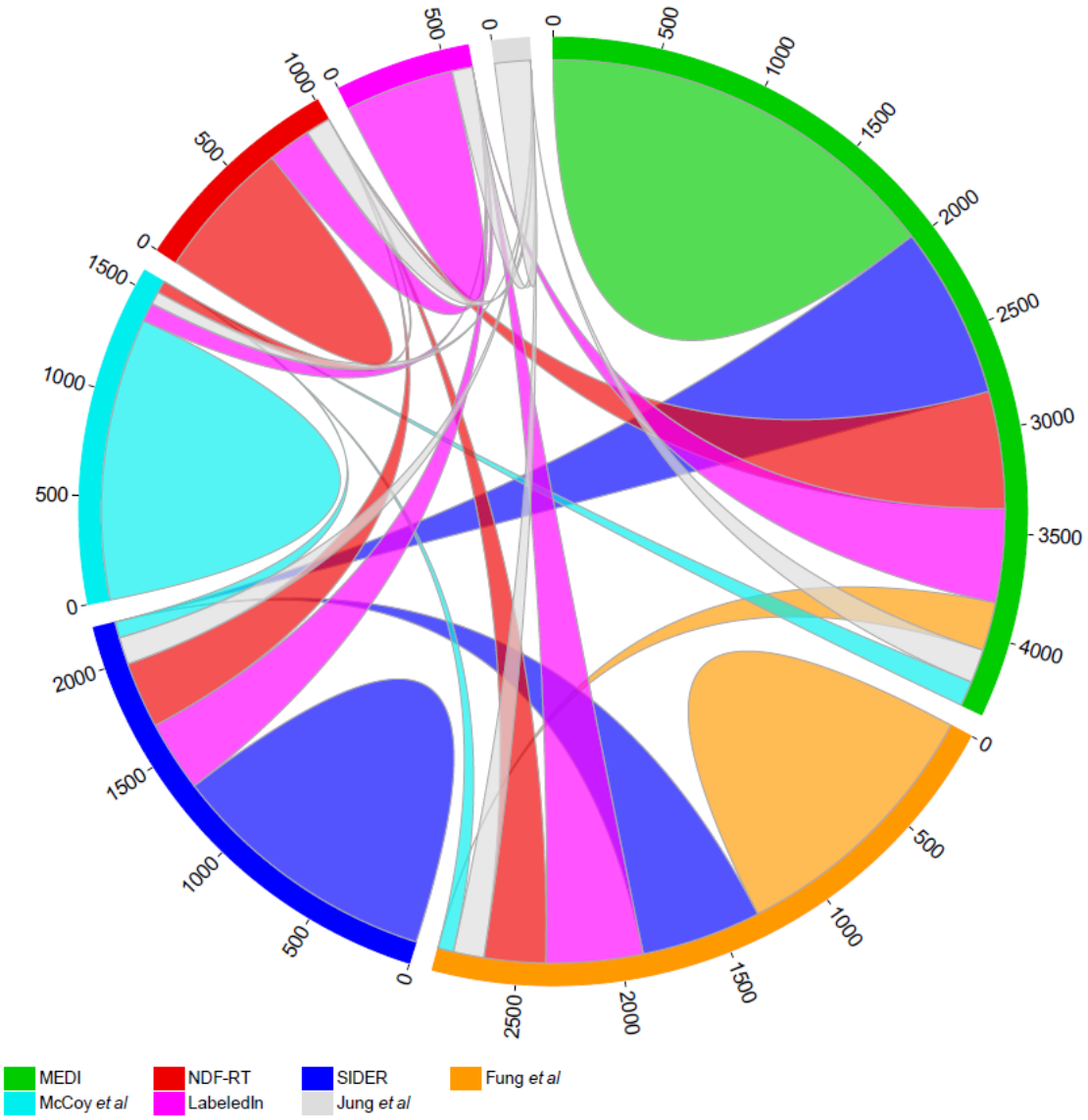


Figure 8 – Overlap between medication-indication KBs (indications).

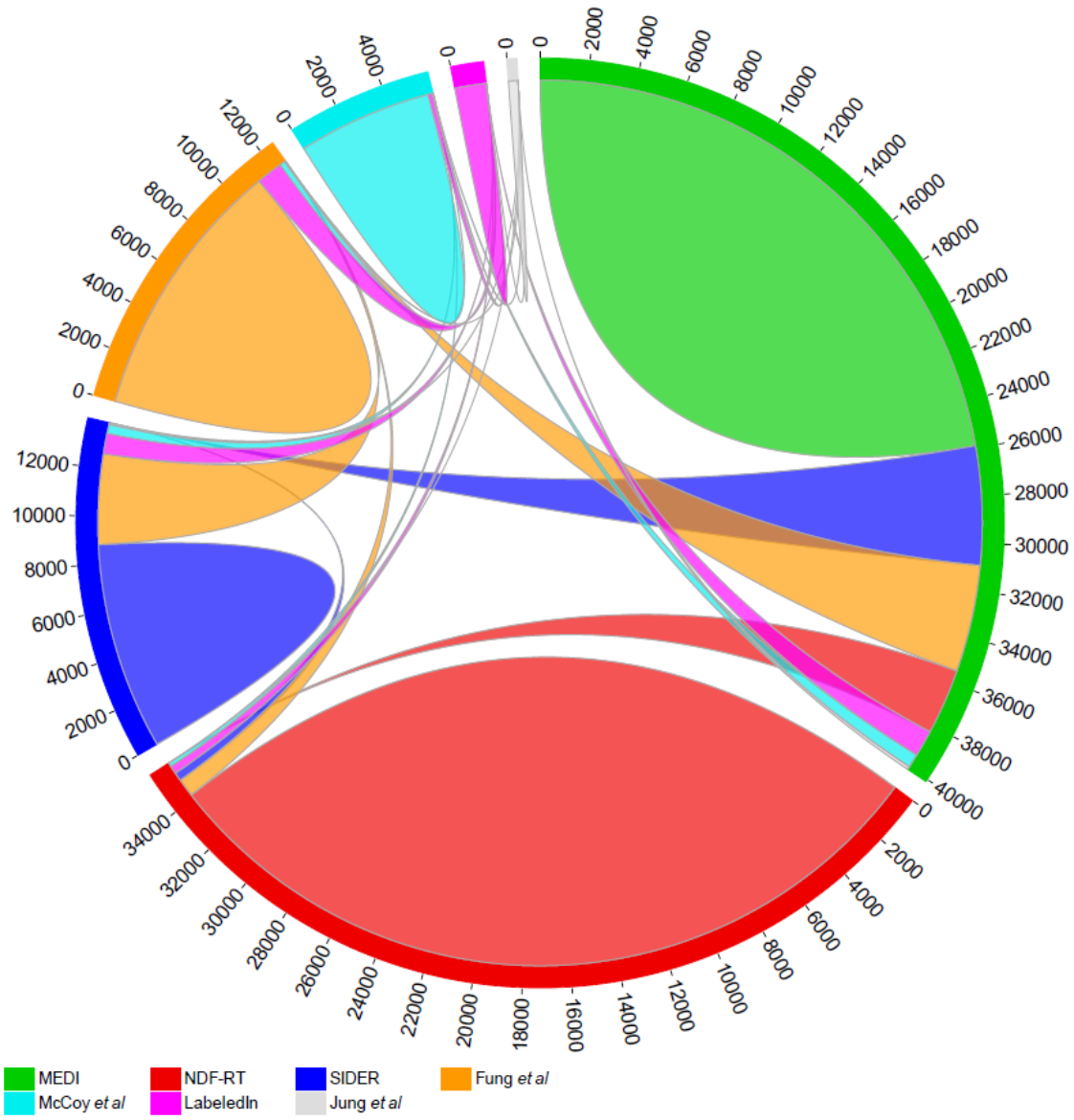


Figure 9 – Overlap between medication-indication KBs (medication-indication pairs).

	MEDI	NDF-RT	SIDER	Fung *et al*	McCoy *et al*	LabeledIn	Jung *et al*
MEDI	2162	1410	1152	1114	580	251	210
NDF-RT	1410	1625	984	934	521	227	179
SIDER	1152	984	1307	979	541	247	199
Fung *et al*	1114	934	979	1181	520	243	200
McCoy *et al*	580	521	541	520	594	232	150
LabeledIn	251	227	247	243	232	251	96
Jung *et al*	210	179	199	200	150	96	212

Table 4 – Overlap between medications included in the KBs (drugs).

	MEDI	Fung *et al*	SIDER	McCoy *et al*	NDF-RT	LabeledIn	Jung *et al*
MEDI	4352	1530	1666	324	887	550	162
Fung *et al*	1530	2885	1309	253	605	565	148
SIDER	1666	1309	2194	263	597	427	132
McCoy *et al*	324	253	263	1580	162	130	61
NDF-RT	887	605	597	162	1010	281	120
LabeledIn	550	565	427	130	281	626	99
Jung *et al*	162	148	132	61	120	99	173

Table 5 – Overlap between medications included in the KBs (indications).

	MEDI	NDF-RT	SIDER	Fung *et al*	McCoy *et al*	LabeledIn	Jung *et al*
MEDI	40096	4388	9653	5756	783	1118	128
NDF-RT	4388	35787	1260	1151	249	333	7
SIDER	9653	1260	13947	487	560	861	71
Fung *et al*	5756	1151	4827	12591	409	1115	41
McCoy *et al*	783	249	560	409	5931	184	36
LabeledIn	1118	333	861	1115	184	1348	7
Jung *et al*	128	7	71	41	36	7	404

Table 6 – Overlap between medications included in the KBs (medication-indication pairs).

To signify the issues with granularity of indications, we also visualized the indications for one randomly selected drug, *Budesonide*, which is a glucocorticoid commonly used for the treatment of asthma and other reactive airway diseases. This medication was randomly picked from the list of 82 medications that have at least one indication in every KB. Almost every KB lists asthma as an indication for *Budesonide* but only some resources include more granular terms such as *intermittent asthma* or *chronic obstructive asthma*, or less granular terms such as *reactive airway disease*. Similarly, while one KB lists *sinusitis* as an indication, another resource only lists more granular terms (i.e. *chronic maxillary sinusitis, chronic frontal sinusitis*¸ etc.) as indications. This signifies why normalizing and aggregating the indications can be challenging. **Figure 10** visualizes the indications for this drug and the overlap between the KBs, and **Table 7** lists the indications and specifies the KBs in which they appeared.

Below, we will first describe the resources based on the review of the literature, and then we will evaluate the content and structure of the resources as well as their strengths and weaknesses in representing complexities of medication indications. KBs are described in chronological order, based on the earliest publication found for each resource.

Review of the literature

The first public medication-indication KB discussed in the publications is the National Drug File – Reference Terminology (NDF-RT), which is developed by the United States Department of Veteran Affairs (VA).[138] The medication-indication data in NDF-RT was initialized using a list of diseases and drugs frequently co-occurring in abstracts cited by MEDLINE (the list can be obtained from http://mbr.nlm.nih.gov/MRCOC.shtml), and subsequently screened by experts. Apart from indications, NDF-RT provides additional information about medications, including but not limited to contraindications, interactions with other medications, pharmacokinetics,

physiologic effect and therapeutic class. NDF-RT is now included in the Unified Medical Language System (UMLS)[153], which can facilitate linking NDF-RT concepts to concepts in other medication related terminologies – such as RxNorm[154] – or disease related vocabularies – such as SNOMED CT.[155] NDF-RT has also been incorporated into RxNorm since June 2010.[156]

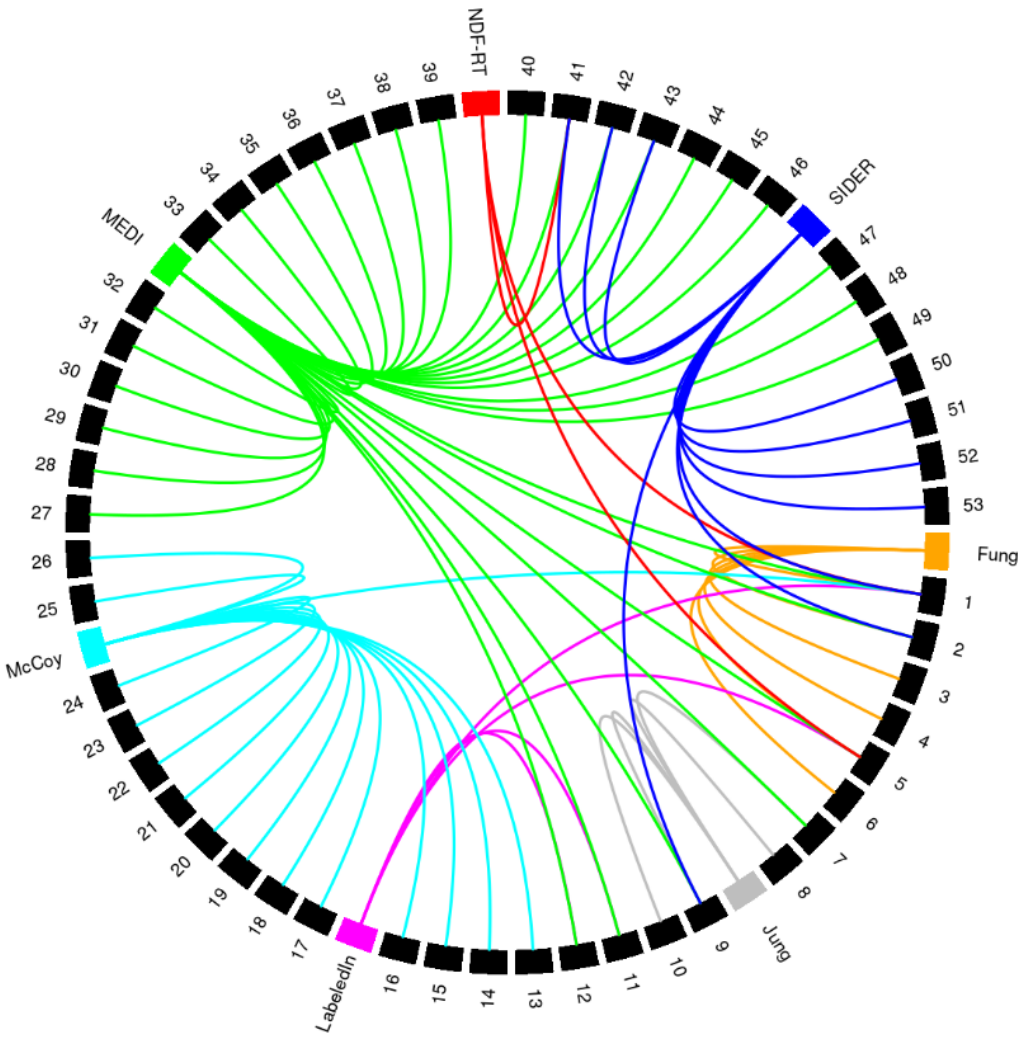


Figure 10 – Overlap between indications listed in the KBs for Budesonide.

	Fung *et al*	Jung *et al*	LabeledIn	McCoy *et al*	MEDI	NDF-RT	SIDER
asthma	1	0	1	1	1	1	1
death (finding)	1	0	0	0	1	0	1
long qt syndrome 2	1	0	0	0	0	0	0
short qt syndrome 2 (disorder)	1	0	0	0	0	0	0
rhinitis, allergic, perennial	1	0	1	0	1	1	0
nose symptoms	1	0	0	0	0	0	0
bronchitis	0	1	0	0	1	0	0
sinusitis	0	1	0	0	0	0	0
bronchial spasm	0	1	0	0	1	0	1
bronchial diseases	0	1	0	0	0	0	0
hay fever	0	0	1	0	1	0	0
crohn disease	0	0	1	0	1	0	0
crohn's disease	0	0	0	1	0	0	0
reactive airway disease	0	0	0	1	0	0	0
chronic obstructive pulmonary disease	0	0	0	1	0	0	0
intermittent asthma	0	0	0	1	0	0	0
difficulty breathing (dyspnea)	0	0	0	1	0	0	0
chronic obstructive asthma	0	0	0	1	0	0	0
emphysema	0	0	0	1	0	0	0
idiopathic pulmonary fibrosis	0	0	0	1	0	0	0
shortness of breath	0	0	0	1	0	0	0
chronic maxillary sinusitis	0	0	0	1	0	0	0
chronic frontal sinusitis	0	0	0	1	0	0	0
chronic ethmoidal sinusitis	0	0	0	1	0	0	0
chronic pansinusitis	0	0	0	1	0	0	0
eosinophilic esophagitis	0	0	0	1	0	0	0
pain	0	0	0	0	1	0	0
allergic rhinitis (disorder)	0	0	0	0	1	0	0
bronchitis, chronic	0	0	0	0	1	0	0
chronic obstructive airway disease	0	0	0	0	1	0	0
lung diseases	0	0	0	0	1	0	0
ulcerative colitis	0	0	0	0	1	0	0
inflammatory bowel diseases	0	0	0	0	1	0	0
symptoms	0	0	0	0	1	0	0
syncope	0	0	0	0	1	0	0
fever	0	0	0	0	1	0	0
swelling	0	0	0	0	1	0	0
body weight decreased	0	0	0	0	1	0	0
dyspnea	0	0	0	0	1	0	0
diarrhea	0	0	0	0	1	0	0
nasal polyps	0	0	0	0	1	1	1
rhinitis	0	0	0	0	1	0	1
rhinitis, vasomotor	0	0	0	0	1	0	1
severe asthma	0	0	0	0	1	0	0
hypersensitivity	0	0	0	0	1	0	0
pulmonary emphysema	0	0	0	0	1	0	0
autopsy	0	0	0	0	1	0	0
swyer-james syndrome	0	0	0	0	1	0	0
asthma attack	0	0	0	0	1	0	0
inflammation	0	0	0	0	0	0	1
polyps	0	0	0	0	0	0	1
cancer remission	0	0	0	0	0	0	1
unresponsive behavior	0	0	0	0	0	0	1

Table 7 – List of indications depicted in Figure 10.

Side Effect Report (SIDER), which was first introduced in 2010,[139] is a KB of medications and their adverse effects which also includes indication data.[152] Medications are represented using their structured product label (SPL) identifiers, and side effects and indications are coded using the Coding Symbols for a Thesaurus of Adverse Reaction Terms (COSTART). Indication information were automatically extracted from the SPLs by performing a simple free-text search in the "indications" section of the SPLs for a list of terms from COSTART which were assigned a semantic type of "Anatomical Abnormality", "Finding" or "Natural Phenomenon or Process" in the UMLS.[157] The method used for identifying these concepts in the SPLs was based on straightforward string matching (e.g. it did not account for negation or other contextual information), and detailed metrics about the accuracy of this method are not available, although a manual review conducted by the authors on a subset of SPLs showed that this method had a sensitivity of 79% in identifying the concepts mentioned in the "Adverse Reactions" section of the SPL (specificity was not reported).

McCoy *et al.* used a crowdsourcing approach to infer medication-problem relationships.[140] They used data from an EHR system where the prescribers were required to connect the medications to one of patient's problems at the time of prescription. The authors used frequently co-occurring concepts to develop a medication-indication KB based on their assumption that frequently recorded medication-problem pairs are likely valid medication-indication pairs. They evaluated the accuracy of the KB using a subset of 100 randomly selected medication-indication pairs, using LexiComp® as the reference standard. They also assessed the impact of two covariates on the accuracy of the medication-indication pairs: patient link (i.e. number of unique patients for whom a specific medication-indication pair was recorded in the EHR), and link ratio (i.e. number of unique patients for whom a specific medication-indication pair was recorded, divided by the

number of unique patients for whom that medication and that problem where recorded [but not necessarily connected in the prescription]). The authors concluded that using any of the following criteria, they can acquire medication-indication pairs that are at least 95% correct: patient link ≥ 10, patient link ≥ 2 and link ratio ≥ 0.2, or patient link ≥ 3 and link ratio ≥ 0.1. These pairs were included in the final KB, and they accounted for 76.47% of all medication-problems found in their EHR.

Fung *et al.* used natural language processing (NLP) to extract drug indication information from SPLs downloaded from the DailyMed website.[141] They configured their approach towards higher sensitivity. After manual evaluation of the results for 300 drugs (corresponding to approximately 3500 medication-indication pairs), the authors concluded that their approach achieved a sensitivity of 95% and specificity of 77% in extracting the indications from drug labels. Primary reasons for errors in specificity included identification of the wrong concept by the NLP system and identifying all disease mentions as indications (including those that are explained in the indication section of the label as comorbidities, exceptions, etc.)

Wei *et al.*[142] developed an ensemble medication-indication KB called MEDI, using four knowledge sources as input: (i) NDF-RT; (ii) SIDER; (iii) MedlinePlus[158] – a website maintained by the National Library of Medicine (NLM) offering health information to consumers; and (iv) Wikipedia[159] – a collaborative encyclopedia on the Internet. The authors determined the accuracy of the KB through manual review by expert physicians, and identified that medication-indication data found in all four resources had a the highest precision (100%) but a very low sensitivity (2%), while data found in only one resource had a lower precision (56% to 97%). No single resource had high sensitivity (20% to 51%). Authors noted that medication-indication pairs appearing in at least two resources had an average precision of 92% and marked

them as the "high-precision subset". This subset had a precision comparable to that of NDF-RT but provided 66% more medication-indication pairs. In a later study, Wei *et al.* used a large clinical dataset and showed that for 97.25% of medications used in outpatient and inpatient settings, MEDI contains at least one indication (the figure was 93.80% for the high-precision set), and medications that were not covered consisted mostly of vaccines, probiotics, nutrition and inert ingredients.[160] Authors did not report for how many of these prescriptions, an indication could actually be found in the medical record.

Jung *et al.* recently published a study which focused on automated detection of novel off-label drug use.[143] They used a large clinical dataset and explored co-mentions of drugs and diseases in the same clinical record. They used a support vector machine classifier to identify positive cases of drug usage, and subsequently removed all known on-label and off-label drug uses (i.e. those medication-indication pairs that were already listed by NDF-RT or MediSpan®) to limit the scope of their study to novel off-label uses only. They also used the side effects list included in the SIDER dataset to remove drug-disease pairs that were likely co-occurring frequently because the disease is a known adverse effect (and not the indication) for the medication. Authors identified 6,142 drug-disease relationships that were categorized as high confidence novel off-label uses of medications. The authors then assessed whether the same indication had been listed in the FDA's Adverse Effect Reporting System (FAERS) data; although FAERS is a resource primarily used for collecting data about adverse effects of medications, the report also includes a field where the intended use of the reported medication can be specified. Previous research had shown that this data can be leveraged to identify indication data.[144] The authors also assessed whether the novel off-label indications have ever appeared in MEDLINE abstracts and reported that out of the 6,142 novel off-label uses found using their method, 766 (12.5%) appeared 10 or

more times in the FAERS reports, and 537 (8.7%) were also co-mentioned in 2 or more abstracts indexed by MEDLINE. However, the final set of novel off-label uses was not manually validated by experts.

Most recently, a group of researchers developed and published a new medication-indication KB called LabeledIn, which tries to address some of the limitations of previous resources.[145] LabeledIn is curated using a semi-automated approach, where NLP is used to facilitate manual annotation of SPLs. They used MetaMap[161] to process all SPLs found in DailyMed, and subsequently presented the results to two professional biomedical annotators in a color coded form. Researchers showed that this process significantly reduces the time needed for annotation, and that agreement between annotators was high ($\kappa = 88.35\%$). LabeledIn lists the medication-indication relationships not only at the level of the active ingredient, but also at finer levels of granularity, including dose form and drug strength. An evaluation on the completeness of LabeledIn using SIDER 2 as the reference standard showed that out of all indications found for a random subset of 50 drug labels, 47.5% appeared in both resources, 10.1% only appeared in LabeledIn, and 42.4% were only found in SIDER; the majority of indications that were only found in SIDER were attributed to the use of less specific terms to describe indications in SIDER (.e.g. “infarction” in SIDER versus “myocardial infarction” in LabeledIn).

Analysis of the KBs

NDF-RT data can also be acquired through UMLS or through RxNorm; we used the version included in UMLS 2014AA, in which medications and indications are both normalized to UMLS concepts. NDF-RT uses an internal unique identifier for medications and indications, called NDF-RT unique identifier (NUI). Two types of relationships can be used to identify indications; they include *may_treat* and *may_prevent*. Overall, 55,704 medication-indication pairs, 1,010

unique indications and 9,579 unique medication concepts were included in NDF-RT, and the medications were mapped to 1,625 main ingredients. For each medication an average of 5.82 indications are listed (range = 1 – 221, SD = 7.74, median = 4). All medications are represented at the main ingredient level.

We used the latest version of SIDER that was available on the web by the end of 2014.[152] This version of SIDER includes information extracted from 18,334 SPLs (which were mapped to 1,307 main ingredients), 165,920 unique label-indication pairs and 2,194 unique indications. SIDER indications are only provided at the SPL level, and searching at the level of main ingredient will require an additional step for mapping the SPL identifiers to the medication concepts.[162] Approximately 18.6% of the indication entries were duplicates (e.g. "infection" and "infections" are both listed as indications for the same SPL, and they are both mapped to the same UMLS concept identifier, i.e. C0021311). After removing duplicates, each medication was associated with an average of 8.93 indications (range = 1 – 129, SD = 12.45, median = 19). SIDER frequently uses nonspecific terms as indications; for instance, wherever *rheumatic fever* was listed as an indication, the nonspecific concept *fever* was also listed as an indication (examples include but are not limited to *Azithromycin*, *Cefprozil*, and *Cefazolin*).

The resource provided by McCoy *et al.* contains 11,166 medication-indication entries, corresponding to 2,537 unique medication concepts and 1,580 unique indications. Medications concepts are defined at the level of dose form and strength (e.g. "Atenolol 50 MG Oral Tablet"), and were mapped to 594 unique main ingredients. For each medication concept an average of 4.40 indications are reported (range = 1 – 122, SD = 7.08, median = 6). Redundancy is observed in the indications; e.g. "Fever (Symptom)" and "Fever (On Exam)" are two separate indications, and no medication is associated to both of these. Redundancy was also observed among

medications (generic and brand names were both present) and several items could not be considered medications (e.g. “Rapid Bacterial Antigen Ident Kit” or “OneTouch Test In Vitro Strip”).

The resource provided by Fung *et al.* contains 19,473 medication-indication pairs, corresponding with 2,885 unique indications and 2,104 unique medications (which mapped to 1,181 unique main ingredients). Medications are represented using RxNorm concepts at the level of semantic clinical dose form (SCDF). Indications are represented in this resource using UMLS concepts. For 3,468 entries a third column indicates whether the medication-indication pair was deemed correct, nearly correct, or incorrect by manual review. On average, for each medication 9.26 indications are reported (range = 1 – 130, SD = 12.56, median = 6). Since medications are represented using their RxNorm identifier, indications are also defined at the level of dose form; for example, separate entries for indications of ‘alprazolam disintegrating tablet’, ‘alprazolam extended release tablet’, ‘alprazolam oral solution’ and ‘alprazolam oral tablet’ could be found, each being associated with a different number of indications.

MEDI contains 65,535 medications-indication relationships, corresponding to 2,897 unique medications and 4,352 unique indications. Medications are represented using both their name and the corresponding RxNorm concept identifier, and they mapped to 2,162 unique main ingredients. Indications are represented using their name, the corresponding UMLS concept identifier and the corresponding concept(s) from International Classification of Diseases, version 9 (ICD9). Each medication is associated with an average of 22.02 indications (range = 1 – 314, SD = 24.52, median = 3). There is no consistency in the use of lower or uppercase letters in the names of diseases and drugs, and some entries are blank. For each medication-indication pair, this resource also specifies how many of the four original resources that were used to create

MEDI had mentioned that specific pair, although the particular resources that are used are not specified.

Jung *et al.* included 407 off-label medication-indication pairs in their data set, corresponding to 173 unique indications and 214 unique medications (mapped to 212 unique main ingredients). They represented the medications and indications only using their names. For each medication-indication relationship they also listed the number of supportive artifacts found in FAERS and in MEDLINE, as well as the cost index and risk index calculated for the respective medication. All medications are represented at the main ingredient level. While Jung *et al.* used UMLS CUIs to represent the medications and indications (as described in their paper), UMLS unique identifiers are not included in the final resource. Although Jung *et al.* collected data about the known uses, they did not provide that data because it was outside the scope of their study. As a result, the resource provided by Jung *et al.* only lists an average of 1.90 indications per medication (range = 1 – 8, SD = 1.31, median = 1).

LabeledIn contains 19,599 unique medication-indication relationships, corresponding to 626 unique indications and 2,376 unique medication concepts, which mapped to 251 unique active ingredients. The data itself comes from 500 labels, but for each label, additional RxNorm identifiers are assigned to the medication when applicable. Each medication concept is associated with an average of 8.29 indication concepts (range = 1 – 144, SD = 11.76, median = 3). Medications are defined using RxNorm concepts, at the level of dose form and strength. Therefore, similar to the resource by Fung *et al.* LabeledIn lists different indications for medications that are essentially different dose forms or strengths of the same ingredient. However, in addition to the previous, the indications are also defined at higher levels, including the main ingredient. Khare *et al.* reviewed SIDER as part of their evaluation on LabeledIn, and

concluded that a lot of the mismatches between SIDER and LabeledIn were because SIDER uses nonspecific terms for indications,[145] a finding that we also observed in our analysis of this KB.

We also analyzed if each KB can represent various types of challenging or complex indications (**Table 8**). Most resources did not distinguish preventive and therapeutic indications from each other. Only three resources could specify indications specific to a particular dose form or route of administration (namely, McCoy, Fung *et al*, and LabeledIn). NDF-RT is the only resource which also specifies contraindications.

Discussion

Out of the seven KBs included in this study, independent assessments were found only for NDF-RT and SIDER. Evaluations of NDF-RT show that it has limitations in drug class information and alignment of its concepts with those in other terminologies such as RxNorm.[147,149] Additionally, the majority of concepts marked as "Chemical & Drugs" in NDF-RT are not associated with an indication,[150] and some researchers have found NDF-RT complex to use for reasons such as lack of meta-data annotations and the use of unfamiliar drug classifications.[151] Khare *et al*. evaluated SIDER and showed that the use of nonspecific terms as indications in SIDER can lead to alignment issues with other resources. [145]

None of the KBs reported the recency of the information they contained, and only SIDER and NDF-RT have been updated by the time of this review. This limits their usefulness, as many of these resources have been developed several years ago, and may now contain outdated information. It should be noted thought that Khare *et al* have proposed a plan for updating LabeledIn on a regular basis and have estimated the amount of work needed to be minimal.[145]

	NDF-RT	SIDER	McCoy *et al.*	Fung *et al.*	MEDI	Jung *et al.*	LabeledIn
Therapy	+	+	+	+	+	+	+
Prevention [a]	+	+/–	+/–	+/–	+/–	+/–	+/–
Co-medication	–	–	–	–	–	–	–
Dose forms	–	–	+	+	–	–	+
Route	–	–	+	+	–	–	+
Strength	–	–	+	–	–	–	+
Duration	–	–	–	–	–	–	–
Primary choice	–	–	–	–	–	–	–
Contraindication	+	–	–	–	–	–	–

Table 8 – Capability of medication-indication KBs to represent complex indications.

[a] Although several resources contain examples of preventative indications in their data, only NDF-RT specifies whether an indication is preventative or therapeutic. Those resources that do not make this delineation are marked as +/–.

Most of the KBs represent the medications only at the level of main ingredient. This can be challenging when different dose forms or strengths of the same medication have different indications. Similarly, resources varied in their support for medications with multiple active ingredients. While in theory KBs that defined the medications at the level of SPLs or using SCDF identifiers from RxNorm should have no difficulty in this regard (because medications with multiple ingredients generally have a separate label identifier and RxNorm identifier associated with them) only one resource clearly marked combination drugs. Another challenge arose concerning medications used in primary or secondary prevention of disease (e.g. vaccines, or *atorvastatin* as a secondary prevention of cardiovascular disorder). Although some of the KBs did contain preventive indications, only NDF-RT discriminated between *may_treat* and *may_prevent* associations.

The granularity of concepts used to describe indications is directly associated with the choice of terminologies. For instance, oral vancomycin is used to treat recurrent infection with Clostridium difficile, and it is not indicated in non-recurrent infections. Since terminologies such as SNOMED CT or ICD9 do not have a separate concept for the recurrent form of this infection, resources using these terminologies cannot accurately specify the use of vancomycin for this infection. Issues with granularity also present themselves at the level of normalization to unique concepts of diverse terms used to describe a disease or condition. For example, most KBs listed myocardial infarction or another synonymous term as an indication for the drug nitroglycerine but because the relationship was defined at this level, these KBs were unable to represent the very important exception, i.e. right-sided myocardial infarction which can lead to lethal hypotension when using nitroglycerine. Terminologies such as SNOMED CT or ICD9 do have

the level of granularity necessary for this example, but the KBs used a coarser concept to represent the indication. A relevant problem occurs for those pathologies that have a spectrum; in this case the challenge would be to include various stages of disease as an indication. For example, Barrett esophagus is commonly treated with proton pump inhibitors (PPIs), but none of the KBs included the association between PPIs and Barret esophagus; some of them, however, included Erosive Esophagitis or a synonymous term as an indication for PPIs, which is the pathophysiological state before Barrett esophagus occurs.

The issues associated with the granularity of concepts used to represent the indications are summarized in **Table 9**.

Finally, it would be impossible to capture all the possible variables of medication-indication knowledge in all circumstances and as a result, any CDS system should allow for exceptions to the rule. Patients with refractory diseases, those who have allergies or contraindications for the "indicated" therapy, and those who have rare diseases for which no drug therapy is established yet (e.g. Ebola infection, for which treatments are all in the experimental stage) may benefit from receiving prescriptions that may contradict with the information in the medication-indication KBs. Additionally, medication-indication KBs are meant to provide evidence-based knowledge, but there will always be new knowledge which may not be incorporated into these resource but still qualify as appropriate use of medications.

Recommendations for future research

One of the key limiting aspects of many of the existing medication-indication KBs is the use of simple binary relationships that do not capture characteristics of the medication or indication. This challenge is partly addressed in those resources that define the medication concepts at a finer level of granularity (e.g. dose form and strength), or use different relationship types to

Complexity	Example	Description of the example
Location	Left-sided myocardial infarction	Nitrates are indicated in the treatment of myocardial infraction, except in right-sided myocardial infarction in which they are contraindicated (because they can cause severe hypotension). If the indications for nitrates are defined at a coarse level, they may not capture this complexity.
Recurrence	Recurrent infection with *Clostridium difficile*	Initial infection with *C. difficile* is treated with metronidazole, but recurrent infection is treated with oral vancomycin. If "recurrence" is not captured in the indication, the KB may wrongly imply that vancomycin is indicated in any *C. difficile* infection.
Symptoms	Acetaminophen for fever	Although fever is not a disease itself, acetaminophen is indicated in patients with fever when it is necessary to alleviate their symptom.

Table 9 – Issues with granularity of the concepts used to represent the indications.

discriminate between preventive and therapeutic indications. Nevertheless, certain nuances of indication knowledge (such as co-medication, co-morbidities or primary choice of therapy) are still not captured by any of the existing resources, indicating that there is a knowledge gap on how to capture this information in an automated way.

Various data sources that have been leveraged to identify adverse effect information may potentially be used to extract indication data as well. Harpaz *et al* recently published a review of various data sources used for extracting adverse effect information.[163] Previous studies have leveraged patient-generated data on the internet to identify adverse effects of medications, for example by analyzing the Internet search logs[164], Twitter posts[165], or user contributions to online health communities.[166,167] Similar data sources and approaches may be plausible for identifying

off-label uses of medications, or other types of information about medication indications, such as their effects on various symptoms.[168] Additionally, while the FAERS database is primarily designed to collect and store adverse effect data, it can be leveraged to extract the "use" information included in the adverse effect reports to identify true indications for medications.[144]

Limitations

Our work has several limitations. First, we restricted our literature review to studies with electronic full texts in English, and only searched PubMed, PubMed Central, EMBASE or CINAHL. Other medication-indication KBs that did not meet our inclusion and exclusion criteria were not reviewed in this study, and the results of our review may not be generalizable to those resources. Our review only includes those publications that are about a publicly available KB for medication indications. In other words, we did not evaluate proprietary KBs that contain medication-indication information, and the strengths and weaknesses of those resources may be different than those described here. Also, we excluded those publications from our review that only discuss a method for developing a medication-indication KB but do not actually provide the resulting KB, because the scope of this review was to compare the KBs not the methods; examples of excluded papers include the works by Wright *et al*[148] and Chen *et al*.[146] Second, the method we used for analyzing the limitations of KBs in representing complex or challenging indications was based only on an informal query sent to a small group of clinicians and pharmacists. It is possible that other categories of complex or challenging indications exist that were not studied in this work. This signifies the need for developing quality metrics for evaluation of medication-indication KBs. Finally, the normalization of medication concepts into main ingredients was not perfect, and it did not handle combination medications properly. Similarly, we aggregated the indications in free-text form, which is not ideal. Moreover, multiple

indications that are assigned different concept identifiers may actually reflect the same true indication (e.g. "*depressive episode, unspecified*" and "*major depressive disorder, single episode, unspecified*") and closely related indications may need to be aggregated (see examples provided in Results section). Further work is needed to normalize all indications, and to aggregate the relevant indications.[169]

Conclusions

Medication-indication KBs are important resources for data-driven research on appropriate uses and adverse effects of medications, as well as for providing automated decision support regarding appropriate use of medications. Each of the KBs reviewed above is a significant step forward in developing a comprehensive computable medication-indication public knowledge. This review also identifies some of the key gaps in the KBs at the level of representation and content, with the aim of motivating researchers to address these important limitations in future research. Characteristics of the ideal KB depend on the task at hand, but the results of this review can help the reader decide which KB meets their needs the most.

Appendix 1

The medication-indication knowledge-bases (KBs) included in this study used various methods to represent the medications, including RxNorm identifiers, UMLS concepts, and free-text. Additionally, those which used RxNorm identifiers or UMLS concepts did not do it identically: some represented the medications at the level of main ingredient (IN), while others represented at the level of specific dose forms (SCDF).

In order to be able to compare these KBs, we normalized them to the same level by mapping every medication concept to the main ingredient using RxNorm's Application Programming Interface (RxNorm API) available at http://rxnav.nlm.nih.gov/RxNormAPIs.html as explained below:

Methods

Whenever the KB defined the medications at the dose form level, we used the ***allRelated*** function from RxNorm API to identify the main ingredient (TTY = IN). For instance, the result of the query http://rxnav.nlm.nih.gov/REST/rxcui/373222/allrelated indicates that the RxNorm identifier for the main ingredient is 40790, where 373222 and 40790 represent *Pantoprazole Delayed Release Oral Tablet* and *Pantoprazole*, respectively.

Whenever the KB defined the medications using UMLS concepts, we used the *idtype* parameter in RxNorm API to find the RxNorm identifier for the medication; then we applied the procedure above to find the main ingredient. For example, the result of the query http://rxnav.nlm.nih.gov/REST/rxcui/?idtype=UMLSCUI&id=C0012772 indicates that the RxNorm identifier for that concept is 3554, where C0012772 and 3554 both represent *Disulfiram*. In this case, *Disulfiram* is the main ingredient itself.

Whenever the KB defined the medications in text form, we used the *approximateTerm* function of RxNorm API to identify the main ingredient. If the resource defined the medications using their dose forms, we first preprocessed the list of medications by removing the suffixes that specify the dose form, including "Oral", "Tablet", "Capsule", etc. as well as numerical values and units (such as "81 MG", etc.). This was done using a set of regular expressions. Finally, the list of medications was fed into the RxNorm API to get the RxNorm identifiers. For example, the query http://rxnav.nlm.nih.gov/REST/approximateTerm?maxEntries=1&term=venlafaxine will return a list of RxNorm identifiers matched for the string "venlafaxine". We then used the methods described above to map these to the main ingredient (which in this example led us to concept identifier 39786).

Finally, SIDER defines the medications using their structured product label identifier (SPL ID). SIDER also provides a mapping from SPL IDs to the generic name of the medication or its ingredients, as a file hosted on their website (http://sideeffects.embl.de/media/download/label_mapping.tsv.gz). We used this mapping to map the SPL IDs to text form, then used the procedures explained above to find the main ingredient.

Results

The table above shows the type of mapping done for each resource, and the percentage of unique medication concepts that were successfully mapped to main ingredients in RxNorm. For two resources, mapping was 100% possible. For two other resources (Fung *et al* and LabeledIn) mapping was more than 98% accurate. Medication concepts in these two KBs that failed to map to main ingredients were all interpreted as invalid identifiers by the current version of RxNorm API.

Manual review of the medication concepts from NDF-RT, that did not map to RxNorm identifier showed that while some of them were actual medications (e.g. C0981111 representing *Vitamin E 13500 U/mL Topical Cream*), others were prescribable items that do not contain any active ingredients (such as *Toothpaste*) or classes of chemical substances which cannot be mapped to an active ingredient (such as *Triglycerides*).

Manual review of the medication concepts from McCoy *et al* that did not map to RxNorm showed that a notable fraction belonged to non-medication concepts (such as test strips, syringes, and blood pressure kits) and many of the concepts that were actual medications were mentioned using a brand name that was potentially not recognized by the RxNorm API.

	From	**To**	**Total**	**Mapped (%)**
NDF-RT	UMLS	RxNorm IN	9579	8076 (84.3%)
SIDER	SPL ID	RxNorm IN	18334	14866 (81.1%)
McCoy *et al*	Free text	RxNorm IN	2537	2230 (87.9%)
Fung *et al*	RxNorm	RxNorm IN	2104	2071 (98.4%)
MEDI	RxNorm	RxNorm IN	2897	2897 (100%)
Jung *et al*	Free text	RxNorm IN	214	214 (100%)
LabeledIn	UMLS	RxNorm IN	2376	2354 (99.1%)

Table 10 – Result of normalizing medication to RxNorm.

Manual review of the medication concepts from SIDER that did not map to RxNorm showed that almost all of them were combination drugs. The mapping provided on the SIDER website lists

all ingredients of a combination drug next to its SPL. However, RxNorm did not return any concepts for these drugs because the search string contained more than one drug name.

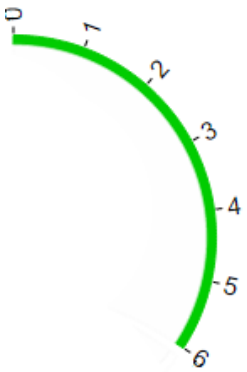
Limitations

The main limitation of our approach above is that when more than one main ingredient was returned for a medication, we only used the first one. This fails to capture the complete data for combination drugs.

Appendix 2

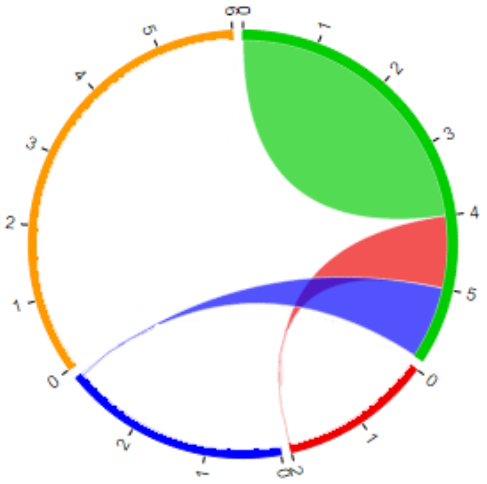
In order to visualize the information from **Tables S1-S3** above, we used chord diagrams (**see Figure 7, Figure 8** and **Figure 9**). The figure consists of three diagrams (top, middle and bottom) showing the overlap for medications, indications, and medication-indications pairs, respectively. The knowledgebases are consistently color-coded in all the three parts (e.g. NDF-RT is always shown in red, MEDI in green, and so forth).

For each diagram, we first ordered the resources based on the unique number of items in them. Then we drew a properly colored sector on the outside of the circle. For instance, in the middle diagram, MEDI is shown first because it has the largest number of unique indication.



Each sector has tick marks to help the reader visually comprehend the size of that resource. For example, tick marks on the example on the right show that it has a size of 6. Similarly, the tick marks on the MEDI sector in the **Figure 8** indicated that MEDI contains between 4,000 and 4,500 unique indications; the actual number is mentioned in **Table S2** which is 4,352.

Next, we compared each sector with its subsequent sectors. The comparison is unidirectional to enhance visibility. If a KB had overlap with its subsequent KBs it was shown using chords. The example on the right is for demonstration. It visualizes four hypothetical resources (shown in green, red, blue and orange). In this figure, not all chords are shown: only chords that are associated with



the first resource (green sector) are shown for demonstration purposes. Two types of chords can be seen: those connecting the green sector to another sector (in this case red and blue) and those connecting it to itself (resembling a green dome). The interpretation of this example will be as follows: the green sector contains 6 items; 4 of those items are unique to the green sector, 1 appears in the red sector as well, and 1 appears not in the red sector but in the blue sector. Note that the origin of the bands (on the green sector) is non-overlapping. It means there may be items in the green sector that appear both in the blue and the red sector, but they will not be counted twice in this diagram.

Similarly, **Figure 8** shows that about 2,000 of the unique indications from MEDIE don't appear in any of the subsequent resources, about 1,600 of them appear in SIDER, about 1,500 of them appear not in SIDER but in Fung *et al*, and so forth.

One of the most important aspects of the chord diagrams are the chords that start and end in the same sector (dome shaped chords). These represent the items that do not appear in any of the subsequent resources. In **Figure 7** and **Figure** 8, these domes are relatively half the size of the corresponding sectors, indicating that about half of the medications and indications found in each KB are unique to that KB. In the bottom diagram, the domes are significantly larger, showing than the majority of medication-indication pairs in each KB are unique to that KB. This is also reflected in **Figure 6** above. The first bar in each diagram in that figure shows the number of medications, indications, and medication-indications to appear exactly in one KB (i.e. they are unique to the KB in which they appear). This bar is always significantly larger than the subsequent bars in the same bar graph.

Study 3: Developing a Formal Representation for Medication Appropriateness Criteria

Introduction

Identification of IUM requires extracting information about the patient, the medication, and other treatments, and comparing them with the MAC.[16] Consequently, developing automated solutions to reduce IUM entails two requirements: a framework to represent the medication appropriateness criteria formally, and methods to extract the information needed to compute these criteria. This article focuses on developing a framework for formal representation of medication appropriateness criteria.

Previous researchers have identified several medication appropriateness criteria and metrics through systematic review of the literature.[44,45] These criteria can be categorized into three groups: the first group enumerates the conditions in which the use of medication is appropriate (e.g. see Choudhrey *et al.*'s criteria for appropriate use of proton pump inhibitors [PPIs]),[110] the second group lists conditions in which the use of medication is deemed inappropriate (e.g. see the Beers' criteria),[39] and the third group provides a combination of both (e.g. see Oborne *et al.*'s criteria on appropriate use of neuroleptics).[112] Since older adults are frequently subject to polypharmacy and therefore more likely to experience the negative impacts of IUM (e.g. drug-drug interactions, adverse drug reactions and increased risk of hospitalization)[170–172] larger collections of medication appropriateness criteria exist for the geriatric population. Examples include the Beers' criteria[39] and the Screening Tool of Older Persons' potentially inappropriate Prescriptions (STOPP)[116] which aim to reduce inappropriate use of medications (overuse), and

the Screening Tool to Alert doctors to Right, i.e. appropriate indicated Treatment (START)[115] which promotes appropriate use of medications that are omitted (underuse).

Medication appropriateness criteria can be described as a special form of clinical guidelines, although they have distinct features that separate them from the majority of clinical guidelines. Clinical guidelines provide best practices for diagnosis and therapy of diseases, but medication appropriateness criteria are focused on proper utilization of a resource (namely, medications). Clinical guidelines are primarily developed by major medical associations, are organized in a common format and are hosted on repositories such as the National Guideline Clearinghouse;[173] in contrast, medication appropriateness criteria are mostly developed by independent groups of researchers and distributed without using a common format or central repository.

Medication appropriateness criteria are currently only available in narrative form, and transforming them into a computable format is challenging because a formal representation for the components of medication appropriateness criteria does not exist. Different criteria have varying levels of granularity and specificity in defining the medications, diagnoses, and symptoms; in addition, some but not all of the criteria are accompanied by information regarding the level of evidence, target population, or extent of clinical relevance. A framework in which these criteria can be explicitly and comprehensively represented is needed. We developed such a representation framework, which we call the Objective Medication Appropriateness Criteria (OMAC).

Methods

In order to study existing medication appropriateness criteria, we started by identifying these criteria by searching PubMed using the following keywords and their variations to identify

published medication appropriateness criteria: inappropriate prescribing, overuse, overtreatment, overutilization, and utilization review. We grouped the relevant studies based on the actual criteria they used to identify IUM. We then curated a collection of published medication appropriateness criteria and used a random subset of those to develop OMAC (for examples, see **Table 11**). To ensure that we didn't have any biases in our component selection and semantic aggregation of the concepts, we used another independent set of criteria for evaluation of OMAC.

Developing OMAC

The purpose of OMAC is to provide a formal representation for appropriateness criteria. We manually analyzed a randomly selected subset of medication appropriateness criteria to identify their components. Each criterion can be described as one or more rules, and We semantically grouped the components we found in the sample criteria to define concepts that comprise the criteria, including high-level elements (such as the general sections of a criterion) and low-level elements (such as modifiers, identifiers, names, etc.) and we also identified the relationships between these concepts and represented them in OMAC.

Different medical concepts are frequently mentioned in the medication appropriateness criteria, such as medications and diseases. OMAC is only a formal representation model and is not a vocabulary itself, therefore we ensured that OMAC takes advantage of previously developed ontologies and terminologies, by linking to external ontologies and terminologies to the extent possible. This will also ensure the concepts are defined in a standard way that can be reused by others. We saved OMAC using frames and properties in Protégé version 3.5.

Evaluating OMAC

After the initial design of OMAC was completed, we presented a separate set of 10 randomly selected MAC to a group of domain experts (physicians and pharmacists) in form of a questionnaire, asking them to identify and categorize the components of these criteria independently. Each item in the questionnaire consisted of one medication appropriateness criterion statement in its original narrative form, and requested that the participant break the statement into basic elements (such as medication names, medication class names, disease names, logical statements, or temporal modifiers). Disagreements in the experts' responses were identified through qualitative analysis of the responses. Three types of disagreements were considered: differences in the classification of the same word or phrase (e.g. classifying "hypertension" as a disease versus a problem), differences in specification of the elements in the statements (e.g. considering "severe hypertension" as two separate concepts versus one), and classification of terms into concepts that are not explicit (e.g. classifying "long-term use of drug X" as one concept of type "overuse"). In a subsequent questionnaire, we presented the experts with the same narrative criteria but clearly marked these areas of disagreement and asked the experts to translate those terms and phrases into more detailed, explicitly defined concepts. Note that the purpose of this process was not to reach perfect agreement, but rather to identify what "elements" constitute the criteria and also to describe the elements so that they are well-defined, so that we can evaluate OMAC's coverage for those elements.

Subsequently, we evaluated whether OMAC could represent all of the explicitly defined concepts provided by the experts. We froze the development of OMAC before we started sending out the questionnaires, to ensure that our knowledge of the results of the previous step would not affect our evaluation of OMAC's completeness. We planned to correct OMAC for any

areas of deficiency that would be found throughout this evaluation, only after the evaluation was completed.

Results

We identified 110 medication appropriateness criteria through literature review, and used a random subset of 40 to develop OMAC. We designed OMAC such that each criterion in this subset could be represented using four functional types of information: '*trigger'*, '*rules'*, '*action'*, and '*meta-data'*. A *trigger* may consist of one or more medications that are the primary focus of the criterion (when prescribing these medications the criterion would be triggered) or one or more clinical conditions in which the use of a certain medication is desirable (in this case the criterion would focus on underuse). *Rules* specify the conditions that a patient must meet to be eligible for the criterion (such as age limit, past medical history, medications prescribed, symptoms, or paraclinical findings). *Action* specifies the recommendation that the criterion makes once the patient meets all the rules; generally, actions are in two forms, either to *avoid* prescribing a medication or to *consider* prescribing a medication. *Meta-data* includes all the additional information that is used to describe the criterion (examples include a name or unique identifier, references to citations, or a justification or concern). As an example, Beers' criteria not only lists medications or combinations of drugs that should be avoided in the elderly, but also specifies what "concern" exists around using these medications, and also provides a "severity rating" for this concern (low vs. high) to help the clinicians determine the importance of each item in this criteria and provides the relevant citations (**Table 11**).[39]. We represented the trigger, action and meta-data components using properties for the "criterion" class (**Figure 11**, right). We used a more complex classification as described below to represent the rules.

Each criterion can contain one or more rules, and there are various types of rules in different criteria. These include *'medication rules'* which specify the medication that is the subject of the criterion as well as co-prescribed medications that need to be considered, and '*clinical rules'* which specify the diseases, symptoms, laboratory tests results, and demographics that have to be present or that should be absent for the patient to meet the criterion. This can be clarified using the third example shown in **Table 11**: "TCA's with cardiac conductive abnormalities"; this item from STOPP criteria states that in elderly patients who have cardiac conductive abnormalities, tricyclic antidepressants (TCAs) should be avoided because of their pro-arrhythmic effects.[32] To apply this criterion to a patient, three rules must be satisfied: (i) the patient must belong to the 'elderly' demographic group (formally defined as age ≥ 65 years), (ii) the patient must have been diagnosed with a cardiac conductive disorder (including, but not limited to Type I heart block, Type II heart block, or right bundle branch block), and (iii) the patient must have been prescribed a medication that belongs to the TCA class. The first two rules in this example are clinical rules, and the latter is a medication rule.

Clinical and medication rules have different properties: clinical rules may focus on the existence, temporality and duration of a clinical finding or condition (including but not limited to diseases, signs and symptoms, allergies, and contraindications), or the result of a measurement (such as a lab test or imaging study), but medication rules may specify the dose, route, frequency and form of a medication. Both clinical and medication rules may include concepts that are externally defined in other ontologies or terminologies (**Table 12**). In the example provided above, "cardiac conductive abnormalities" can be represented as a clinical rule, which can refer to a pertinent concept in International

Classification of Diseases, version 10 (ICD-10) or Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT), and thus, it is possible to link the concept to a standardized external knowledge source. A link to an external concept consists of four parts: the name of the concept as mentioned in the criteria (e.g. 'cardiac conductive abnormalities'), the name of the external ontology or terminology and its version number (e.g. ICD-10), URL of the external ontology or terminology (e.g., http://purl.bioontology.org/ontology/ICD10) and the unique identifier of the corresponding concept in that external ontology or terminology (in this case 'I44').

	Source	**Narrative criterion**
1	2002 Beers' criteria	Disease: Seizures or epilepsy Drug: Clozapine, chlorpromazine, thioridazine, and thiothixene Concern: May lower seizure thresholds Severity Rating: High
2	STOPP, section A, item 2	Loop diuretic for dependent ankle edema only i.e. no clinical signs of heart failure
2	STOPP, section B, item 3	TCA's with cardiac conductive abnormalities
3	STOPP, section A, item 6	Beta-blocker in combination with verapamil
4	STOPP, section A, item 7	Use of diltiazem or verapamil with NYHA Class III or IV heart failure
5	STOPP, section B, item 9	Use of aspirin and warfarin in combination without histamine H2 receptor antagonist (except cimetidine because of interaction with warfarin) or proton pump inhibitor

Table 11 – Examples of MAC previously published in the literature.

Concept Type	**External ontology or terminology**
Medication	RxNorm, ATC, NDC
Disease	ICD, SNOMED CT
Symptom	SNOMED CT, Symptom Ontology
Procedure	CPT, ICD-9 or ICD-10, SNOMED CT

Table 12 – Examples of external terminologies used to define concepts in MAC.

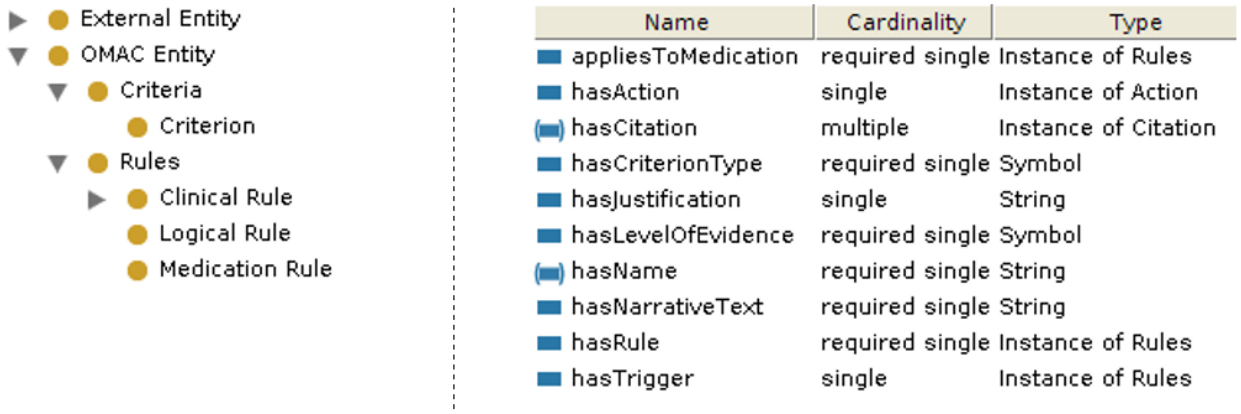


Figure 11 – Main concepts in the OMAC (left) and the properties of the criterion class (right).

Clinical and medication rules can be combined with each other using *'logical rules'*. Each logical rule has a mandatory field which specifies the Boolean operator it is representing ('AND', 'OR', or 'NOT'). In the example above, the clinical and medication rules are combined using a logical rule with 'AND' logic (i.e. the patient must be among the elderly AND have a cardiac conductive disorder to be eligible for this criterion).

We grouped all the three aforementioned types of rules under a parent class called *'rules'* (**Figure 11**¸ left). To represent complex statements, these rules can be nested to create 'rule trees'. Clinical and medication rules can only appear as the leaves of the rule tree. Logical rules appear as branches of the tree, and each logical rule references one or more rules of any type. The latter enables nested rules which allow representation of complex logical statements. The last example in **Table 11** (item B9 from STOPP) demonstrates a criterion with a complex logic. This complex statement can be encoded through nesting different types of rules, as shown in **Figure 12**.

Evaluating OMAC

Eight domain experts collaborated in the first questionnaire. There was no disagreement among experts in their responses for simple and well-defined criteria; for instance, all collaborators described STOPP criteria item A6 (**Table 11**) using similar components. We observed disagreements with concepts that are not explicitly defined (NED); for example, there was lower agreement on how the terms 'dependent ankle edema' and 'no clinical sign of heart failure' were categorized by different experts. When experts clarified the areas of vagueness using detailed explicit concepts, we noticed that although they clarified these vague terms using different sets of explicit concepts, they used similar 'types' of concepts to describe them. For example, each expert used a different set of 'signs', 'symptoms' and 'paraclinical findings' to describe the phrase 'clinical signs of heart failure', but all experts used exactly those three types of information.

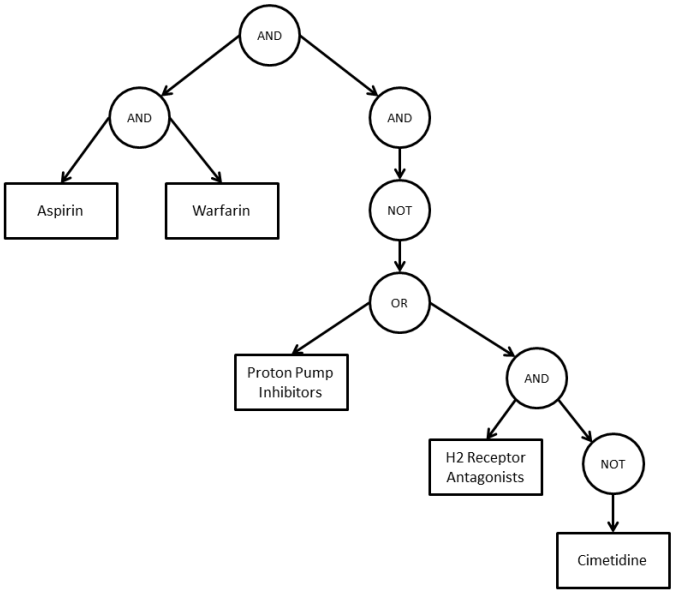


Figure 12 – Nesting logical statements representing a complex criterion.

All of the types of information that experts used to transform the vague phrases into explicit forms were consistent with the types of information that we had already incorporated into OMAC's clinical or medication rules. In other words, OMAC had complete coverage for all of the criteria that were coded by the experts, and as a result we did not modify OMAC after this evaluation.

Discussion

Developing a representation format for medication appropriateness criteria is the first step towards developing computable, interchangeable and reusable solutions to prevent inappropriate medications use. OMAC formally defines the structure of explicitly defined medication appropriateness criteria, and allows referencing to external ontologies and terminologies when applicable.

The results of our questionnaire study indicate that at least some of the medication appropriateness criteria are defined using vague terms that were interpreted differently by the experts. These criteria only provide guidelines for appropriate use of medications, and variability in the application of guidelines is a well-established phenomenon in health care practice; however, ideally the guideline itself should be interpreted identically by all of its users so that the variability should be only due to the specific characteristics of the patient or the settings in which the guideline is used, and not due to different interpretations of the appropriate care.[174,175] Although our questionnaire study has a small sample size, it signifies the need for well-defined medication appropriateness criteria. OMAC can facilitate this process, as encoding the criteria into OMAC requires translating all terms into explicitly defined medication, clinical or logical rules.

OMAC is designed to be flexible, and allow for multiple ways of defining concepts and their relationships. Its design does not limit it only to medication appropriateness criteria; other appropriateness criteria (e.g. those for appropriate use of tests and procedures) can also be formally represented using the same approach. Through the use of logical rules, it is possible to model the steps that are used to implement medication appropriateness criteria in clinical practice and encode these steps in a computable way. When a clinical or medication concept is in fact referring to a class of diseases or medications, logical steps can be used to internally define these sets instead of referencing external knowledge sources, which is important when defining a concept that does not exist in any external knowledge source. Therefore, the user has the choice of either specifying a medication class by referencing an external entity, or by defining external references to each member of that class and then combining them using an 'OR' logic (**Figure 13**). Each approach has its own advantages: using an external reference for each of the elements in that class makes the local definition of the criteria more explicit, while using an external reference for the class itself reduces the amount of effort needed to encode the criteria in OMAC.

Using an external ontology or terminology to define the concepts in OMAC also has the advantage of reusing knowledge that has been vetted by a group of experts, but a suitable external knowledge source may not be available in all cases, or it may not be as accurate or complete. In addition, not all of the concepts that are found in medication appropriateness criteria can be identically found in external knowledge sources. For instance, one medication appropriateness criteria may specify the severity levels for heart failure using the classification provided by the New York Health Association, but this classification may not be already defined in any existing disease ontologies and terminologies. OMAC flexibly supports defining these

complex concepts either by external links (when possible) or locally, and the users can choose their preferred method based on the task at hand (**Figure 13**).

OMAC is different from a guideline representation language. While medication appropriateness criteria can be described as a special form of guidelines, guideline representation languages (e.g. GEM[176], GLIF[177], EON[178], PRO*forma*[179], and SAGE,[180] among others) do not enforce the mandate level of detail in their formalism that is needed for representing medication appropriateness criteria. Guideline representation languages provide a structured way to encode the "flow" of decisions in a guideline. However, to ensure that they can support different types of decision and various forms of guidelines, they provide a significant amount of flexibility as to how each decision step is defined. Previous research has shown that these guideline representation models have limitations when applied to medication related guidelines used for chemotherapy, and that representing medication related guidelines as rules can address this limitation.[181] OMAC combines this rule-based approach with specific features of guideline representation language (such as the inclusion of meta-data about provenance of the guidelines), to provide a more strict structure to represent the medication appropriateness criteria than guideline representation models, thereby providing a common framework for encoding all such criteria in a similar, interchangeable way. In that sense, OMAC complements the guideline representation languages by providing the formalism that is necessary for a certain type of decisions, namely the decision about appropriateness of medications.

One potential challenge in interchanging OMAC-encoded criteria is that a criterion may be encoded using an external ontology or terminology which may be different from what is desirable for a second user of the criterion. This challenge can be addressed by creating cross-walks between these external knowledge sources; in many cases, this can be easily possible

using the Unified Medical Language System (UMLS). Finally, our study is also limited in that we did not conduct a large scale evaluation of the completeness of OMAC. We intend to address this limitation in future research. We also intend to use OMAC to develop structured representations of well-established medication appropriateness criteria and then export them into a format supported by HL7 Clinical Decision Support (CDS) standards. Namely, we intend to use the virtual medical record (vMR) format[182] to represent the patient data, and use the OpenCDS platform[183] to integrate the computable medication appropriateness criteria with the medical records and evaluate the accuracy and impact of using this approach to provide decision support regarding appropriateness of medications**.**

Conclusions

OMAC provides the necessary flexibility for defining concepts using external ontologies or terminologies whenever applicable, and through the use of rules, it enforces the necessary formalism to ensure that all essential concepts of the medication appropriateness criteria are

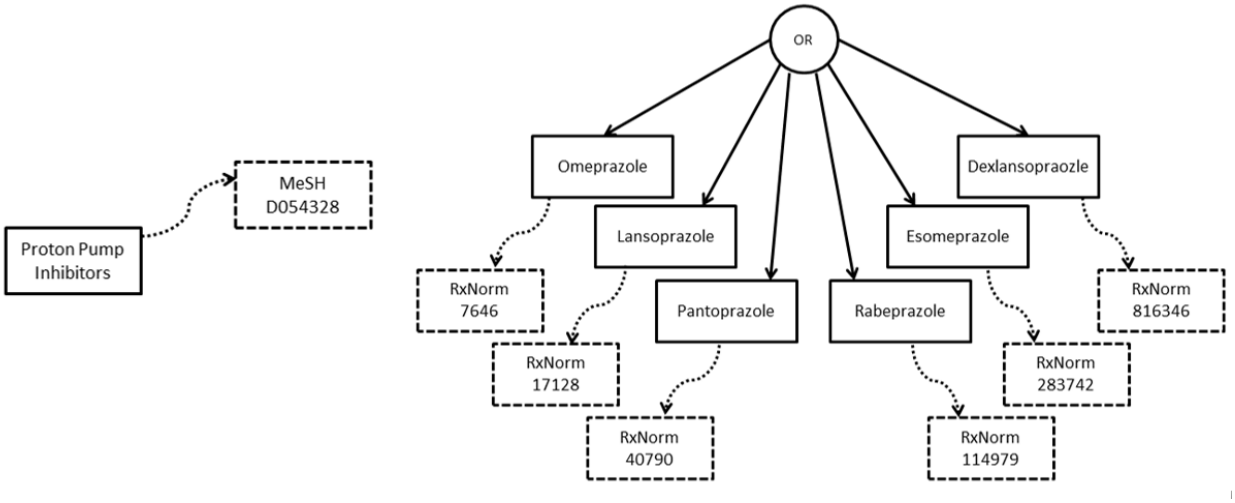


Figure 13 – Linking concepts to external terminologies.

represented using a common structure. To the best of our knowledge OMAC is the first framework that specifies encoding the medication appropriateness criteria into a formal, structured form, which is necessary to incorporate a decision support component aimed at reducing IUM.

Chapter 4: Translating appropriateness criteria into the formal representation

Background

Once the formal representation for MAC was developed, the next challenge was to develop a workflow through which existing narrative can be accurately translated into structured form. However, as shown in previous chapters, the translation task is often non-trivial. Narrative MACs commonly contain not explicitly defined (NED) concepts and therefore translating them into the formal representation will require input from domain experts (pharmacists and clinicians).

In this chapter, we describe a tool called *MAC Annotator* which was designed particularly for that purpose. We demonstrate that using a simple but real-time natural language processing approach can assist with translating of some but not all of the concepts in the MAC, and that using a real-time search feature that allows domain experts to browse existing biomedical terminologies can further facilitate the translation process.

Study 4: Translating MAC into structured form using MAC Annotator

Introduction

Medication appropriateness criteria (MAC) are narrative guidelines that are used to identify inappropriate use of medication (IUM). Currently, MAC are only available in narrative form, therefore automated implementation of them is limited. Ideally, a repository of MAC in computable form is needed for scaling the automated application of MAC for real clinical settings, but creating such a repository would be practical only if a process exists that would facilitate translating the MAC into computable form.

In our previous work, we showed that each MAC is comprised of four major components: *trigger*, *rules*, *action*, and *meta-data* (see Chapter 3).[184] While the trigger, action and meta-data are generally easy to represent, translating rules into computable form introduces challenges. We have shown that the rules can ultimately be represented as logical statements (combined using Boolean operators *AND*, *OR*, and *NOT*), and the elements of the logic consist of either clinical rules, medication rules, or another logical rule. The latter will allow nesting of logical rules, which is essential for the formal representation of the MAC. We have also shown that the medical rules – in which patient characteristics, indications, contraindications, symptoms, etc. are mentioned – and the medication rules – in which the specific therapy and other co-medications are explained – frequently contain concepts that are not explicitly defined (NED); these terms may be vague (e.g. "long-term") or context dependent (e.g. "monotherapy") which need to be substituted with explicit definitions before the MAC can be completely represented in computable form.

In our previous work we also showed that using the help of experts it is possible to translate the MAC into a computable form and implement it to identify IUM with high accuracy.[37] In that study, one of the rate limiting factors in the implementation of MAC was the process of translating the narrative MAC into computable form. Specifically, the approach used in that study included focus group sessions between domain experts and separate meetings between clinical and informatics experts. During the focus group sessions, domain experts would review the narrative MAC and resolve issues with NED, add any clinical details that were necessary for implementation of the MAC, and resolve all disagreements. In a subsequent meeting between clinical and informatics experts, the vetted MAC was then translated from narrative form into computable form. Because this process was serialized and required several iterations until the translation process was complete, it was not efficiently scalable.

In this study, we developed and validated a tool we call MAC Annotator, which is an interactive tool for annotation of MAC. This tool allows domain experts (clinicians and pharmacists) to annotate all clinical concepts in the narrative MAC in a quick and easy way. This tool allows the domain experts to not only resolve issues with the NED, but also involves them in the translation of concepts into computable form, by linking them to concepts defined in existing biomedical terminologies. We assessed whether this novel tool and workflow would enable faster translation of MAC into computable form, thereby mitigating this bottleneck in the automated application of MAC.

Methods

MAC Annotator is a client-side web-based program that is programmed using the jQuery JavaScript library (https://jquery.com/) and the Bootstrap framework (http://getbootstrap.com/). It allows the user to highlight parts of the narrative MAC and to associate each part with a

concept in an existing biomedical terminology. It provides basic editing capabilities to add, modify and remove annotations, and a real-time search feature allowing the users to browse concepts in existing biomedical terminologies.

To access the existing biomedical terminologies, MAC Annotator interacts with the NCBO Annotator application programming interface (API) using asynchronous JavaScript and XML (AJAX) calls. MAC Annotator first uses the API to "pre-annotate" the narrative MAC by identifying mentions of medications, diseases, signs, symptoms and problems using only three terminologies (MeSH, RxNorm and SNOMED CT). Pre-annotated terms are then highlighted in the text (**Figure 14**) and the interface allows the user to review those annotations and verify that they are correct. MAC Annotator also allows users to add, modify or delete annotations, and facilitates this by providing a straightforward search mechanism for all terminologies available (**Figure 15**). Ideally, all of the clinical and medication concepts in the narrative MAC will be annotated by the end of this process.

To evaluate MAC Annotator, we asked a group of experts to annotate a set of narrative MAC using this tool, and we made several measurements throughout this process. The measurements made in this study are summarized in **Table 13**. First, we measured how many of the concepts in the MAC could be automatically identified using the pre-annotation. To determine the accuracy of the pre-annotation we measured how many of the pre-annotated concepts needed to be modified or deleted by the annotators. We also measured how many of the concepts could only be annotated by the user (i.e. they were not identified in pre-annotation), and how many had to be defined manually (i.e. could not be associated with a concept in any existing terminology). Additionally, we measured the time spent for each of these activities.

In order to ensure that MAC Annotator facilitates achieving consensus on the definitions of the MAC, we also measured agreement between annotations. Each narrative MAC was annotated by two independent annotators, and we measured the agreement between annotators by dividing twice the number of concepts that were identically annotated by both annotators at the end of annotation to the total number of concepts identified by both annotators. This proportion can range between 0 to 100%, where higher numbers indicate more agreement between annotations. We also manually analyzed the types of disagreements in annotations.

Results

Six experts (including three pharmacists and three physicians) each annotated six narrative MAC using MAC Annotator. The average time spent to complete each annotation task was 121 seconds, including 0.50 seconds spent for automated pre-annotation. Pre-annotation resulted in identification of 75 concepts, from which experts modified seven (9%) and accepted the remainder (91%) as correct annotations. In the end, 180 concepts were annotated in the MAC, of which 107 (91%) were associated to a concept in an existing biomedical terminology while the remainder (9%) were manually defined by the experts. The agreement between annotations was 77.3%.

1	Number of concepts in the MAC that could be automatically identified
2	Number of pre-annotated concepts that needed to be modified or deleted.
3	Number of concepts that could only be identified through manual annotation
4	Number of concepts that could not be associated with existing standard terminologies
5	Time spent for each annotation activity

Table 13 – Measurements conducted in study 4.

Agreement between annotators was 100% for medication names, and 100% for medication class names. With regard to diseases and problems, agreement was exactly 50%, i.e. half of the concepts identified by one annotator were not identically annotated by the other annotator. The most common reason for disagreement was that annotators used different approaches when the disease was mentioned along with a modifier. For instance, one annotator broke down the phrase *New York Heart Association Class III or IV heart failure* to several concepts so that *heart failure* could be associated with the respective concept in SNOMED, while the other annotator used a manual definition for the entire phrase and therefore did not link it to any concepts in existing biomedical terminologies. Similarly, one annotator provided a manual definition for the term *uncomplicated pulmonary embolus* while the other annotator mapped *pulmonary embolus* to SNOMED and provided an explanation for the modifier *uncomplicated*.

For all context-dependent terms, annotators were asked to provide a manual definition that would allow substituting the concept with a set of context-independent terms in a second round of annotations. For instance, the modifiers *uncomplicated* (as a modifier before a disease name) or *monotherapy* (which can be translated as "excluding other therapies for the same disease", where the other therapies change for each disease) could be annotated with a manual definition listing all the complications and all the other therapies, respectively. While some annotators provided such definitions, others provided abstract definitions (e.g. defining *uncomplicated* as "with no complication") that would be context-agnostic but not objective or computable.

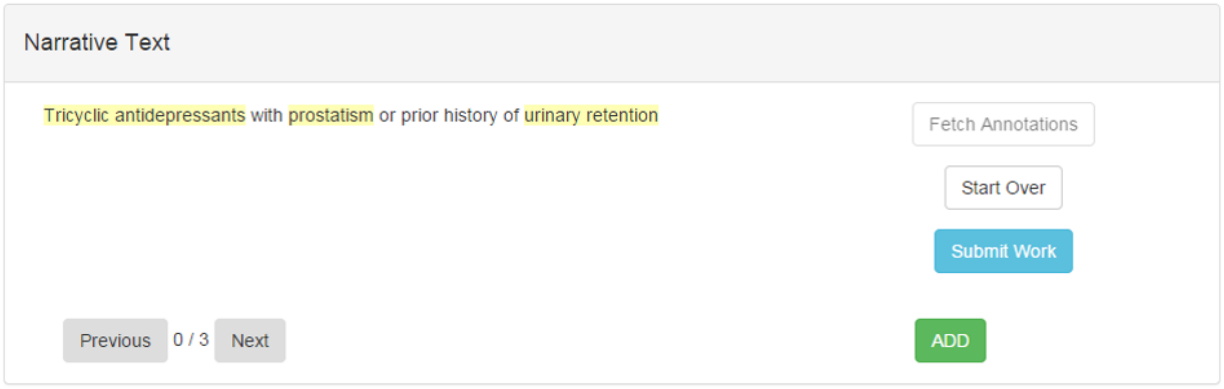


Figure 14 – Screenshot of MAC Annotator after completion of pre-annotation.

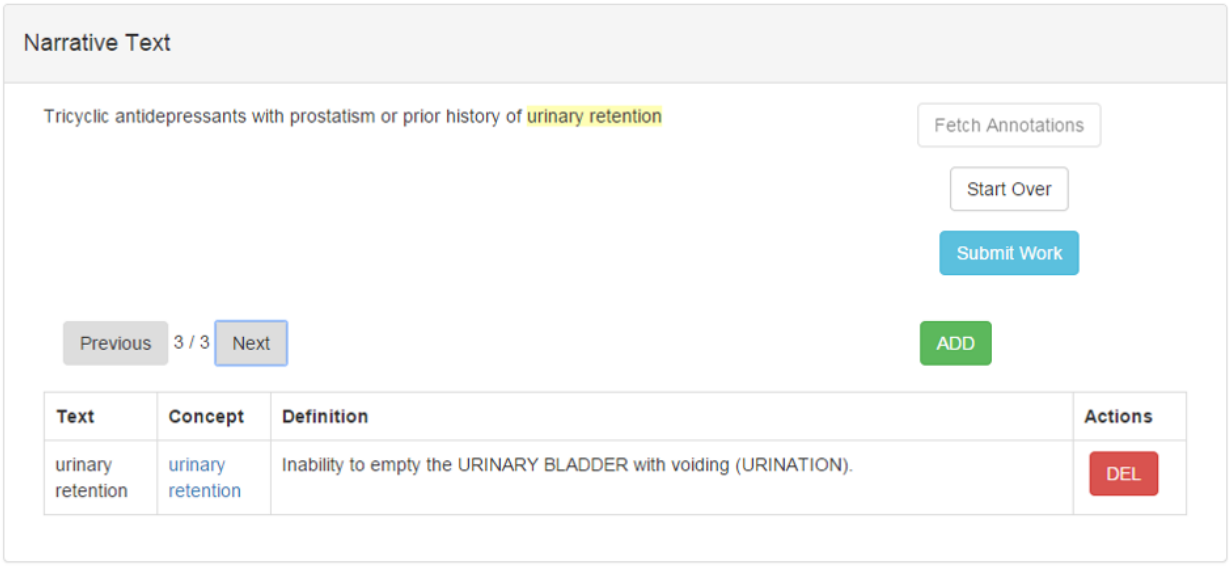


Figure 15 – Screenshot of MAC Annotator during manual annotation.

Discussion

MAC Annotator facilitated the annotation of MAC through the pre-annotation step, and by providing an easy to use, interactive interface for searching existing biomedical terminologies for relevant concepts. It also helped identify concepts that were not well-defined in the original narrative MAC.

Our results indicate that MAC Annotator can facilitate rapid annotation of MAC by domain experts, without the need to use any other software. Our results also indicate that different experts may interpret the MAC inconsistently when the MAC contain concepts that are context-dependent or not well-defined.

Traditionally, annotation of biomedical text is performed by a small number of expert annotators over extended periods of time.[185] Previous work on creating structured versions of decision support rules also relies on manual annotation.[186] MAC Annotator may help this process by speeding up the translation of these narratives into structured form. Pérez-Pérez et al have recently conducted a thorough review of public and well-known annotation tools used for biomedical text annotation.[187] Compared to existing tools, MAC Annotator has the advantage of pre-annotating the text in a time-efficient manner with high accuracy, allowing annotators to define terms that are not explicit, and platform-agnostic design which doesn't require any software installation by the user.

This work is not without limitations. First, we only conducted one iteration of annotations. Disagreements in annotations could be resolved through multiple iterations of annotation, as we have shown in our previous work.[184] However, the scope of this study was limited to facilitating each annotation session, and not the consensus building process across iterations of annotation.

Second, MAC Annotator can only access those biomedical terminologies that are available through NCBO Annotator. While this includes most of the commonly used biomedical terminologies (e.g. SNOMED, MeSH, RxNorm, or ICD-9) it does not include all existing biomedical terminologies, and it is possible that certain concepts in the MAC could be associated with concepts that are in the other terminologies. Third, MAC Annotator currently only defines the links to the concepts in the target terminology; ideally, the link should be expanded to also include the unique concept identifier for that concept as included in the Unified Medical Language System (UMLS) when possible. This can be addressed in future versions of the tool. Finally, the sample size in this study was small, and it is possible that other challenges with annotation of MAC might have been identified if a larger collection of MAC was analyzed. However, we have manually evaluated larger collections of MAC in our previous work and the challenges identified in the current study are in alignment with those identified in our previous studies.[37,184]

Conclusions

MAC Annotator facilitates the translation of MAC into structured form. It allows domain experts to more easily annotate the medical concepts in narrative MAC, and substitute the NED concepts with objective definitions. The use of a real-time NLP approach along with an interactive design enabled annotation of concepts in only a few minutes per MAC.

Chapter 5: Deploying computable appropriateness criteria

Background

Deploying computable appropriateness criteria can be discussed from three different perspectives. First are technical issues with deployment. While a formal representation of the appropriateness criteria is a necessary piece for deploying the MAC in an automated way, the data regarding the patient and the medication should also be delivered in a compatible format so that the appropriateness criteria can be applied to the data. This can be as simple as restructuring the data from one format to the other, or as complex as having to abstract certain variables from the data, e.g. using phenotype definitions or other complex information abstraction approaches.

The second issue is with regard to deployment of the appropriateness criteria in a manner consistent with the clinical workflow. Review of literature shows that many methods of delivery have been studied for implementation of appropriateness criteria, both in a manual and an automated way (for citations please see Chapter 2). Last but not least, once the appropriateness criteria are deployed, their impact needs to be measured. These measurement should not only include verifying the accuracy of the appropriateness criteria in action, but also focus on changes in various process and clinical outcomes.

Regardless of how the MAC are implemented, one of the key challenges in their implementation is to acquire the clinical data about the patient that are necessary for the rules in the MAC to be computed. This includes knowing where the information can be found and also knowing how to identify it in that source. For example, to implement a MAC targeting the use of antihypertensive drugs in the elderly, various types of information may need to be acquired and made available to the execution engine that runs the MAC; these include patient age (which can be calculated from

patient's date of birth as long as that is available and is in structured form), information regarding the administration of the specific medications or classes that are subject to that MAC (which may include structured data from the EHR as well as unstructured data such as mentions of the medication in visit notes), and establishing the fact that the patient has hypertension.[188] The latter may require applying some form of phenotyping algorithm that would utilize various forms of structured and unstructured data (including problem lists, claims data, and observations) to accurately identify presence of hypertension.

While the challenge of identifying and abstracting clinical data (including the use of phenotyping approaches) deserves tremendous attention, it is beyond the scope of this research. Nevertheless, to demonstrate that given suitable data extraction methods, it is possible to implement the MAC with high accuracy, we conducted a study in which the MAC for a single group of medications was used in computable form to identify IUM and the results were compared with manual chart review by the experts (Study 4). We used a very simple approach for identifying clinical data, which only relied on narrative reports. We also did not use the formal representation based on OMAC, but rather simply used a list of indications in computable form. However, we demonstrated that even a simple approach like this can achieve high accuracy.

A number of automated approaches have been studied for reducing IUM, and achieved various levels of success.[89] The mostly commonly used method is to provide decision support using electronic alerts within the CPOE; this method has been associated with variable levels of success in reducing IUM.[89] Indication-based prescribing is another approach which relies on a mandatory "indication" field in the CPOE system; this approach has also been inconsistently effective in reducing IUM.[80] The sole focus of these automated solutions is always the prescriber, and they all share an interruptive approach: the prescriber is interrupted during the

process of making prescriptions and required to address an alert about a potentially inappropriate prescription or specify the indication for said prescription.

A review of the literature on studies aimed at reducing inappropriate medication use shows that while feedback to the provider is a common component in these studies, it does not need to be provided in an interruptive way at the time of prescription. "Academic detailing" is an alternative approach, which is carried out as part of the workflow of patient care, but not necessarily during the time prescriptions are made. It is a pharmacist-mediated approach that aims at educating providers towards more appropriate use of medication.[82] Rigorous studies have proven that pharmacist intervention is an effective approach in reducing inappropriate prescribing[189] and multiple studies have shown that academic detailing can successfully reduce IUM.[67,85,190]

Developing automated interventions for reducing IUM can be challenging for at least two reasons: impact can vary among different patient populations, and a fully automated process can be difficult because the information necessary for this process may not readily available in a computable form.

While inappropriate use of medications can increase the potential for harm and excessive cost in any patient, its effect may be difficult to measure among heterogeneous patient populations. One approach used in more recent studies is to narrow the focus of study to patients who are subject to polypharmacy, especially the elderly.[191] Polypharmacy is commonly defined as receiving five or more medications.[192–194] By focusing on older adults that are subject to polypharmacy, these studies target the portion of population that are more likely to suffer from negative impacts of IUM, including drug-drug interactions, adverse drug reactions and increased risk of hospitalization.[170–172]

Developing and implementing a fully automated solution for reducing IUM is difficult, because much of the information that needs to be incorporated into the automated process is not readily available in a structured form that computers can understand. As an example, despite the wide use of EHR systems that allow the storage of patient's problems in a structured problem list, these problem lists are often incomplete, inaccurate and out of date.[133,195] One potential solution to this problem is to use a semi-automated process, including an automated *screening* step followed by manual *verification* by experts. In the case of IUM in the inpatient setting, this can be achieved by using an automated approach to identify the patients that are potentially subject to IUM, and presenting this information to pharmacists to review and take action if necessary. In effect, this approach combines the benefits of an automated informatics solution and the process of "academic detailing". Rigorous studies have demonstrated that pharmacist intervention is an effective approach in reducing inappropriate prescribing,[189] and the semi-automated process involving pharmacists is also previously evaluated in one study, where it was effective in reducing medication *underuse*.[196] However, this hybrid method has not been studied in reducing overuse or other types of IUM.

It is critical that approaches used to reduce IUM are specific, and can capture the "alternative" forms of IUM that may take place when one specific form of IUM is subject to an intervention. For example, in a retrospective analysis of the effects of an electronic alert on reducing inappropriate use of PPIs in the intravenous (IV) form, we not only measured the reduction in the rate of intravenous PPI use but also measured if the total PPI use was changed, and whether a switch to other antacid drugs (namely, histamine 2 receptor blockers) happened after the initiation of the intervention.[197] While we observed a decrease in the proportion of PPI orders made for the IV form (Figure X), the number of IV PPI orders that were appropriately indicated

was not significantly higher after the implementation of the alert (88.0% indicated after vs 74.0% before; $p = 0.07$) and we did not observe any concurrent change in the utilization of histamine 2 receptor blockers in the same patients.

On the back-end, various methods can be used to implement the appropriateness criteria which would then feed the information into any of the delivery methods described above. Because MAC can be inherently described both as a "guideline" for medication use and as a tool for "decision support", we analyzed how guideline execution methods and decision support systems can be used to implement the MAC. Specifically, we looked at the Guideline Execution Engine (GLEE) and OpenCDS systems, as described in Study 5 below.

Study 5: Benchmark implementation of structured MAC

Introduction

Previous studies have shown that proton pump inhibitors (PPIs), antidepressants, antipsychotics, statins, non-steroidal anti-inflammatory drugs (NSAIDs) and cyclooxygenase-2 (COX-2) inhibitors are subject to widespread overuse.[4–6] For the purpose of this study, we focused on the overuse of proton pump inhibitors (PPIs). PPIs are a group of gastric acid-suppressing drugs with high potency. Established indications for use include eradication of *Helicobacter pylori* in patients with peptic ulcers, prevention of gastric ulcers induced by NSAIDs, treatment of gastric ulcers, Zollinger-Ellison syndrome, acid-induced esophagitis, Barrett's esophagitis and severe gastroesophageal reflux disease (GERD).[7] Earlier studies showed that PPIs are more potent and effective than other acid-suppressing drugs;[8,9] consequently, PPIs have gradually replaced histamine receptor antagonists (H_2RAs) over the last two decades.[10,11] In 2009, PPIs ranked third in the sales of medicines in the United States (US); in 2010 esomeprazole (Nexium®) was second in total sales among prescription drugs.[12,13] In the US, some PPIs are now available without prescription, and their use is not reflected in the aforementioned statistics.

PPIs are frequently overused and compliance to guidelines of appropriate use has been reported to be as low as 31-33%.[4,14,15] Aside from the direct costs associated with overuse,[16,17] serious adverse effects have been reported from long-term acid suppression with PPIs. Several studies have demonstrated associations between PPIs and *Clostridium difficile* colitis.[18–20] Long-term use of PPIs may also lead to Vitamin B12 malabsorption,[21] bone fractures,[22–24] iron deficiency,[25] interstitial nephritis,[26,27] and gastric carcinoids,[28] although these associations are not supported by strong evidence.[29] PPIs may increase the risk of pneumonia,[30] and may decrease the effectiveness of clopidogrel – a frequently used anti-platelet agent – to cause fatal cardiovascular

events.[31–33] Most of these side effects of PPIs arise after long-term use of these drugs; however, PPIs are very frequently used for long periods.[4,34] Recent guidelines on PPI use include warnings regarding these long-term effects,[7] and in February 2012 the Food and Drug Administration issued a warning specifically regarding the association between PPIs and *C. difficile* colitis.[35]

Overuse of PPIs has been studied by manual review of patient records or via patient interviews.[4,11,14,15,36,37] In all these studies, the primary method has been to rule out any established indication for PPI use, thereby identifying patients receiving the drug without an appropriate indication. This requires collecting a list of all existing conditions and diseases the patient has, and comparing it to the set of established indications. Although conditions and diseases affecting a patient can be recorded in the electronic health records (EHRs) in a structured form by means of electronic problem lists, studies have shown that these sources of information are neither comprehensive nor reliable.[38–41] Thus the information needed to make the decision about the appropriateness of use of PPIs must be extracted from narrative, unstructured notes in a process that is typically difficult to automate. This issue is not unique to studies of overuse, and has been a well-known limitation for using electronic health records in clinical decision support (CDS).[42,43]

Natural language processing (NLP) can be used to convert narrative information into a computable format and has been proven to be accurate and efficient in clinical studies.[42,44] Previous studies have shown that using an NLP engine can assist with automated generation of problem lists for patients with high accuracy.[45,46] This study develops a framework for automated identification of patients who are potentially subject to medication overuse and applies the framework to the overuse of PPIs. We also aim to identify the potential challenges in

generalizing this framework to overuse associated with other groups of medications or other types of treatments.

Methods

Our framework consists of two arms (**Figure 16**): a knowledgebase of established indications for using a medication, and a list of current problems and conditions that a patient has. Here, we describe each of these components in more detail. Although we describe the process for the specific case of PPIs, the framework can similarly be used for other groups of medications.

To develop the knowledgebase of indications (**Figure 16**, top-left), we began with collecting a list of medical conditions in which PPIs are used by reviewing drug labels and clinical guidelines. We also analyzed the reasons for PPIs use reported by healthcare professionals in the Food and Drug Administration (FDA) Adverse Effects Reporting System (FAERS).[47] While the primary purpose for FAERS is to report adverse effects, one of the data fields in submissions to FAERS is the intended use of medication reported; we used this information to generate a list of common uses for PPIs. We also retrieved the indications listed for PPIs in the National Drug File Reference Terminology (NDF-RT).[48]

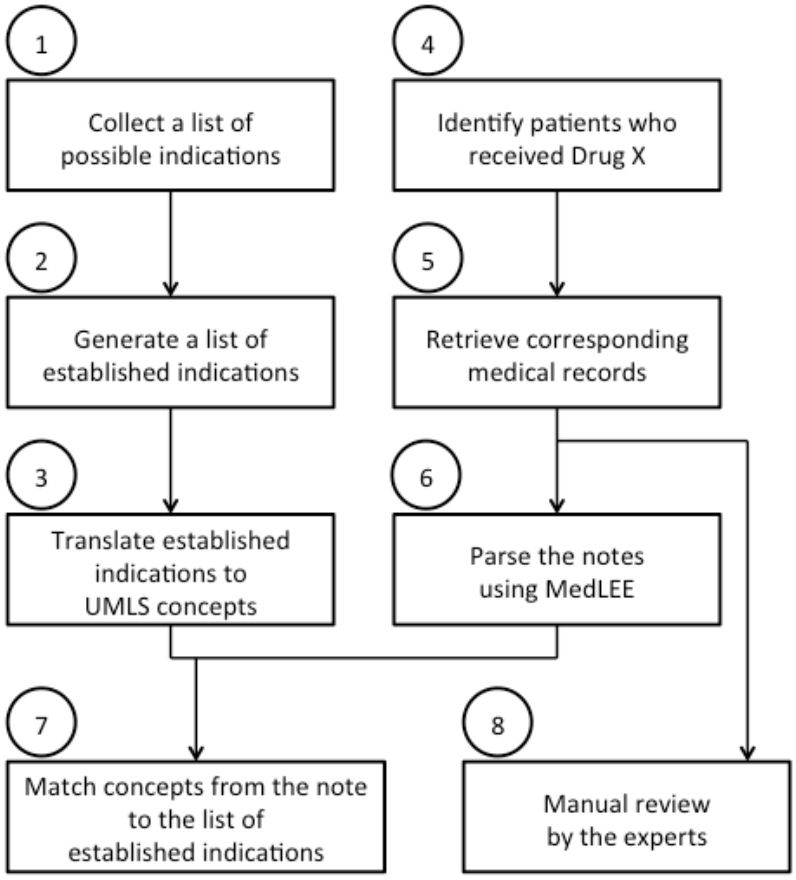


Figure 16 – The framework for identifying overuse of a medication.

Subsequently, a committee of three attending faculty gastroenterology specialists reviewed the list of possible indications with respect to the scientific evidence behind them, and selected a smaller list of all appropriate on-label and off-label indications for PPI use in adults, which we refer to as "established indications". Each of the established indications was mapped to a single corresponding concept in the Unified Medical Language System (UMLS). The UMLS is a compendium of numerous controlled vocabularies in the biomedical sciences where each concept (e.g. a disease, a drug, a procedure, or a gene) is identified using a unique identifier.[49] Mapping the indications to concepts in the UMLS ensures that these indications are coded in a

consistent form that can also be used to represent patient problems in the next step. The list of established indications of PPI use and the respective UMLS concepts is shown in **Table 14**.

The other component of our framework is to collect current problems for patients who are receiving the medication in question (in this case, PPIs) using the EHR (**Figure 16**, top-right). We identified such patients by searching records of medication orders in the EHR, and retrieved the medical records collected for those patients during their hospitalization period. Depending on the availability and quality of the EHR data, patient problems can be obtained from structured problem lists or from narrative reports. In our data set, comprehensive and reliable problem lists were not available; therefore, we used narrative reports as the only source for obtaining patient problems. For the purpose of this study, we only used discharge summaries, and we used the MedLEE (Medical Language Extraction and Encoding) NLP system to extract the concepts in the notes. MedLEE is a rule-based NLP program that is especially designed to parse clinical texts, i.e. it takes natural language clinical text as an input, and represents the data in a standardized computable form as the output to facilitate accurate retrieval of information [50]. It is capable of identifying negation and temporality identifiers, and of mapping the medical concepts found in the notes to the UMLS concepts. We generated a list of patient's current conditions coded into UMLS concepts, taking care to exclude negated conditions, family history, and conditions strictly happening in the past (i.e. a past medical history of an acute disease was not considered as a current problem, but if a chronic problem such as diabetes mellitus was mentioned in the past medical history, we included it in the patient's current problems).

In a pilot study, the accuracy of our system in identifying indications in a training set of notes was measured by comparing it to manual review by the corresponding author. We estimated our method has an agreement of 75% in predicting overuse (95% CI = 63 – 86%), and used this

information in sample size calculation for the current study. Assuming a null accuracy of 50% (pure chance) and using sample size calculation for proportions, we determined that a total sample size of 115 was needed to be able to identify accuracy of 0.75 using a type-I error threshold of 0.05 and a statistical power threshold of 0.80. Since some of the exclusion criteria (see below) could only be assessed after the sampling was done, to allow for possible exclusions and subgroup analyses without critical decline in statistical power, we used a sample size of 200. The sample used in this study did not contain any of the records or patients included in the pilot phase.

We used a computer program to randomly select 200 patients who were admitted to the New York Presbyterian Hospital in 2010 and had one or more orders to receive PPIs during admission. For patients with multiple hospitalizations in 2010, we used a computer program to randomly choose one of the hospitalizations in which PPIs were prescribed, and discarded the others. We used this data to evaluate our framework. Patients whose current conditions did not include an established indication for PPI use were classified as "overuse candidates". We call them candidates because establishing overuse is not completely possible by reviewing discharge summaries alone (even by manual review) since a valid indication might not be documented properly in the discharge summary. We used a single inclusion criterion: to have an order in the EHR to receive any PPI during their hospitalization. Our exclusion criteria included not having a discharge summary recorded in the EHR (that is because discharge summaries are usually not recorded for hospitalizations shorter than two days at our center), history of intubation during the hospitalization, and age less than 18 years.

CUI	Condition
C0013295	Duodenal ulcer
C0013298	Duodenitis
C0013299	Duodenogastric reflux
C0030920	Peptic ulcer
C0030922	Peptic ulcer hemorrhage
C0151966	Duodenal ulcer hemorrhage
C0341245	Erosive duodenitis
C0854225	Duodenal ulcer, obstructive
C2741638	Stress ulcer
C0004763	Barrett's esophagus
C0014868	Esophagitis
C0014869	Reflux esophagitis
C0017168	Gastroesophageal reflux disease
C0151970	Oesophageal ulcer
C0155789	Oesophageal varices hemorrhage
C0267055	Erosive esophagitis
C0267075	Esophagitis ulcerative
C0013395	Dyspepsia
C0017181	Gastrointestinal hemorrhage
C0038358	Gastric ulcer
C0079487	Helicobacter infection
C0235325	Gastric hemorrhage
C0237938	Gastrointestinal ulcer
C0267158	Reflux gastritis
C0267166	Gastroduodenitis
C0341163	Gastric ulcer perforation
C0341164	Gastric ulcer hemorrhage
C0343378	Helicobacter gastritis
C2010560	Gastritis hemorrhagic
C2243088	Gastritis erosive
C0041909	Upper gastrointestinal hemorrhage

Table 14 – Reference list of established indications for proton pump inhibitors.

A gastroenterologist manually reviewed the notes and identified whether an established indication was present. A gastroenterology fellow independently reviewed 40 of these notes so that we could assess inter-rater reliability. We facilitated the review process by providing the

experts with an interface that helped them browse the notes in a secured and encrypted environment and record the indications they found in the notes using a free-text input box. Our reviewers were not involved in the development of the framework. We measured inter-rater reliability using concordance and Cohen's Kappa.[51] We finally compared the output of the framework with the results of manual review by the experts to calculate the sensitivity and specificity of our method. The Institutional Review Board of Columbia University approved the protocol of this study.

Results

From the original 200 notes selected for manual review by experts, 23 were excluded because they were associated with children. We also excluded 40 notes because they were associated with patients with a history of intubation. There is evidence suggesting that PPIs may be beneficial for preventing stress ulcers in patients receiving prolonged intubation [52], but guidelines are vague and the decision is generally made on a case-to-case basis. All remaining 137 notes were manually reviewed by experts and categorized by them as either "appropriate use" or "overuse candidate". The inter-rater reliability of two reviewers was high (agreement = 92.1%, Cohen's κ = 0.773). **Table 15** summarizes the baseline characteristics of patients included in this study. It should be noted that distribution is impacted by the choice of PPI recommended by the hospital formulary.

Out of all 137 notes reviewed, only 43 contained an indication based on manual review. We measured the accuracy of the system using the well-known performance metrics sensitivity and specificity. Sensitivity is defined as TP/(TP+FN) where TP is the number of times the system correctly found that there is an appropriate indication, and FN is the number of times the system did not find an appropriate indication but manual review did. Likewise, specificity is defined as

TN/(TN+FP) where TN is the number of times the system correctly identified the absence of appropriate indications, and FP is the number of times the system found an appropriate indication but manual review did not.

Demographics	
Age (years)	65.5 (19.8)
Gender	
Female	79 (58%)
Male	58 (42%)
Ethnicity	
African-American	17 (12%)
Caucasian	39 (28%)
Hispanic	32 (23%)
Other*	49 (36%)
Medication use	
Esomeprazole	116 (85%)
Pantoprazole	6 (4%)
Omeprazole	2 (2%)
Lansoprazole	1 (1%)
Multiple PPIs	12 (9%)
Total	137 (100%)

Table 15 – Baseline characteristics of patients included in the study.

Our framework identified indications in 37 of the notes and in comparison to the manual review it had a moderate sensitivity (74%, 95%CI = 59% – 86%) and a high specificity of (95%, 95%CI = 87% – 98%). In cases that were not classified as overuse candidates, the number of indications returned by our framework ranged from 1 to 4 (mean = 1.43, standard deviation = 0.74, median = 1, mode =1). In 28 cases, only one indication was returned. In all true positive cases, the indication documented by manual review was correctly returned by the framework. There was no difference in the number of indications returned for true positives versus false positives (Mann-

Whitney P-value = 0.3491). The main reasons for receiving PPIs included GERD (17 cases), gastrointestinal hemorrhage (13 cases) and peptic ulcer (4 cases).

The mean ± standard deviation of the time spent by manual reviewers reviewing each discharge summary was 83 ± 52 seconds. Processing the same discharge summaries using the NLP system took 1.88 ± 1.38 seconds, and the matching query took 27 milliseconds on average (all measurements were done on a commodity server with 43 GB of memory and one 4-core 2.93 GHz processor).

Discussion

We were able to create a framework using NLP that can be used for automated identification of established indications of medication use in narrative reports with high accuracy, and an automated framework to identify overuse candidates. The success of our approach is a product of the comprehensiveness of the list of indications, accuracy of extracting patients' problems, completeness of documentation, and ability to translate the established indications determined by clinicians into unique concepts in the knowledgebase.

Identifying the reasons for the prescription of a medication using NLP was previously studied by multiple groups of researchers as part of the Institute for Integrating Biology and the Bedside (I2B2) "medication extraction challenge". Studies conducted in response to that challenge all showed a low performance in extracting the reasons for prescription of medications, with their F-measure ranging from 0.03 to 0.525.[53,54] In contrast, our system was able to identify the indications for PPI administration in adults with high sensitivity and specificity (F-measure = 0.80). While our results are not directly comparable with those studies because they are obtained using different data sets, we believe the higher accuracy of our approach has in part resulted

from intentionally restricting our search for certain concepts (i.e. the list of established indications). Another study in which the search for medication-problem associations was limited to those previously described in a knowledgebase also yielded comparable results (sensitivity of 67.5% and specificity of 86%).[48]

Given the subjective nature of the task as illustrated by the inter-rater agreement of 92% on the reference standard, any system would be unlikely to be able to perform with 100% accuracy. Therefore, we further analyzed the output of our framework to investigate the false positives. We found that at least in 2 of the 5 false positive cases, a second reviewer could interpret the indication found by the system as appropriate. Other reasons for false positives included differential diagnoses that were suggested in one section of the note (e.g. "*patient complained of epigastric pain during admission, likely secondary to GERD*") and ruled out later in the note, indicating that more complex reasoning concerning the section of note where the information occurred would be required to determine appropriateness of medication use. Another cause of error was the NLP errors in interpreting temporality (e.g. *history of gastritis* from several years ago was incorrectly identified as an ongoing chronic event).

Similarly, false negatives were predominantly cases in which a manual reviewer documented an indication that was not in the original list of appropriate indications (e.g. hiatus hernia, coagulopathy, post-surgical ileus, and gastric outlet obstruction), an inevitable consequence of lack of expert agreement on an inclusive list of indications. Most of the remaining false negatives were the result of NLP limitations; for instance, the term "*GI ulcer*" was not recognized as "*peptic ulcer disease*" because it was not obvious to the NLP system that the location of bleed was in the upper gastrointestinal tract (as determined by the expert reviewers). Additionally, in some cases the decision about the appropriateness of PPI use was unusually too complex to be

implemented in our framework. For instance, in one patient who had a low risk for gastric ulcer, the expert reviewer determined that prescribing a PPI prophylactically would be appropriate because that patient was a Jehovah's Witness who would be likely to refuse transfusions if significant gastric bleeding occurred.

There were some limitations in our work. First, we processed only discharge summaries for possible indications. Further indications might be revealed in other documents such as admission notes or progress notes. Some researchers have also expanded the information used to identify indications by complementing the data from medical records with data collected directly from the patients, for example through interviews.[4] While this may serve to eliminate under-documentation of patient's conditions, it introduces recall bias and precludes automation.

Our method is unable to distinguish overuse from under-documentation. While the majority of indications for PPI use are severe and significant events (such as gastrointestinal bleeds, or peptic ulcers), one can still expect that some less serious indications (e.g. GERD) may not be documented in the medical records if they are not the primary reason for admission, or if the patient started receiving PPIs long before they were admitted to the hospital. However, this issue also affects manual review of the notes by experts, and should not affect the results of our evaluation. Last, our method currently ignores indirect indications; for instance, if the PPIs were prescribed solely to prevent or diminish the gastro-irritant effects of a co-prescribed medication (e.g. NSAIDs, steroids, aspirin, or alendronate) our framework would mark this as overuse.

The performance of the system could be improved by promoting better documentation, extending the list of indications to include less frequent indications, and introducing more complex logic into the process of identifying patient problems. We also anticipate that the performance would be higher if structured problem lists were in use and were kept up to date. In

future work, we will try to address these limitations and demonstrate the generalizability of this framework by applying it to other cases of overuse.

In summary, our automated framework compares favorably with expert manual review. In the future, this framework could provide clinical decision support for identifying overuse of medications both to reduce overuse and to encourage better documentation of indications.

Study 6: Utilizing the Guideline Execution Engine (GLEE) and OpenCDS for implementation of structured MAC

Background

Clinical practice guidelines are intended to improve the quality and cost effectiveness of patient care by fostering best practices.[198] Similarly, medication appropriateness criteria (MAC) intend to foster appropriate use of medications in ways that would lead to improve outcomes and reduce cost and harm. Therefore, it can be argued that MAC can be viewed as a special form of clinical guidelines. Under this proposition, it would be reasonable to consider representing the MAC using guideline representation languages.

In this study, we intended to assess whether it is possible to represent the MAC using guideline representation languages, and if so, would representing the MAC in this form be efficient. We conducted this study using two guideline modeling methodologies, namely the Guideline Interchange Format (GLIF) Version 3[199] and OpenCDS.[183] These models were selected because we wanted to assess issues with representation of MAC as well as issues with implementation of MAC, since the latter is only possible if an interpreter for the guideline representation language is available. There is a proprietary interpreter for GLIF called Guideline Execution Engine (GLEE).[177] In contrast, OpenCDS comes with a set of free and open-source tools for implementing the decision support guidelines. We had access to both these executables therefore we focused our study to guideline modeling methodologies that have the potential to be implemented in real clinical settings.

Methods

We first analyzed the examples of guidelines that are bundled with GLIF version 3, as well as those described in the numerous publications about this resource.[177,198–200] Subsequently, we assessed if and how MAC can be implemented into GLIF format. Our analysis focused on whether all aspects of MAC can be defined using GLIF, i.e. we identified which aspects of MAC are not readily captured by GLIF and therefore would require extending GLIF. Next, we analyzed how GLEE can be used to implement such guidelines. Here, our analysis focused on whether it is possible to link GLEE to an arbitrary data source, and specifications of the data that needs to be provided to GLEE.

Similar steps were used for analyzing OpenCDS. We first started by analyzing the decision support rules that were already bundled with OpenCDS. Next we assessed whether all aspects of MAC can be defined using OpenCDS concepts and predicates. We also assessed how OpenCDS can connect to an arbitrary data source and the specifications of the data the needs to be provide to it.

Results

Analysis of GLIF

In GLIF, guidelines are represented by nesting *steps* of a guideline where each step may constitute an *action*, *decision*, or *branching/synchronization*. When the steps are linked, they form an *algorithm* and each guideline itself may comprise of one or more algorithms (*subguidelines*). Figure 17 shows the components of the GLIF representation.

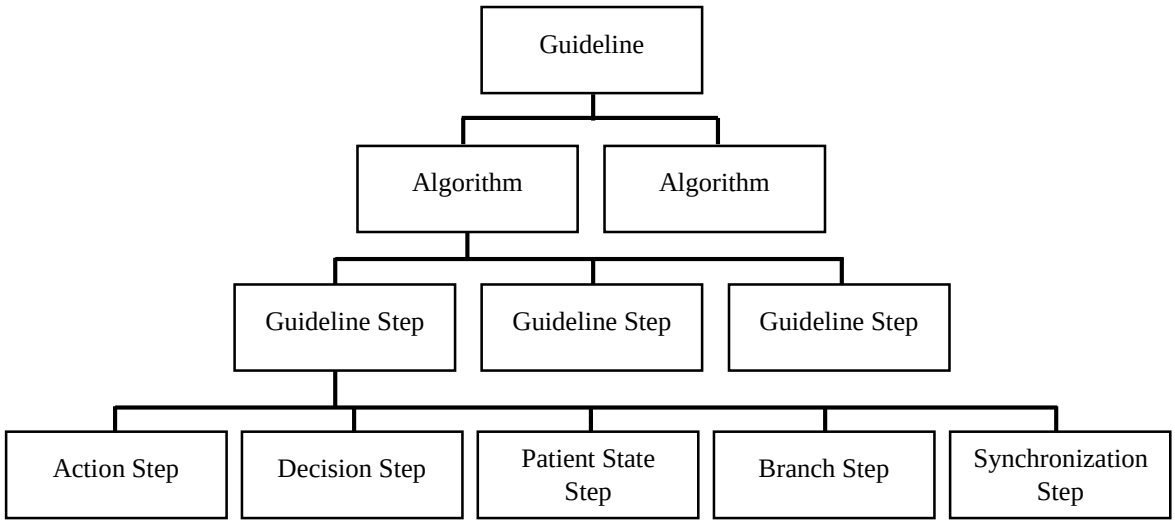


Figure 17 – Components of Guideline Interchange Format (GLIF).

By design, GLIF focuses only on representing the guideline steps, and therefore it only defines the steps in the highest level (e.g. only discriminating between "decisions" and "actions"). For those steps that involved input of data (e.g. Patient State step and Decision step), a "variable" is defined in GLIF which would be later linked to a data source within GLEE. GLEE plays two roles: it walks through the guideline and executes each step (which includes executing the branching and synchronization steps), and it also reads the input from a data source and returns the output back to a similar data source. In its current version, GLEE can only read from and write into plain text files. For each variable defined in the guideline, one plain text file with a similar name must be placed in the appropriate directory. GLEE will take the contents of that file as an input; it can also apply simple arithmetic functions (such as equality, addition, subtraction, etc.) for numeric variables. Through the use of branching steps in GLIF, GLEE can also apply simple "AND" and "OR" logic to the steps. However, it cannot perform any higher level predicates (such as aggregation) nor does GLIF support any specification for the type of the data (e.g. it does not discriminate diagnoses from medications, and doesn't verify if the type of

information passed in a variable matches the desired type). From GLEE's perspective, variables can only contain numbers, strings or true/false statements.

Figure 18 shows how GLIF could be used to represent the same criteria that were used in Study 4 to identify inappropriate use of PPIs. The guidelines starts with a *patient state* step that is triggered when the patient receives an order for a PPI medication. The next step will search the patient records to identify any appropriate indications for PPI use. This is followed by a *decision step* which checks if at least one appropriate indications has been found. If none is found, an *action step* will be executed that will mark the patient is potentially subject to overuse of PPIs.

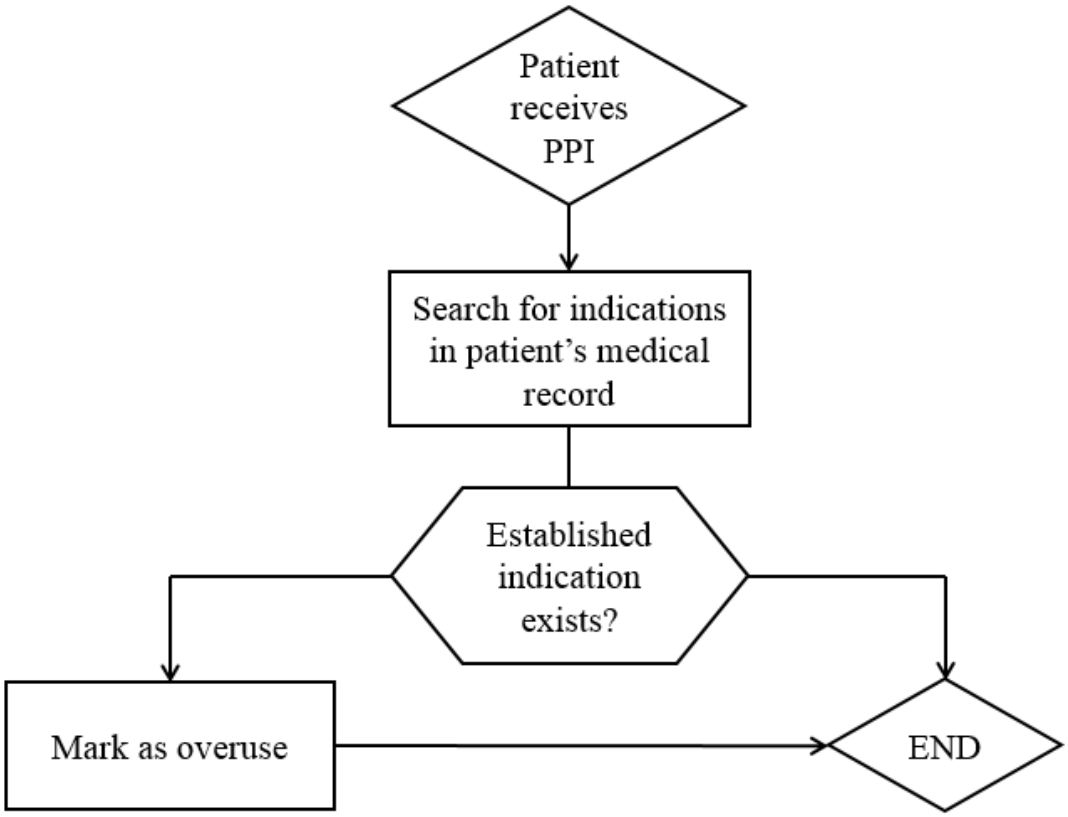


Figure 18 – GLIF representation for guidelines for identifying IUM.

The decision step in the guideline above can be further represented using a sub-guideline that defines the indications and combines them using the "OR" operation. However, GLIF doesn't enforce any specifications as to how the indications should be defined, and how the medication ordering event should be defined. For instance, if a more sophisticated guideline requires specifying the dose form or strength, GLIF does not have any built-in solutions for representing these modifiers and therefore needs to be extended to support them.

OpenCDS takes a different approach in several ways. First, OpenCDS does not use a custom guideline representation model for medical concepts; instead, it uses an existing "business rule management" system called Drools,[201] which is written in Java, and extends it to support medical concepts as input. Similar to GLIF and other guideline representation models, Drools allows defining a flow of decisions, with branching and synchronization steps. The Drools-based flowcharts in OpenCDS are encoded and executed using jBPM, which is an open-source workflow engine written in Java. Each of the steps in the workflow will then be linked to one or more "rules" that are specified using domain-specific language (DSL); the DSL is also adopted from Drools, but they have been expanded in the OpenCDS to also accept concepts from the OpenCDS terminology.

Every medical concept is locally defined in an instance of the OpenCDS terminology. This terminology will also provide mappings to other known terminologies. This is notably different from GLIF where the only three data types (numeric, Boolean or string) are accepted as input. In OpenCDS medications, diseases, classes of drugs, patient characteristics, etc. can be defined using the OpenCDS terminology. OpenCDS also comes with an execution engine. This execution engine is capable of accepting data using web service calls using the Simple Object Access Protocol (SOAP). On the backend, the rules are compiled and stored as "knowledge

modules" each having a unique identifier. On the front-end, the program that accesses the web service should specify which version of which rule is intended to be called. The clinical data is also enclosed with the web service call using the Virtual Medical Records (VMR) format,[182] which is a standard data model for supporting clinical decision support that is sponsored by Health Level 7 (HL7). OpenCDS server can then parse the input in VMR format to extract all clinical data and extract the variables necessary for executing of the specified rules.

As an example, to execute the criteria from Study 4 using the OpenCDS tools, first all medical records needed to be converted into the VMR format. Next, the criteria needed to be defined using DSLs in the OpenCDS. Unlike GLIF, the DSLs in OpenCDS do have built-in support for certain modifiers (such as time interval between two events), but they still do not have specifications for dose form, route, strength etc. In fact, the OpenCDS vocabulary only defines the medication as the main ingredient level, therefore its application for specific dose forms and strength requires expanding the vocabulary to contain those concepts.

Discussion

Our analysis of GLIF shows that while it can support representing the "flow" of decisions of a MAC, it does not provide the necessary level of detail that would be needed to represent MAC in structured form. In other words, because GLIF does not use any data models for clinical concepts, representing MAC using GLIF would require selecting such data models, and expanding GLIF specifications to include references to those data models. GLIF and GLEE will not be able to provide any data type validation beyond basic types, and unlike OpenCDS, GLEE is not capable of parsing clinical data in a conglomerate format (such as VMR). In summary, to encode MAC using GLIF and execute them using GLEE, two additional steps would be necessary: expanding GLIF to include specifications for the components of MAC, and adding

another layer between GLEE and the clinical data which would carry out the abstraction of clinical data and extraction of the variables used by GLEE. Arguably, the same amount of work can lead to creation of a novel, more consistent guideline execution system which completely supports representation and implementation of MAC.

Our analysis of OpenCDS indicates that it has several advantages over GLIF, namely the use of a data model and the ability to process clinical data in form of VMR reports. However, because most clinical data repositories currently store data in formats other than the VMR format, using OpenCDS for implementing MAC will require adding an additional layer to the process which would access the clinical data in its original format and transform it into VMR format. Similarly, the output of OpenCDS is also in VMR form and the interfacing program will require to use some form of abstraction to identify the actionable pieces of this output (e.g. identify a textual recommendation and show it in an alert in the EHR). Finally, while OpenCDS allows for specifying various clinical concepts (such as medications, diseases, dose forms, etc.) these specifications are not part of the OpenCDS itself, but they are defined as DSL rules which then connect those concepts to the appropriate VMR data type. This means to encode MAC into OpenCDS in an efficient way, first a library of rules that define all of the elements of MAC (essentially, a DSL representation of OMAC) should be encoded into OpenCDS form and then used to represent the MAC itself.

Overall, our analysis indicates that both GLIF and OpenCDS provide a solid abstraction at the higher level but do not provide the necessary details for representing and implementing MAC. In both cases, additional steps are necessary to augment these representations to support MAC.

Limitations

The major limitation of this study is that it only uses one guideline modeling methodology (GLIF), and it could be argued that the results would differ using another methodology. However, a review by Peleg *et al*[202] shows that six different commonly-used guideline modeling methodologies (namely, Asbru[203], EON[178], GLIF[199], GUIDE[204], PRODIGY[205], and PRO*forma*[179]) all use similar components to represent the actions, decisions, branching, etc. and therefore our results are potentially generalizable to those methodologies as well.

Conclusions

One of the key limitations of both GLIF and OpenCDS in our use case is that the components of MAC are not readily defined in either model. However, this can be addressed by using OMAC. An additional challenge is that both GLIF and OpenCDS expect the input data (clinical data) to be provided in a very specific format (i.e. plain text files containing variable values, and VMR artifacts, respectively). Therefore implementation of MAC would require creating tools that would translate the clinical data into the format accepted by the GLEE or OpenCDS. This is a classical interoperability challenge and falls outside the scope of this thesis.

Chapter 6: Conclusions and future work

Conclusions

This thesis investigates the problem of IUM and methods needed for automated identification of IUM using electronic health care data. While several approaches for reducing IUM exist (Chapter 2), demonstration of automated applications of all of those approaches relies on the availability of a common framework and formal representation for appropriateness criteria. In this thesis, such a formal representation model was created (Chapter 3). Using this framework, concepts in the MAC can be linked to existing standard biomedical terminologies, and the logic can be represented using standard logical statements.

Additionally, because "indications" are an important part of MAC, we systematically reviewed public medication-indications KBs to identify the advantages and deficiencies of each KB. We identified the knowledge gaps that should be addressed in future work in that area. Some of these limitations can be addressed by using MAC instead of binary medication-indication KBs. This thesis provides the foundational work for implementing MAC in electronic form, into the clinical workflow.

Additionally, a workflow was developed that facilitates translating the existing MAC into this formal structured representation as well as handling issues with concepts that are not explicitly defined (Chapter 4). The utility of formally represented MAC in clinical application was demonstrated (Chapter 5) and an analysis of advantages and limitations of existing informatics solutions for implementing the formally represented MAC, including guideline execution engines and clinical decision support systems, was conducted (Chapter 5).

Impact in Biomedical Informatics

We developed a formal representation for appropriateness criteria. This representation model is primarily design for MAC, but can also be used for formal representation of other clinical appropriateness criteria.

We developed an evaluation method for assessing the coverage of medication-indication knowledgebases for complexities of indication knowledge. This tool, along with the knowledge gaps identified in existing medication-indication KBs, can shed light on the future directions of research and development of such KBs.

We developed a real-time annotation tool that can be used not only for translation of MAC into structured form, but also for annotating clinical guidelines, clinical trial eligibility criteria, and other similar biomedical narratives in a quick and easy way.

We demonstrated the challenges and shortcomings of guideline representation models and clinical decision support tools in implementing the MAC. This enables enhancing these systems to gain better coverage for medication-indication KBs.

Impact in Biomedicine

We developed and demonstrated the complete workflow of translating MAC into computable form, implementing them in an automated way, and evaluating its accuracy and impact.[37] Tools and methodologies that were developed as part of this research have contributed to clinical studies on evaluating the impact of electronic alerts in reducing IUM in real clinical setting,[197] and assessing the adverse effects associated with frequently overused medications.[206] Finally, we curated a collection of published MAC, which can enhance clinical research on identification and reduction of IUM.

Future Work

Although we developed a formal representation for MAC and developed a workflow for translating MAC into structured form, the structured MAC is still one step away from being executable. In future work, a repository of executable MAC should be developed to facilitate its implementation across different settings. Specifically, future work will include encoding the structured MAC into executable form using the tools from OHDSI. Each MAC will be encoded as a cohort definition where the concepts are normalized to the Observational Medical Outcome Partnership (OMOP) standard vocabulary, and the logic will be encoded as a cohort definition that is stored in JSON format inside the CIRCE tool (which is an OHDSI tool for cohort definitions). This allows plug-and-play use of the MAC in all data sources that have been exposed using the OHDSI data model.

Bibliography

1. Sahlan S, Wollny A, Brockmann S, Fuchs A, Altiner A. Reducing unnecessary prescriptions of antibiotics for acute cough: adaptation of a leaflet aimed at Turkish immigrants in Germany. *BMC Fam Pract.* 2008 Jan;9:57.

2. Craig DGN, Thimappa R, Anand V, Sebastian S. Inappropriate utilization of intravenous proton pump inhibitors in hospital practice--a prospective study of the extent of the problem and predictive factors. *QJM.* 2010 May;103(5):327–35.

3. Parente F, Cucino C, Gallus S, Bargiggia S, Greco S, Pastore L, et al. Hospital use of acid-suppressive medications and its fall-out on prescribing in general practice: a 1-month survey. *Aliment Pharmacol Ther.* 2003 Jun 15;17(12):1503–6.

4. Issa I a, Soubra O, Nakkash H, Soubra L. Variables associated with stress ulcer prophylaxis misuse: a retrospective analysis. *Dig Dis Sci.* 2012 Oct;57(10):2633–41.

5. Eid SM, Boueiz A, Paranji S, Mativo C, Landis R, Abougergi MS. Patterns and predictors of proton pump inhibitor overuse among academic and non-academic hospitalists. *Intern Med.* 2010 Jan;49(23):2561–8.

6. Thomson F, Masters IB, Chang a B. Persistent cough in children and the overuse of medications. *J Paediatr Child Health.* 2002 Dec;38(6):578–81.

7. Kapur PA. Pharmacy Acquisition Costs: Responsible Choices Versus Overutilization. 1994;617–618.

8. Lindley CM, Tully MP, Paramsothy V, Tallis RC. Inappropriate medication is a major cause of adverse drug reactions in elderly patients. *Age Ageing.* 1992 Jul;21(4):294–300.

9. Smith M, Saunders R, Stuckhardt L, Mcginnis JM. Best care at lower cost: The path to continuously learning health care in America. Institute of Medicine; 2012.

10. Naunton M, Peterson GM, Bleasel MD. Overuse of proton pump inhibitors. *J Clin Pharm Ther.* 2000 Oct;25(5):333–40.

11. Conti R, Busch AB, Cutler DM. Overuse of antidepressants in a nationally representative adult patient population in 2005. *Psychiatr Serv.* 2011 Jul;62(7):720–6.

12. Autier P, Creplet J, Vansant G, Brohet C, Paquot N, Muls E, et al. The impact of reimbursement criteria on the appropriateness of “statin” prescribing. *Eur J Cardiovasc Prev Rehabil.* 2003 Dec;10(6):456–62.

13. Mainous AG, Hueston WJ, Love MM, Evans ME, Finger R. An evaluation of statewide strategies to reduce antibiotic overuse. *Fam Med.* 2000 Jan;32(1):22–9.

14. Kester L, Stoller JK. Ordering respiratory care services for hospitalized patients: practices of overuse and underuse. *Cleve Clin J Med.* 1992;59(6):581–5.

15. Korenstein D, Falk R, Howell EA, Bishop T, Keyhani S. Overuse of health care services in the United States: an understudied problem. *Arch Intern Med.* 2012 Jan 23;172(2):171–8.

16. Chan KS, Chang E, Nassery N, Chang H-Y, Segal JB. The State of Overuse Measurement: A Critical Review. *Med Care Res Rev.* 2013 Jun 26;

17. Ahwah I, Karpiel M. Using profiling for cost and quality management in the emergency department. *Healthc Financ Manage.* 1997 Jul;51(7):48, 50–3.

18. Sketris I, Ingram EL, Lummis H. Strategic opportunities for effective optimal prescribing and medication management. *Can J Clin Pharmacol.* 2009 Jan;16(1):e103–25.

19. Lysdahl KB, Børretzen I. Geographical variation in radiological services: a nationwide survey. *BMC Health Serv Res.* 2007 Jan;7:21.

20. Behnke LM, Solis A, Shulman SA, Skoufalos A. A Targeted Approach to Reducing Overutilization: Use of Percutaneous Coronary Intervention in Stable Coronary Artery Disease. *Popul Health Manag.* 2012 Oct 31;

21. Emanuel EJ, Fuchs VR. The Perfect Storm of Overutilization Overutilization of Health Care. 2013;299(23):2789–2791.

22. Shapiro DW, Lasker RD, Bindman AB, Lee PR. Containing costs while improving quality of care: the role of profiling and practice guidelines. *Annu Rev ....* 1993 Jan;14:219–41.

23. Berwick DM, Hackbarth AD. Eliminating waste in US health care. *JAMA.* 2012 Apr 11;307(14):1513–6.

24. Ciric IS. U.S. Health Care: A Conundrum and a Challenge. *World Neurosurg.* 2012 Nov 3;(January):1–8.

25. Wang H, Zhang L, Yip W, Hsiao W. An experiment in payment reform for doctors in rural China reduced some unnecessary care but did not lower total costs. *Health Aff (Millwood).* 2011 Dec;30(12):2427–36.

26. Keyhani S, Falk R, Howell EA, Bishop T, Korenstein D. Overuse and systems of care: a systematic review. *Med Care.* 2013 Jun;51(6):503–8.

27. Douglas PS, Khandheria B, Stainback RF, Weissman NJ, Brindis RG, Patel MR, et al. ACCF/ASE/ACEP/ASNC/SCAI/SCCT/SCMR 2007 appropriateness criteria for transthoracic and transesophageal echocardiography: a report of the American College of

Cardiology Foundation Quality Strategic Directions Committee Appropriateness Criteria Working Group. *J Am Coll Cardiol.* 2007 Jul 10;50(2):187–204.

28. Patel MR, Bailey SR, Bonow RO, Chambers CE, Chan PS, Dehmer GJ, et al. ACCF/SCAI/AATS/AHA/ASE/ASNC/HFSA/HRS/SCCM/SCCT/SCMR/STS 2012 appropriate use criteria for diagnostic catheterization: a report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, Society for Cardiovascular Angiography and. *J Thorac Cardiovasc Surg.* 2012 Jul;144(1):39–71.

29. Park RE, Fink a, Brook RH, Chassin MR, Kahn KL, Merrick NJ, et al. Physician ratings of appropriate indications for six medical and surgical procedures. *Am J Public Health.* 1986 Jul;76(7):766–72.

30. Hanlon JT, Schmader KE, Samsa GP, Weinberger M, Uttech KM, Lewis IK, et al. A method for assessing drug therapy appropriateness. *J Clin Epidemiol.* 1992 Oct;45(10):1045–51.

31. Beers MH. Explicit criteria for determining potentially inappropriate medication use by the elderly. An update. *Arch Intern Med.* 1997 Jul 28;157(14):1531–6.

32. Gallagher P, Ryan C, Byrne S. STOPP (screening tool of older person's prescriptions) and START (screening tool to alert doctors to right treatment). Consensus validation. *Int J Clin Pharmacol Ther.* 2008 Feb;46(2):72–83.

33. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs.* 2000 Oct;32(4):1008–15.

34. Observational Health Data Science and Informatics (OHDSI) [Internet]. Available from: http://ohdsi.org

35. Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. *Syst Rev.* 2012;1(1):10.

36. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol.* 2006 Jan;3(2):77–101.

37. Salmasian H, Freedberg DE, Abrams JA, Friedman C. An automated tool for detecting medication overuse based on the electronic health records. *Pharmacoepidemiol Drug Saf.* 2012 Dec 12;

38. Fournier J, DeRubeis R. Antidepressant drug effects and depression severity. *JAMA J ….* 2010;303(1):47–53.

39. Fick DM, Cooper JW, Wade WE, Waller JL, Maclean JR, Beers MH. Updating the Beers criteria for potentially inappropriate medication use in older adults: results of a US consensus panel of experts. *Arch Intern Med.* 2003;163(22):2716–24.

40. Oborne CA, Hooper R, Swift CG, Jackson SHD. Explicit, evidence-based criteria to assess the quality of prescribing to elderly nursing home residents. *Age Ageing.* 2003 Jan;32(1):102–8.

41. Ahrens D, Behrens G, Himmel W, Kochen MM, Chenot J-F. Appropriateness of proton pump inhibitor recommendations at hospital discharge and continuation in primary care. *Int J Clin Pract.* 2012 Aug;66(8):767–773.

42. Oborne C a, Batty GM, Maskrey V, Swift CG, Jackson SH. Development of prescribing indicators for elderly medical inpatients. *Br J Clin Pharmacol.* 1997 Jan;43(1):91–7.

43. Wenger NS, Shekelle PG. Assessing care of vulnerable elders: ACOVE project overview. *Ann Intern Med.* 2001 Oct 16;135(8 Pt 2):642–6.

44. Kaufmann CP, Tremp R, Hersberger KE, Lampert ML. Inappropriate prescribing: a systematic overview of published assessment tools. *Eur J Clin Pharmacol.* 2013 Sep 10;

45. Spinewine A, Schmader KE, Barber N, Hughes C, Lapane KL, Swine C, et al. Appropriate prescribing in elderly people: how well can it be measured and optimised? *Lancet.* 2007 Jul 14;370(9582):173–84.

46. Harris CM, Jarman B, Woodman E, White P, Fry JS. Prescribing--a suitable case for treatment. *J R Coll Gen Pract Occas Pap.* 1984 Jan;(24):1–39.

47. Meyer TJ, Van Kooten D, Marsh S, Prochazka a V. Reduction of polypharmacy by feedback to clinicians. *J Gen Intern Med.* 1989;6(2):133–6.

48. Rokstad K, Straand J, Fugelli P. Can drug treatment be improved by feedback on prescribing profiles combined with therapeutic recommendations? A prospective, controlled trial in general practice. *J Clin Epidemiol.* 1995 Aug;48(8):1061–8.

49. Hershey CO, Porter DK, Breslau D, Cohen DI. Influence of simple computerized feedback on prescription charges in an ambulatory clinic. A randomized clinical trial. *Med Care.* 1986 Jun;24(6):472–81.

50. Lassen LC, Kristensen FB. Peer comparison feedback to achieve rational and economical drug therapy in general practice: a controlled intervention study. *Scand J Prim Health Care.* 1992 Mar;10(1):76–80.

51. Naughton C, Feely J, Bennett K. A RCT evaluating the effectiveness and cost-effectiveness of academic detailing versus postal prescribing feedback in changing GP antibiotic prescribing. *J Eval Clin Pract.* 2009 Oct;15(5):807–12.

52. Schectman JM, Kanwal NK, Schroth WS, Elinsky EG. The effect of an education and feedback intervention on group-model and network-model health maintenance organization physician prescribing behavior. *Med Care.* 1995 Feb;33(2):139–44.

53. Søndergaard J, Andersen M, Vach K, Kragstrup J, Maclure M, Gram LF. Detailed postal feedback about prescribing to asthma patients combined with a guideline statement showed no impact: a randomised controlled trial. *Eur J Clin Pharmacol.* 2002 May;58(2):127–32.

54. Vægter K, Wahlström R, Wedel H, Svärdsudd K. Effect of mailed feedback on drug prescribing profiles in general practice: a seven-year longitudinal study in Storstrøm County, Denmark. *Ups J Med Sci.* 2010 Nov;115(4):238–44.

55. Haaijer-Ruskamp FM, Denig P. Impact of feedback and peer review on prescribing. *Occas Pap R Coll Gen Pract.* 1995 Feb;(69):13–9.

56. Zwar N, Wolk J, Gordon J, Sanson-Fisher R, Kehoe L. Influencing antibiotic prescribing in general practice: a trial of prescriber feedback and management guidelines. *Fam Pract.* 1999 Oct;16(5):495–500.

57. Lagerløv P, Loeb M, Andrew M, Hjortdahl P. Improving doctors' prescribing behaviour through reflection on guidelines and prescription feedback: a randomised controlled study. *Qual Health Care.* 2000 Sep;9(3):159–65.

58. Nilsson G, Hjemdahl P, Hässler a., Vitols S, Wallén NH, Krakau I. Feedback on prescribing rate combined with problem-oriented pharmacotherapy education as a model to improve prescribing behaviour among general practitioners. *Eur J Clin Pharmacol.* 2001 Jan 1;56(11):843–848.

59. Elliott R a, Woodward MC, Oborne CA. Antithrombotic prescribing in atrial fibrillation: application of a prescribing indicator and multidisciplinary feedback to improve prescribing. *Age Ageing.* 2002 Sep;31(5):391–6.

60. Elliott R a, Woodward MC, Oborne C a. Improving benzodiazepine prescribing for elderly hospital inpatients using audit and multidisciplinary feedback. *Intern Med J.* 2001 Dec;31(9):529–35.

61. Madridejos-Mora R, Amado-Guirado E, P??rez-Rodr??guez MT. Effectiveness of the Combination of Feedback and Educational Recommendations for Improving Drug Prescription in General Practice. *Med Care.* 2004 Jul;42(7):643–648.

62. Arnold FW, McDonald LC, Smith RS, Newman D, Ramirez JA. Improving antimicrobial use in the hospital setting by providing usage feedback to prescribing physicians. *Infect Control Hosp Epidemiol.* 2006 Apr;27(4):378–82.

63. Fowler S, Webber a, Cooper BS, Phimister a, Price K, Carter Y, et al. Successful use of feedback to improve antibiotic prescribing and reduce Clostridium difficile infection: a controlled interrupted time series. *J Antimicrob Chemother.* 2007 May;59(5):990–5.

64. Camins BC, King MD, Wells JB, Googe HL, Patel M, Kourbatova E V, et al. Impact of an antimicrobial utilization program on antimicrobial use at a large teaching hospital: a randomized controlled trial. *Infect Control Hosp Epidemiol.* 2009 Oct;30(10):931–8.

65. Navarro HJ, Shakeshaft A, Doran CM, Petrie DJ. The cost-effectiveness of tailored, postal feedback on general practitioners' prescribing of pharmacotherapies for alcohol dependence. *Drug Alcohol Depend.* 2012 Aug 1;124(3):207–15.

66. Dormuth CR, Carney G, Taylor S, Bassett K, Maclure M. A randomized trial assessing the impact of a personal printed feedback portrait on statin prescribing in primary care. *J Contin Educ Health Prof.* 2012 Jan;32(3):153–62.

67. Kisuule F, Wright S, Barreto J, Zenilman J. Improving antibiotic utilization among hospitalists: a pilot academic detailing project with a public health approach. *J Hosp Med.* 2008 Jan;3(1):64–70.

68. Hux JE, Melady MP, DeBoer D. Confidential prescriber feedback and education to improve antibiotic use in primary care: a controlled trial. *CMAJ.* 1999 Aug 24;161(4):388–92.

69. Mandryk J a, Mackson JM, Horn FE, Wutzke SE, Badcock C-A, Hyndman RJ, et al. Measuring change in prescription drug utilization in Australia. *Pharmacoepidemiol Drug Saf.* 2006 Jul;15(7):477–84.

70. Beilby J, Wutzke SE, Bowman J, Mackson JM, Weekes LM. Evaluation of a national quality use of medicines service in Australia: an evolving model. *J Eval Clin Pract.* 2006 Apr;12(2):202–17.

71. Kiyatkin DE, Wright S, Zenilman J, Kisuule F. Improving utilization of antimicrobial drugs among physician assistants in the ED. *JAAPA.* 2011 Oct;24(10):44, 47–8, 50–1.

72. Smith DH, Christensen DB, Stergachis A, Holmes G. A randomized controlled trial of a drug use review intervention for sedative hypnotic medications. *Med Care.* 1998 Jul;36(7):1013–21.

73. Schoenbaum SC. Feedback of clinical performance information. *HMO Pract.* 1993 Mar;7(1):5–11.

74. Chapman SR. Prescribing information systems; making sense of primary care data. *J Clin Pharm Ther.* 2001 Aug;26(4):235–9.

75. Lomas J, Anderson GM, Domnick-Pierre K, Vayda E, Enkin MW, Hannah WJ. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N Engl J Med.* 1989 Nov 9;321(19):1306–11.

76. Eguale T, Winslade N, Hanley J a, Buckeridge DL, Tamblyn R. Enhancing pharmacosurveillance with systematic collection of treatment indication in electronic prescribing: a validation study in Canada. *Drug Saf.* 2010 Jul 1;33(7):559–67.

77. Shojania KG, Yokoe D, Platt R, Fiskio J, Ma'luf N, Bates DW. Reducing vancomycin use utilizing a computer guideline: results of a randomized controlled trial. *J Am Med Inform Assoc.* 1993;5(6):554–62.

78. Eguale T, Buckeridge DL, Winslade NE, Benedetti A, Hanley J a, Tamblyn R. Drug, patient, and physician characteristics associated with off-label prescribing in primary care. *Arch Intern Med.* 2012 May 28;172(10):781–8.

79. Ho Y-F, Hsieh L-L, Lu W-C, Hu F-C, Hale KM, Lee S-J, et al. Appropriateness of ambulatory prescriptions in Taiwan: translating claims data into initiatives. *Int J Clin Pharm.* 2012 Feb;34(1):72–80.

80. Walton SM, Galanter WL, Rosencranz H, Meltzer D, Stafford RS, Tiryaki F, et al. A trial of inpatient indication based prescribing during computerized order entry with medications commonly used off-label. *Appl Clin Inform.* 2011 Mar 9;2(1):94–103.

81. Bashford JN, Norwood J, Chapman SR. Why are patients prescribed proton pump inhibitors? Retrospective analysis of link between morbidity and prescribing in the General Practice Research Database. *BMJ.* 1998 Aug 15;317(7156):452–6.

82. Soumerai SB, Avorn J. Principles of educational outreach ('academic detailing') to improve clinical decision making. *JAMA.* 1990 Jan 26;263(4):549–56.

83. Erstad BL, Camamo JM, Miller MJ, Webber AM, Fortune J, Pharmd BL, et al. Impacting cost and appropriateness of stress ulcer prophylaxis at a university medical center. *Crit Care Med.* 1997 Oct;25(10):1678–84.

84. Sanal S, Aronow W. Effect of an educational program on the prevalence of use of antiplatelet drugs, beta blockers, angiotensin-converting enzyme inhibitors, lipid-lowering drugs, and calcium channel blockers prescribed during hospitalization and at hospital discharge in pat. *J Gerontol A Biol Sci Med Sci.* 2003 Nov;58(11):1046–8.

85. Awad a I, Eltayeb IB, Baraka OZ. Changing antibiotics prescribing practices in health centers of Khartoum State, Sudan. *Eur J Clin Pharmacol.* 2006 Feb;62(2):135–42.

86. Khalili H, Dashti-Khavidaki S, Hossein Talasaz AH, Tabeefar H, Hendoiee N. Descriptive analysis of a clinical pharmacy intervention to improve the appropriate use of stress ulcer prophylaxis in a hospital infectious disease ward. *J Manag Care Pharm.* 2010 Mar;16(2):114–21.

87. Hatch JB, Schulz L, Fish JT. Stress ulcer prophylaxis: reducing non-indicated prescribing after hospital discharge. *Ann Pharmacother.* 2010 Oct;44(10):1565–71.

88. Gupta R, Marshall J, Munoz JC, Kottoor R, Jamal MM, Vega KJ. Decreased acid suppression therapy overuse after education and medication reconciliation. *Int J Clin Pract.* 2013 Jan;67(1):60–65.

89. Clyne B, Bradley MC, Hughes C, Fahey T, Lapane KL. Electronic prescribing and other forms of technology to reduce inappropriate medication use and polypharmacy in older people: a review of current evidence. *Clin Geriatr Med.* 2012 May;28(2):301–22.

90. Goodman K, Grad R, Pluye P, Nowacki A, Hickner J. Impact of knowledge resources linked to an electronic health record on frequency of unnecessary tests and treatments. *J Contin Educ Health Prof.* 2012 Jan;32(2):108–15.

91. Kuperman GJ, Gibson RF. Computer physician order entry: benefits, costs, and issues. *Ann Intern Med.* 2003 Jul 1;139(1):31–9.

92. Monane M, Matthias DM, Nagle B a, Kelly M a. Improving prescribing patterns for the elderly through an online drug utilization review intervention: a system linking the physician, pharmacist, and computer. *JAMA.* 1998 Oct 14;280(14):1249–52.

93. Raebel M a, Charles J, Dugan J, Carroll NM, Korner EJ, Brand DW, et al. Randomized trial to improve prescribing safety in ambulatory elderly patients. *J Am Geriatr Soc.* 2007 Jul;55(7):977–85.

94. Richards MJ, Robertson MB, Dartnell JG a, Duarte MM, Jones NR, Kerr D a, et al. Impact of a web-based antimicrobial approval system on broad-spectrum cephalosporin use at a teaching hospital. *Med J Aust.* 2003 Apr 21;178(8):386–90.

95. Samore MH, Bateman K, Alder SC, Hannah E, Donnelly S, Stoddard GJ, et al. Clinical decision support and appropriateness of antimicrobial prescribing: a randomized trial. *JAMA.* 2005 Nov 9;294(18):2305–14.

96. Simon SR, Smith DH, Feldstein AC, Perrin N, Yang X, Zhou Y, et al. Computerized prescribing alerts and group academic detailing to reduce the use of potentially inappropriate medications in older people. *J Am Geriatr Soc.* 2006 Jun;54(6):963–8.

97. Smith DH, Perrin N, Feldstein A, Yang X, Kuang D, Simon SR, et al. The impact of prescribing safety alerts for elderly persons in an electronic medical record: an interrupted time series evaluation. *Arch Intern Med.* 2006 May 22;166(10):1098–104.

98. Tamblyn R, Huang A, Perreault R, Jacques A, Roy D, Hanley J, et al. The medical office of the 21st century (MOXXI): effectiveness of computerized decision-making support in reducing inappropriate prescribing in primary care. *CMAJ.* 2003 Sep 16;169(6):549–56.

99. Ahn EK, Cho S-Y, Shin D, Jang C, Park RW. Differences of Reasons for Alert Overrides on Contraindicated Co-prescriptions by Admitting Department. *Healthc Inform Res.* 2014;20(4):280.

100. Sijs H Van Der, Aarts J, Vulto a, Berg M. Overriding of drug safety alerts in computerized physician order entry. *J Am Med …* 2006;138–148.

101. Slight SP, Seger DL, Nanji KC, Cho I, Maniam N, Dykes PC, et al. Are we heeding the warning signs? Examining providers' overrides of computerized drug-drug interaction alerts in primary care. *PLoS One.* 2013;8(12):1–7.

102. Payne TH, Hines LE, Chan RC, Hartman S, Kapusnik-uner J, Russ AL, et al. Recommendations to Improve the Usability of Drug-Drug Interaction Clinical Decision Support Alerts. 2015;(520):1–10.

103. Ash JS, Sittig DF, Campbell EM, Guappone KP, Dykstra RH. Some unintended consequences of clinical decision support systems. *AMIA Annu Symp Proc.* 2007 Jan;26–30.

104. Wright A, Maloney FL, Feblowitz JC. Clinician attitudes toward and use of electronic problem lists: a thematic analysis. *BMC Med Inform Decis Mak.* 2011 Jan;11:36.

105. Wright A, Feblowitz J, Maloney FL, Henkin S, Bates DW. Use of an electronic problem list by primary care providers and specialists. *J Gen Intern Med.* 2012 Aug;27(8):968–73.

106. American Geriatrics Society 2012 Beers Criteria Update Expert Panel. American Geriatrics Society updated Beers Criteria for potentially inappropriate medication use in older adults. *J Am Geriatr Soc.* 2012 Apr;60(4):616–31.

107. Batty GM, Oborne C a, Swift CG, Jackson SH. Development of an indicator to identify inappropriate use of benzodiazepines in elderly medical in-patients. *Int J Geriatr Psychiatry.* 2000 Oct;15(10):892–6.

108. Batuwitage BT, Kingham JGC, Morgan NE, Bartlett RL. Inappropriate prescribing of proton pump inhibitors in primary care. *Postgrad Med J.* 2007 Jan;83(975):66–8.

109. Bez C, Perrottet N, Zingg T, Leung Ki E-L, Demartines N, Pannatier A. Stress ulcer prophylaxis in non-critically ill patients: a prospective evaluation of current practice in a general surgery department. *J Eval Clin Pract.* 2012 Mar 15;

110. Choudhry MN, Soran H, Ziglam HM. Overuse and inappropriate prescribing of proton pump inhibitors in patients with Clostridium difficile-associated disease. *QJM.* 2008 Jun;101(6):445–8.

111. Larson EL, Mayhall CG, Nichols RL, Russell BJ, Adams AB, Fleming DW. Recommendations for preventing the spread of vancomycin resistance: recommendations of the Hospital Infection Control Practices Advisory Committee (HICPAC). *Am J Infect Control.* 1995 Apr;23(2):87–94.

112. Oborne CA, Hooper R, Li KC, Swift CG, Jackson SHD. An indicator of appropriate neuroleptic prescribing in nursing homes. *Age Ageing.* 2002 Nov;31(6):435–9.

113. Pillans PI, Kubler PA, Radford JM, Overland V. Concordance between use of proton pump inhibitors and prescribing guidelines. *Med J Aust.* 2000 Jan 3;172(1):16–8.

114. Williams AW, Dwyer AC, Eddy A a, Fink JC, Jaber BL, Linas SL, et al. Critical and honest conversations: the evidence behind the "Choosing Wisely" campaign recommendations by the American Society of Nephrology. *Clin J Am Soc Nephrol.* 2012 Oct;7(10):1664–72.

115. Barry PJ, Gallagher P, Ryan C, O'mahony D. START (screening tool to alert doctors to the right treatment)--an evidence-based screening tool to detect prescribing omissions in elderly patients. *Age Ageing.* 2007 Nov;36(6):632–8.

116. Gallagher P, O'Mahony D. STOPP (Screening Tool of Older Persons' potentially inappropriate Prescriptions): application to acutely ill elderly patients and comparison with Beers' criteria. *Age Ageing.* 2008 Nov;37(6):673–9.

117. Cameron IM, Lawton K, Reid IC. Appropriateness of antidepressant prescribing: an observational study in a Scottish primary-care setting. *Br J Gen Pract.* 2009 Sep;59(566):644–9.

118. Werner NL, Hecker MT, Sethi AK, Donskey CJ. Unnecessary use of fluoroquinolone antibiotics in hospitalized patients. *BMC Infect Dis.* 2011 Jan;11(1):187.

119. Nardino RJ, Vender RJ, Herbert PN. Overuse of acid-suppressive therapy in hospitalized patients. *Am J Gastroenterol.* 2000 Nov;95(11):3118–22.

120. Lipitz-Snyderman A, Bach PB. Overuse of health care services: when less is more … more or less. *JAMA Intern Med.* 2013 Jul 22;173(14):1277–8.

121. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q.* 2004 Jan;82(4):661–87.

122. Van Walraven C, Naylor CD. Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits. *JAMA.* 1998 Aug 12;280(6):550–8.

123. Liu X, Mills A. Evaluating payment mechanisms: how can we measure unnecessary care? *Health Policy Plan.* 1999 Dec;14(4):409–13.

124. Patel MR. Appropriate Use Criteria to Reduce Underuse and Overuse: Striking the Right Balance. *J Am Coll Cardiol.* 2012 Sep 28;60(19):1885–1887.

125. Weingart SN, Simchowitz B, Padolsky H, Isaac T, Seger AC, Massagli M, et al. An empirical model to estimate the potential impact of medication safety alerts on patient

safety, health care utilization, and cost in ambulatory care. *Arch Intern Med.* 2009 Sep 14;169(16):1465–73.

126. Fitzgerald LS, Hanlon JT, Shelton PS, Landsman PB, Schmader KE, Pulliam CC, et al. Reliability of a modified medication appropriateness index in ambulatory older persons. *Ann Pharmacother.* 1997 May;31(5):543–8.

127. Radley DC, Finkelstein SN, Stafford RS. Off-label prescribing among office-based physicians. *Arch Intern Med.* 2006 May 8;166(9):1021–6.

128. McIntyre J, Conroy S, Avery a, Corns H, Choonara I. Unlicensed and off label prescribing of drugs in general practice. *Arch Dis Child.* 2000 Dec;83(6):498–501.

129. Turner S, Longworth a, Nunn a J, Choonara I. Unlicensed and off label drug use in paediatric wards: prospective study. *BMJ.* 1998 Jan 31;316(7128):343–5.

130. Pandolfini C, Bonati M. A literature review on off-label drug use in children. *Eur J Pediatr.* 2005 Sep;164(9):552–8.

131. Cebul RD, Love TE, Jain AK, Hebert CJ. Electronic health records and quality of diabetes care. *N Engl J Med.* 2011 Sep 1;365(9):825–33.

132. Roth MT, Weinberger M, Campbell WH. Measuring the quality of medication use in older adults. *J Am Geriatr Soc.* 2009 Jun;57(6):1096–102.

133. Falck S, Adimadhyam S, Meltzer DO, Walton SM, Galanter WL. A trial of indication based prescribing of antihypertensive medications during computerized order entry to improve problem list documentation. *Int J Med Inform.* 2013 Aug 8;82(10):996–1003.

134. MedicineNet [Internet]. Available from: http://www.medicinenet.com

135. DrugBank [Internet]. Available from: http://www.drugbank.ca

136. DailyMed: Current Medical Information [Internet]. Available from: http://dailymed.nlm.nih.gov

137. RxNorm API [Internet]. Available from: http://rxnav.nlm.nih.gov/RxNormAPIs.html

138. Carter JS, Brown SH, Erlbaum MS, Gregg W, Elkin PL, Speroff T, et al. Initializing the VA medication reference terminology using UMLS metathesaurus co-occurrences. *Proc AMIA Symp.* 2002 Jan;116–20.

139. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol.* 2010 Jan;6(343):343.

140. McCoy AB, Wright A, Laxmisan A, Ottosen MJ, McCoy J a, Butten D, et al. Development and evaluation of a crowdsourcing methodology for knowledge base construction: identifying relationships between clinical problems and medications. *J Am Med Inform Assoc.* 2012;19(5):713–8.

141. Fung KW, Jao CS, Demner-Fushman D. Extracting drug indication information from structured product labels using natural language processing. *J Am Med Inform Assoc.* 2013 May 1;20(3):482–8.

142. Wei W-Q, Cronin RM, Xu H, Lasko T a, Bastarache L, Denny JC. Development and evaluation of an ensemble resource linking medications to their indications. *J Am Med Inform Assoc.* 2013 Apr 10;1–8.

143. Jung K, LePendu P, Chen WS, Iyer S V, Readhead B, Dudley JT, et al. Automated detection of off-label drug use. *PLoS One.* 2014 Jan;9(2):e89324.

144. Li Y, Salmasian H, Harpaz R, Chase H, Friedman C. Determining the reasons for medication prescriptions in the EHR using knowledge and natural language processing. *AMIA Annu Symp Proc.* 2011 Jan;2011:768–76.

145. Khare R, Li J, Lu Z. LabeledIn: Cataloging labeled indications for human drugs. *J Biomed Inform.* 2014 Aug 23;

146. Chen ES, Hripcsak G, Xu H, Markatou M, Friedman C. Automated Acquisition of Disease-Drug Knowledge from Biomedical and Clinical Documents: An Initial Study. *J Am Med Informatics Assoc.* 2008;15(1):87–98.

147. Pathak J, Chute CG. Analyzing categorical information in two publicly available drug terminologies: RxNorm and NDF-RT. *J Am Med Inform Assoc.* 2010;17(4):432–9.

148. Wright A, Chen ES, Maloney FL. An automated technique for identifying associations between medications, laboratory results and problems. *J Biomed Inform.* 2010;43(6):891–901.

149. Pathak J, Chute CG. Further revamping VA's NDF-RT drug terminology for clinical research. *J Am Med Inform Assoc.* 2011 May 1;18(3):347–8.

150. Barrière C, Gagnon M. Drugs and disorders: From specialized resources to web data [Internet]. In: The 10th international semantic web conference. 2011 [cited 2014 Oct 3]. Available from: https://files.ifi.uzh.ch/ddis/iswc_archive/iswc/pps/web/iswc2011.semanticweb.org/fileadmin/iswc/Papers/Workshops/WeKEx/paper_7.pdf

151. Pathak J, Murphy SP, Willaert BN, Kremers HM, Yawn BP, Rocca W a, et al. Using RxNorm and NDF-RT to classify medication data extracted from electronic health

records: experiences from the Rochester Epidemiology Project. *AMIA Annu Symp Proc.* 2011 Jan;2011:1089–98.

152. SIDER 2 Side Effect Resource [Internet]. Available from: http://sideeffects.embl.de/

153. National Library of Medicine. Unified Medical Language System [Internet]. Available from: http://www.nlm.nih.gov/research/umls/>

154. National Library of Medicine. RxNorm [Internet]. Available from: http://www.nlm.nih.gov/research/umls/rxnorm/

155. National Library of Medicine. SNOMED Clinical Terms [Internet]. Available from: http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html

156. Nelson SJ, Zeng K, Kilbourne J, Powell T, Moore R. Normalized names for clinical drugs: RxNorm at 6 years. *J Am Med Inform Assoc.* 18(4):441–8.

157. Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P. Drug target identification using side-effect similarity. *Science.* 2008 Jul 11;321(5886):263–6.

158. MedlinePlus [Internet]. Available from: http://www.nlm.nih.gov/medlineplus/

159. English Wikipedia [Internet]. Available from: http://en.wikipedia.org

160. Wei W-Q, Mosley JD, Bastarache L, Denny JC. Validation and enhancement of a computable medication indication resource (MEDI) using a large practice-based dataset. *AMIA Annu Symp Proc.* 2013 Jan;2013:1448–56.

161. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp.* 2001 Jan;17–21.

162. Zhu Q, Jiang G, Chute CG. Profiling structured product labeling with NDF-RT and RxNorm. *J Biomed Semantics.* 2012 Jan;3(1):16.

163. Harpaz R, Callahan A, Tamang S, Low Y, Odgers D, Finlayson S, et al. Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug Saf.* 2014 Oct;37(10):777–90.

164. White RW, Harpaz R, Shah NH, DuMouchel W, Horvitz E. Toward enhanced pharmacovigilance using patient-generated data on the internet. *Clin Pharmacol Ther.* 2014 Aug;96(2):239–46.

165. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. *Drug Saf.* 2014 May;37(5):343–50.

166. Mukherjee S, Weikum G, Danescu-Niculescu-Mizil C. People on drugs [Internet]. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14. New York, New York, USA: ACM Press; 2014 [cited 2014 Sep 11]. p. 65–74.Available from: http://dl.acm.org/citation.cfm?doid=2623330.2623714

167. Yang CC, Yang H, Jiang L, Zhang M. Social media mining for drug safety signal detection [Internet]. In: Proceedings of the 2012 international workshop on Smart health and wellbeing - SHB '12. New York, New York, USA: ACM Press; 2012 [cited 2014 Sep 11]. p. 33.Available from: http://dl.acm.org/citation.cfm?doid=2389707.2389714

168. Hughes S, Cohen D. Can online consumers contribute to drug knowledge? A mixed-methods comparison of consumer-generated and professionally controlled psychotropic medication information on the internet. *J Med Internet Res.* 2011 Jan;13(3):e53.

169. Pivovarov R, Elhadad N. A hybrid knowledge-based and data-driven approach to identifying semantically similar concepts. *J Biomed Inform.* 2012;45(3):471–481.

170. Hanlon JT, Schmader KE, Ruby CM, Weinberger M. Suboptimal prescribing in older inpatients and outpatients. *J Am Geriatr Soc.* 2001 Feb;49(2):200–9.

171. Rancourt C, Moisan J, Baillargeon L, Verreault R, Laurin D, Grégoire J-P. Potentially inappropriate prescriptions for older patients in long-term care. *BMC Geriatr.* 2004 Oct 15;4:9.

172. Rollason V, Vogt N. Reduction of polypharmacy in the elderly: a systematic review of the role of the pharmacist. *Drugs Aging.* 2003 Jan;20(11):817–32.

173. National Guideline Clearinghouse [Internet]. Available from: http://www.guideline.gov/

174. Woolf SH. Practice guidelines: a new reality in medicine. I. Recent developments. *Arch Intern Med.* 1990 Sep;150(9):1811–8.

175. Kahol K, Vankipuram M, Patel VL, Smith ML. Deviations from protocol in a complex trauma environment: errors or innovations? *J Biomed Inform.* 2011 Jun;44(3):425–31.

176. Shiffman RN, Karras BT, Agrawal a, Chen R, Marenco L, Nath S. GEM: a proposal for a more comprehensive guideline document model using XML. *J Am Med Inform Assoc.* 7(5):488–98.

177. Wang D, Peleg M, Tu SW, Boxwala A a, Ogunyemi O, Zeng Q, et al. Design and implementation of the GLIF3 guideline execution engine. *J Biomed Inform.* 2004 Oct;37(5):305–18.

178. Tu SW, Musen M a. A flexible approach to guideline modeling. *Proc AMIA Symp.* 1999 Jan;420–4.

179. Sutton DR, Fox J. The syntax and semantics of the PROforma guideline modeling language. *J Am Med Inform Assoc.* 2003;10(5):433–43.

180. Tu S, Campbell J, Glasgow J. The SAGE Guideline Model: achievements and overview. *J Am Med Inform Assoc.* 2007;14(5):589–598.

181. Chen R, Georgii-Hemming P, Ahlfeldt H. Representing a chemotherapy guideline using openEHR and rules. *Stud Health Technol Inform.* 2009 Jan;150:653–7.

182. Johnson PD, Tu SW, Musen MA, Purves I. A virtual medical record for guideline-based decision support. *Proc AMIA Symp.* 2001 Jan;294–8.

183. OpenCDS: Open Clinical Decisoin Support Tools and Resources. [Internet]. Available from: http://www.opencds.org/

184. Salmasian H, Tran TH, Friedman C. Developing a formal representation for medicatoin appropriateness criteria. *AMIA Annu Symp Proc.* 2014;

185. Good BM, Nanis M, Su AI. Microtask crowdsourcing for disease mention annotation in PubMed abstracts. *arXiv Prepr arXiv14081928.* 2014;282–293.

186. Zhou L, Hongsermeier T, Boxwala A, Lewis J, Kawamoto K, Maviglia S, et al. Structured representation for core elements of common clinical decision support interventions to facilitate knowledge sharing. *Stud Health Technol Inform.* 2013;192(1-2):195–199.

187. Pérez-Pérez M, Glez-Peña D, Fdez-Riverola F, Lourenço A. Marky: A tool supporting annotation consistency in multi-user and iterative document annotation projects. *Comput Methods Programs Biomed.* 2015;118(2):242–251.

188. Nadkarni GN, Gottesman O, Linneman JG, Berg RL, Farouk S, Nadukuru R, et al. Development and validation of an electronic phenotyping algorithm for chronic kidney disease. *AMIA Annu Symp Proc.* 2014;907–916.

189. Hanlon JT, Weinberger M, Samsa GP, Schmader KE, Uttech KM, Lewis IK, et al. A randomized, controlled trial of a clinical pharmacist intervention to improve inappropriate prescribing in elderly outpatients with polypharmacy. *Am J Med.* 1996 Apr;100(4):428–37.

190. Arnold SR, Straus SE. Interventions to improve antibiotic prescribing practices in ambulatory care. *Cochrane Database Syst Rev.* 2005 Jan;(4):CD003539.

191. Martin P, Tamblyn R, Ahmed S, Tannenbaum C. An educational intervention to reduce the use of potentially inappropriate medications among older adults (EMPOWER study): protocol for a cluster randomized trial. *Trials.* 2013 Jan;14(1):80.

192. Lechevallier-Michel N, Gautier-Bertrand M, Alpérovitch A, Berr C, Belmin J, Legrain S, et al. Frequency and risk factors of potentially inappropriate medication use in a community-dwelling elderly population: results from the 3C Study. *Eur J Clin Pharmacol.* 2005 Jan;60(11):813–9.

193. Liu C-L, Peng L-N, Chen Y-T, Lin M-H, Liu L-K, Chen L-K. Potentially inappropriate prescribing (IP) for elderly medical inpatients in Taiwan: a hospital-based study. *Arch Gerontol Geriatr.* 2012;55(1):148–51.

194. Nobili A, Marengoni A, Tettamanti M, Salerno F, Pasina L, Franchi C, et al. Association between clusters of diseases and polypharmacy in hospitalized elderly patients: results from the REPOSI study. *Eur J Intern Med.* 2011 Dec;22(6):597–602.

195. Wright A, Pang J, Feblowitz JC, Maloney FL, Wilcox AR, McLoughlin KS, et al. Improving completeness of electronic problem lists through clinical decision support: a randomized, controlled trial. *J Am Med Inform Assoc.* 2012;19(4):555–61.

196. Bailey TC, Noirot L a, Blickensderfer A, Rachmiel E, Schaiff R, Kessels A, et al. An intervention to improve secondary prevention of coronary heart disease. *Arch Intern Med.* 2007 Mar 26;167(6):586–90.

197. Freedberg DE, Salmasian H, Abrams JA, Green R a. Orders for Intravenous Proton Pump Inhibitors After Implementation of an Electronic Alert. *JAMA Intern Med.* 2015;175(3):452–4.

198. Boxwala A a, Peleg M, Tu S, Ogunyemi O, Zeng QT, Wang D, et al. GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines. *J Biomed Inform.* 2004 Jun;37(3):147–61.

199. Peleg M, Boxwala a a, Bernstam E, Tu S, Greenes R a, Shortliffe EH. Sharable representation of clinical guidelines in GLIF: relationship to the Arden Syntax. *J Biomed Inform.* 2001 Jun;34(3):170–81.

200. Choi J, Currie LM, Wang D, Bakken S. Encoding a clinical practice guideline using guideline interchange format: a case study of a depression screening and management guideline. *Int J Med Inform.* 2007 Oct;76 Suppl 2:S302–7.

201. Drools Business Rules Management System [Internet]. Available from: http://www.drools.org

202. Peleg M. Comparing Computer-interpretable Guideline Models: A Case-study Approach. *J Am Med Informatics Assoc.* 2002 Sep 23;10(1):52–68.

203. Seyfang A, Kosara R, Miksch S. Asbru's Referance Manual, Version 7.3. Vienna: Vienna University of Technology; 2002.

204. Quaglini S, Stefanelli M, Lanzola G, Caporusso V, Panzarasa S. Flexible guideline-based patient careflow systems. *Artif Intell Med.* 2001;22(1):65–80.

205. Johnson PD, Tu S, Booth N, Sugden B, Purves IN. Using scenarios in chronic disease management guidelines for primary care. *Proc AMIA Symp.* 2000;389–393.

206. Freedberg DE, Salmasian H, Friedman C, Abrams J a. Proton pump inhibitors and risk for recurrent Clostridium difficile infection among inpatients. *Am J Gastroenterol.* 2013;108(11):1794–801.