“RXNORM CODING OF EHR MEDICATION ORDERS USING AN NLP-BASED APPROACH”

A THESIS SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA

SAJEDA K. ABDO, RPH

IN PARTIAL FULFILLMENT OF THE REQUIERMENTS

FOR THE DEGREE OF

MASTER OFSCIENCE IN HEALTH INFORMATICS

ADVISER: DR. BONNIE WESTRA

OCTOBER 2013

Acknowledgements

I want to express my deeply-felt thanks to my advisor, Dr. Bonnie Westra for her encouragement, thoughtful guidance and patience. I also extend my deep gratitude to Dr. Genevieve Melton-Meaux, Dr. Terrance Adam and Dr.Saif Khairat for their substantial assistance and support. Exceptional thanks goes to my Father (May his soul rest in peace), Khamis Abdo, and My mother, Fatima, for having instilled in me the confidence to face difficult challenges and come out as a winner. They provided me with solace in times of despair and encouraged me to carry on. My heart-felt thanks to my amazing and brilliant Husband Ahmad for his ongoing support, quiet patience and unwavering devotion and love were undeniably the bedrock upon which graduate years have been built. I would never forget the fun and joy my lovely daughters Layan and Raneem gave to me during the period of my study. Above all, I thank God for providing me with such a great people around me who have contributed generously to my success.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction and Background

1. 1 Introduction

1.1.1 Importance of EHR Data.

Prior to the implementation and use of EHRs, claims data has traditionally been used as the main source of electronic health information to support clinical research. It is limited however, since claims data are missing many key data elements that are required to better reflect the patient's conditions and care plan [1]. It only captures demographics, diagnoses and procedures recorded for billing purposes [2]. To fill in this knowledge gap, EHRs can provide important details about vital signs, diagnostic test results, social and family history, prescriptions and physical examination findings. Perhaps one main advantage of claims data over EHRs is that it is only through claims data that a holistic view of the patient's interactions with the health care system can be seen given that only 57% of healthcare providers utilized EHRs in 2011 [3, 4]. The current trend is to utilize EHRs to generate and test hypotheses about the relationships among patients, diseases, practice styles, therapeutic modalities and clinical outcomes [5, 1, 2].

1.1.2 Importance of Pharmacy Information.

Medication information is among the most important types of clinical data in EHR. It is critical for healthcare safety and quality, as well as for clinical research that uses EHR Data [6]. Medication orders are often represented by two parts; medication name and medication signature (information about drug administration, such as dose, route, frequency, and duration). [7]. Sharing medication names and signatures among providers is necessary to create an accurate medication profile for a patient [6]. It is also an essential step for the support of accurate medication reconciliation process; which compares a patient's medication orders to all of the medications that the patient has been taking to avoid medication errors such as omissions, duplications, dosing errors, or drug interactions [8]. This has the potential to reduce adverse drug reactions (ADR); which is one of the leading causes of morbidity and mortality in health care [9]. The Institute of Medicine reported in January of 2000 that from 44,000 to 98,000 deaths occur annually from medical errors. Of this total, an estimated 7,000 deaths occur due to ADRs [10].

To facilitate sharing of EHR medication data, medication terminology and coding standards have been utilized. The National Library of Medicine (NLM) created RxNorm, a standardized nomenclature for clinical drugs that is used for the electronic exchange of medication name and signature, as well as other drug-related information [11]. The goal of RxNorm is to enable pharmacy management systems and drug interaction software,

using different drug nomenclatures, to share and exchange data efficiently by mapping their drug vocabularies to standard RxNorm clinical drug names. The scope of RxNorm contains the names of prescription and many over-the-counter drugs available in the United States. Radiopharmaceuticals, bulk powders, contrast media, food, dietary supplements, and medical devices, such as bandages and crutches, are all out of scope for RxNorm [12].

To support accurate and efficient RxNorm encoding of EHR medications, medication names and signatures can be used [13,14]. The main drawback of using a medication name and signature alone for RxNorm coding is that the medication could be mapped to different RxNorm concept unique identifier (RxCUI) which would need further manual review by a domain expert to assign the right RxCUI to the medication order [13]. Therefore, EHR medication orders usually populate medication's National Drug Code (NDC) at the time the order is being created and recorded by the system. A single NDC code is associated with only one single RxCUI but a single RxCUI is associated with multiple NDCs [15]. Drug NDC serves as a universal product identifier for drugs. Registered drug establishments regularly provides The Food and Drug Administration (FDA) with a current list of all drugs manufactured, prepared, propagated, compounded, or processed by it for commercial distribution. The NDC is a commercially oriented coding system identifying the company, the drug trade name, strength, and dosage form and package size. [16].

Studies have found that the NDC Directory is unreliable . A study in 2011 found that 27% of the 123,856 codes in the NDC Directory were erroneous, and that 14,337 additional prescription drug products were missing codes [17,18]. With missing or inaccurate NDCs, drug names and signatures would be employed for coding medication records.

1.1.3 EHR Medication Data issues and Impact to Mapping and Usability.

Medications orders in EHR databases are usually created using computer systems that populate medication information as well as NDC codes. However, there are cases where medication orders are missing key information such as an NDC code or any of the medication signature information at the time of data entry. For example, providers often use free text fields to store clinical drug data in electronic health records. The use of free text facilitates rapid data entry by the clinician. Errors in spelling, abbreviations, and jargon, however, limit the utility of mapping medications to RxNorm identifiers [19]. Furthermore, patients' self-reporting medications may have missing NDCs and incomplete data [20]. EHR data in general suffers from at least one of the three common data quality issues in Table (1) [21]. This might impact the quality and efficiency of automated RxNorm encoding tools [22, 23].

Table 1 -EHR data quality issues

Data Quality Issue	Definition
Incompleteness	Missing information
Inconsistency	Information mismatch between various or within the same EHR data source;
Inaccuracy	Non-specific, non-standards-based, inexact, incorrect, or imprecise information.

EHR databases contain millions of medication orders that are coded to RxNorm using tools such as Medi-Span®, © First Databank, Drug Indications Database™ and RxNav (a browser for RxNorm by NLM) [24]. The accuracy of the automated mapping relies heavily on the completeness and uniqueness of the medication information (name, signature, and NDC). Due to missing information, the mapping task becomes suboptimal and cannot be fulfilled by only using an exact string match or simple mapping algorithm since this would have the potential to misrepresent the RxNorm dataset through the mapping to many target RxNorm concepts [25]. Manual mapping would also be costly and time consuming. Little research has been done to study the mapping between RxNorm and EHR medication orders developed. Such studies are critical for the future adoption and integration of RxNorm in EHRs [23].

In this thesis, the impact of incomplete EHR medication orders on the automation of RxNorm coding/mapping was analyzed.

1.2 Background

Diverse medication vocabularies are used to improve the semantic interoperability between electronic health records (EHRs). For example, a Computerized Physician Order Entry (CPOE) or a Pharmacy Information System (IS) might use a local terminology or provided by a vendor or both [26]. When a medication order s submitted by the provider, it gets to the electronic Medication Administration Record (eMAR), medication names will probably be modified to match the package description at the Food and Drug Administration (FDA)'s National Drug Code (NDC) level [27]. Consequently, it is likely that the same medication concept can be linked to more than one identifier.

In 2004, the initial release of RxNorm was created as a mean to provide an available and reliable standard medication terminology to mediate messages between systems not using the same software and vocabulary. Eventually, this will improve the semantic interoperability of information in heterogeneous systems. [28,29].

There have been several studies that mapped medication terminology across institutions prior to the development of RxNorm. For example, Kannry et al. [30] conducted a study in 1996 to examine the issues involved in mapping an existing structured controlled vocabulary, the Medical Entities Dictionary (MED) developed at Columbia University, to an institutional vocabulary, the laboratory and pharmacy vocabularies of the Yale New Haven Medical Center. With a match rate of 73%, the researchers highlighted the critical need of standardization of local pharmacy vocabulary subsets, standardization of attribute representation, and term granularity.

In another study, Sherertz et al. [31] utilized a lexical mapping that matches terms on an "exact word by word equivalence of phrase." They demonstrated the approach by lexically mapping 834 descriptions from the University of Southern California at San Francisco (UCSF) to Medical Subject Headings (MeSH) [32] terms. They successfully mapped 47.8% of the descriptions to MeSH terms.

Another mapping approach was proposed by Evans et.al. [33] and Cimino et.al [34]. Their utilized a frame-based approached for mapping descriptions and terms. A frame is a self- contained "unit of knowledge representation" that contains a term and its attributes. The mapping approach identifies semantic and hierarchical relationships between terms. In both studies a match rate of 40% to 50% was achieved.

Mapping studies that utilized algorithms and tools prior to the existence of RxNorm are in general small in size; i.e. the number of mapped terms is usually below 1000 terms. This is due to the fact the mapping algorithms relied heavily on manual interactions that make it a very time consuming process. Furthermore, the match rate is usually <50%. This is referred to incomplete information that impacted the mapping. For example, Kannry et al. [30] reported that many extracted pharmacy terms do not specify values for the attributes dose and dosage in advance.

With the development of RxNorm and given its emerging role as a national standard, its use within the informatics infrastructure of many organizations was felt to offer a scalable strategy for representing drug orders obtained from different EHR systems using different drug vendor information models. A study in 2009 by Hernandez et.al. [35],

merged pharmacy data from two Stanford hospitals that use different drug representation systems into a single, standards-based model supporting research by mapping HL7 pharmacy orders to RxNorm concepts, the output was utilized by the Stanford Translational Research Integrated Database Environment (STRIDE) clinical data warehouse. Compared to the studies conducted prior to the development of RxNorm, we notice that this study mapped around 9000 pharmacy orders and achieved a mapping rate of 93%. The matching rate was impacted by the data quality of the text and as a result manual mapping was required in some cases.

RxNorm was also used to exchange standardized, codified patient drug allergy information between VA and DoD [36]. RxNorm is designated as the national standard for exchanging allergies to branded drugs. After both agency files were mapped to RxNorm, DoD will understand 74 percent of VA terms and VA will understand 58 percent of DoD terms. One major drawback identified in this study for using RxNorm is that RxNorm is not built to be a complete drug reference terminology. As a result, it does not model drug classes and multiple ingredient generics (which make up a portion of the drug allergies file in both agencies) as single terms.

RxNorm has been as well used to capture medication history of the patient. Two methods have been usually used to capture patient's medication history:

Mapping patient's medical data into RxNorm after the data is being collected. Most of the studies in this domain achieved a certain percentage of medications map correctly and usually a manual interaction is needed. In addition, many successful mapping efforts have

been limited to well-defined subsets containing only fully specified prescriptions incorporating National Drug Codes (NDCs) [37]. Even the most sophisticated algorithms cannot overcome all the limitations of poorly collected medication history at the point-of-care, such as open-ended free text fields, non-standard abbreviations, or invalid combinations due to uncontrolled data capture (e.g., recording a medication in a non-existent dose unit or dose form) [38].

Capturing patient's medication data in RxNorm-compatible format at the point of data entry. While this approach avoids some of the issues that the above approach faces, it still has its own challenges. For example, medication coverage can be an issue if a medication is missing in a search list which may prevent data entry from occurring. In addition, data entry systems must allow the capturing of partial data when the patient does not recall specific information or lacks the abilities to report complete information (e.g., an individual experiencing psychotic schizophrenia) [38, 39].

In general, mapping medication information to terminology standards is a challenging process due to many reasons [23, 40, 41]:

- The drug naming conventions as well as drug attributes are represented differently among institutions, or even among various applications (e.g., CPOE, Pharmacy IS).
- The various representations of drug names (e.g., ingredient name, generic name, multiple brand names, and longstanding "nicknames" like HCTZ that means

hydrochlorothiazide and APAP that stands for acetaminophen). This yields a number of mappings for one term.

- The incompleteness and various representations of drug signature elements (e.g., dose form and strength information), yielding a number of mappings for one term (e.g., "Acetaminophen 325 MG / Oxycodone Hydrochloride 5 MG Oral Tablet [Percocet 5/325] ").
- The maintenance of mapping results is another challenge. With the regular updates to terminologies (e.g., addition, refinement, and obsolescence), the mapping results needs to be maintained and updated as per the new releases of terminology standards. Consequently, the initial mapping effort might not be sufficient. Continuous maintenance of the mapping results might be needed.
- The mapping effort becomes costly and time consuming, if only using approaches that rely heavily on manual interactions. An automated method would make the process manageable and allow institutional dictionaries to stay "in synch" with the standards." [42].

Little research has been done to study the mapping between RxNorm and medication terminologies developed at local institutions or other organizations [36]. Further analysis is needed to understand the underlying data issues that impact the mapping rate and to identify recommendations to enhance mapping rates. Such studies are critical for the future adoption and integration of RxNorm in EHRs [42].

Chapter 2 – Research purpose and methods

2.1 Research Purpose

The main goal of this study was to evaluate the impact of incomplete EHR medication orders on the automation of RxNorm coding/mapping. Specifically, the research question is: how to develop a semi-automated mapping process for incomplete EHR medication orders using existing RxNorm API?

To answer the research question, the following specific objectives were addressed:

- Describe EHR issues that affect the ability to automate mapping of medication records to RxNorm.
- Map EHR medication records that have missing National Drug Codes (NDC) to RxNorm.
- Create a list of rules to enhance automated RxNorm coding (help finding the most accurate target RxNorm concept without manual review) of EHR medication based on observations from the manual review step.
- Validate the set of rules by applying the set of rules to RxNav API for all EHR medications and compare results to Medi-Span® coding tool.
- Describe data quality issues that may affect mapping medication orders to RxCUIs.

2.2 Data Source

In this study, about 25 million medication orders with missing NDCs were analyzed. The source data was from the Academic Health Center Information Exchange Platform at the University of Minnesota (AHC-IE). The AHC-IE is a secure environment to hold protected health information provided by the Clinical Translational Science Institute, Biomedical Informatics (CTSI-BMI) and managed by the Academic Health Center (AHC) at the University of Minnesota. AHC-IE contains clinical data from Fairview Health System's (FHS) EHR with data from 8 hospitals and over 40 clinics. The AHC_IE platform [43] is shown in Figure (1).

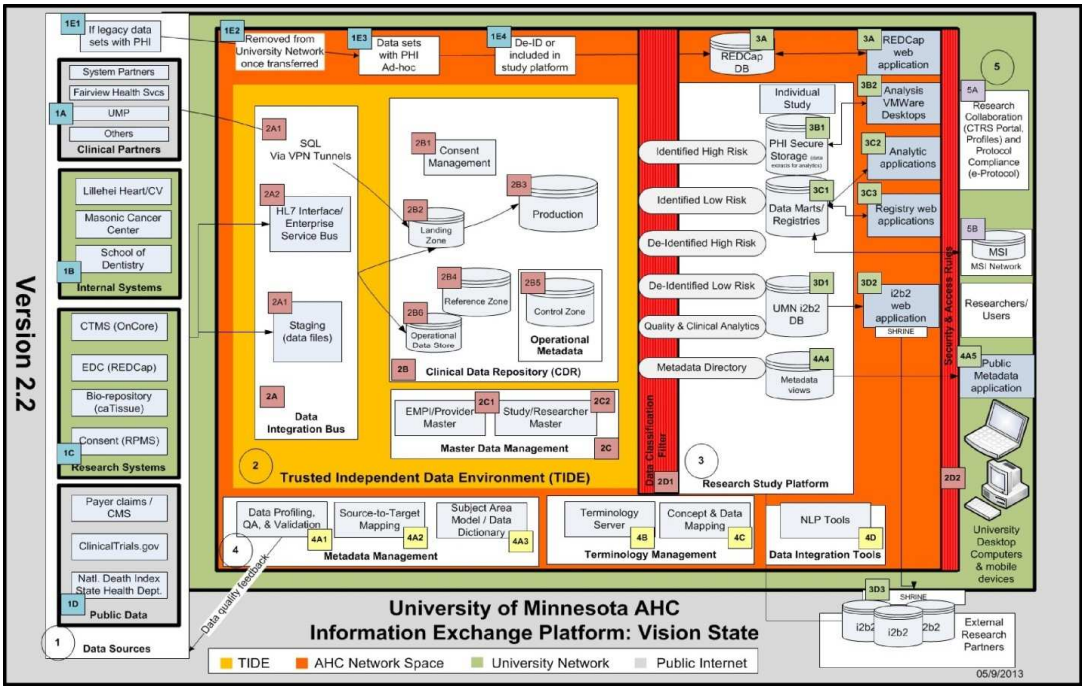

Figure 1 - The Academic Health center Information Exchange Platform at the University of Minnesota

The details associated with de-identified medication orders were extracted and provided in an Excel file for the investigator. The medication orders where then filtered to exclude RxNorm out-of-scope items. This covers radiopharmaceuticals, bulk powders, contrast media, food, dietary supplements, and medical devices, such as bandages and crutches. This resulted in approximately 20 million records.

The most frequent medications without NDC codes covered approximately 95% of the EHR medication orders, which represents around 4000 records, were considered for the purpose of the RxNorm mapping effort. The medications were mapped to RxNorm using RxNav; which is a browser and Application Programming Interfaces (API) for RxNorm provided by NLM to resolve drug information into an RxNorm concept [44].

2.3 Methods

This work was exempt from the Institutional Review Board (IRB) approval since the data utilized were sufficiently de-identified. The analysis was performed using a secure, HIPAA-compliant environment.

The automated mapping results were manually reviewed by a registered pharmacist (RPH) under the supervision of a registered nurse (RN) and a medical doctor (MD) to create a gold standard. To validate and further refine the gold standard, we ran the same set of medications (pertained to 95% of the EHR medication orders) through Medi-Span tool and results were compared to the manual review results. We evaluated the accuracy of the automated mapping by comparing the results of RxNav to the gold standard.

2.3.1 Framework

We followed a framework similar to Zhou, et. al's. study design where they measured the accuracy of an NLP-based approach for RxNorm coding [42]. While the study framework is similar, we customized it to fit the aims of this research (Figure 2).

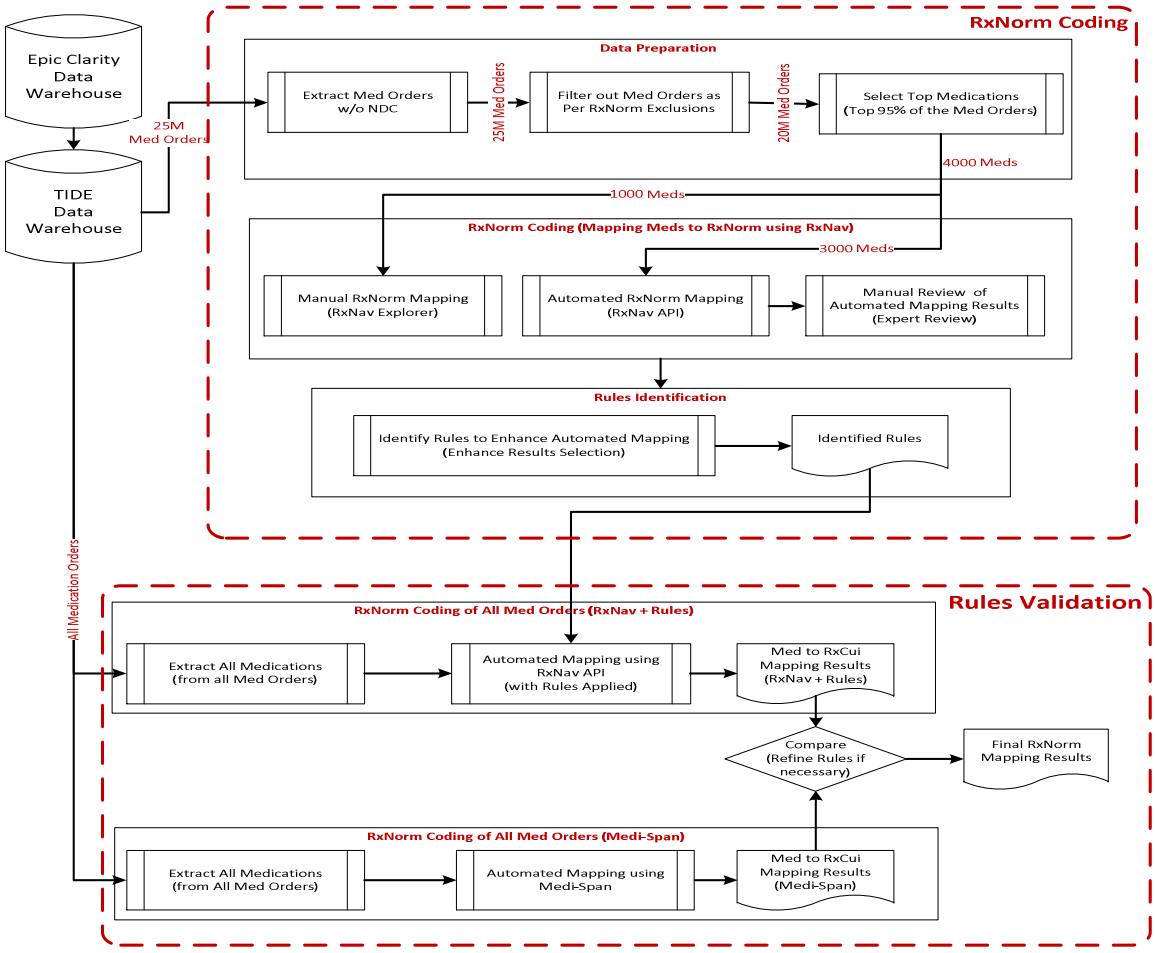

Figure 2 - RxNorm Coding Framework

2.3.2 Data Extraction and Preparation

There were a combination of 19033 unique medication names and signatures without NDC codes representing 24,994,195 medication orders extracted using a SQL query from Clarity tables and were listed in a descending order according to the count of medication orders. The query pulls the data fields described in Table (2) from the landing zone of the AHD-IE data warehouse in which EPIC Clarity data are loaded. The fields were selected to match the format of the normalized drug name in RxNorm standard; which consists of name, active ingredient, strength, dose, form and route.

Table 2 - Extracted EHR medication order dataset

Field Name	Description
Medication ID	The unique ID of the medication record.
Name	The name of the medication. It could be generic, brand or a general name.
Form	The form of the medication, such as tablet or suspension.
Route	The route of administration of the medication, such as intramuscular or subcutaneous.
Strength	The strength of this version of the drug, for example, 10%, or 50 mg/ml.
Med_order_count	Added field to count the frequency of the Med_ID appeared in the table.

An expert (Pharmacist) manually filtered the extracted list of medications to exclude medications that do not map to RxNorm as per RxNorm exclusion criteria. This includes radiopharmaceuticals, bulk powders, contrast media, food, dietary supplements, and medical devices, such as bandages and crutches, which are all out of scope for RxNorm [12].

Consequently, 1,010 (5.3%) unique medication identifiers pertained to 5,437,573 (21.1%) medication orders were excluded from the list of 19,033 unique medication identifiers. The list of medications now contains 18,023 records representing 20,556,815 medication orders.

We further manually filtered the list of medications, for the RxNorm mapping effort, to focus on the most frequent medications identifiers; which we define as medication identifiers that are related to 95%, Figure (3) displays a chart that illustrates the cumulative count f the medications when sorted in a descending order according to their occurrence frequency in Clarity tables.

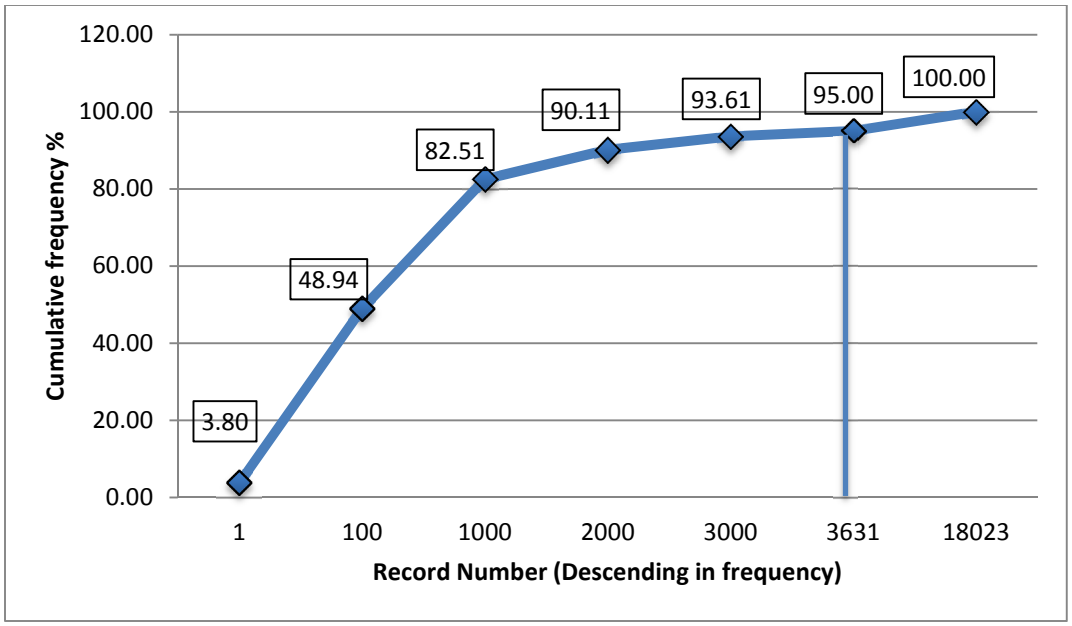

Figure 3 - Cumulative frequency of medication records showing the 95% cut point

The resulting list contained 3713 medication identifiers representing 18,578,791 EHR medication orders. This covers the top 3713 medications from the list of medications that are arranged in a descending order according to the count of the medication orders per medication identifier.

2.3.4 RxNorm Coding of EHR Medications

Initially, the RxNorm coding effort of the identified 3,713 medications started manually (849 medications) and then was automated to speed up the process (2,864 medications).

2.3.4.1 Manual Mapping using RxNav Explorer

Mapping of 849 medication records was done manually using RxNav Explorer; an NLM tool for mapping drugs to RxNorm standard using a Java Web Start Technology [45]. Figure (4) shows a snapshot of RxNav explorer results screen.

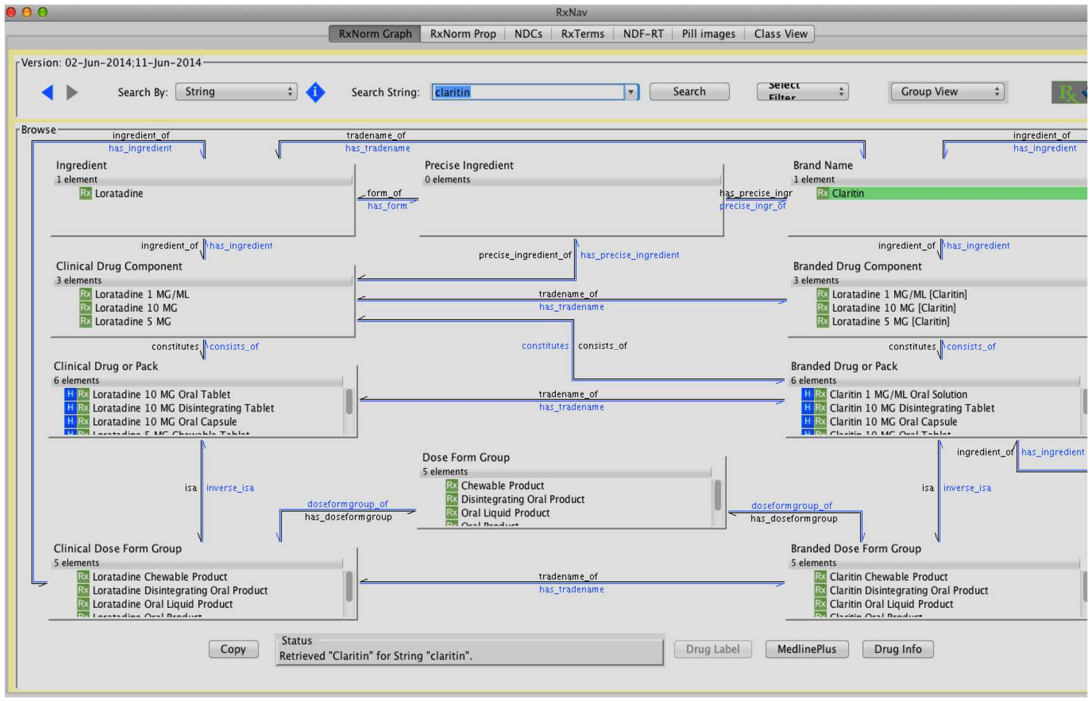

Figure 4 - RxNAV Explorer search results screen

We used the string-based search process of the Explorer. The process requires manual data entry of medication name as well as signature information. The available provided information determines whether a specific RxNorm mapping can be found. If any of the signature information is missing, then expert intervention is needed to select the most appropriate RxNorm code from the Explorer's list of suggested Target RxCUI. In this case, the selected RxCUI would represent a less specific drug mapping as in the example in Figure (5).

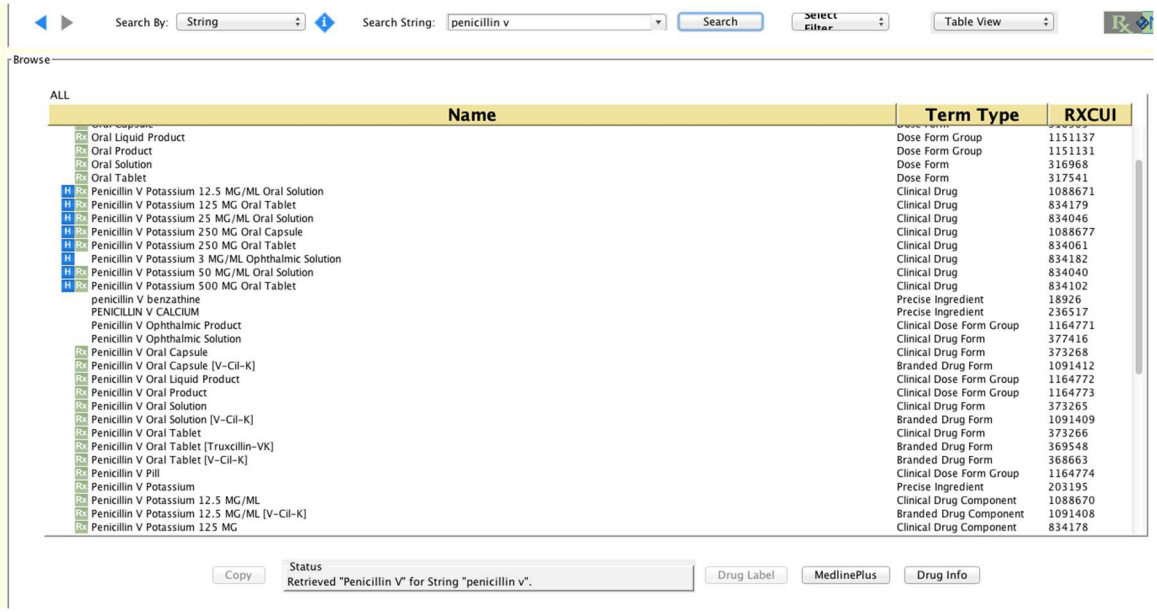

Figure 5 - List of RxCUI mapping options retrieved by RxNAV for a medication with incomplete information

In some cases, the medication information would need to be edited prior to mapping to RxNorm. For example (Figure 6), RxNav would not map "SULFATRIM DS 800-160 MG OR TABS" to RxNorm since it does not recognize the format of the medication.

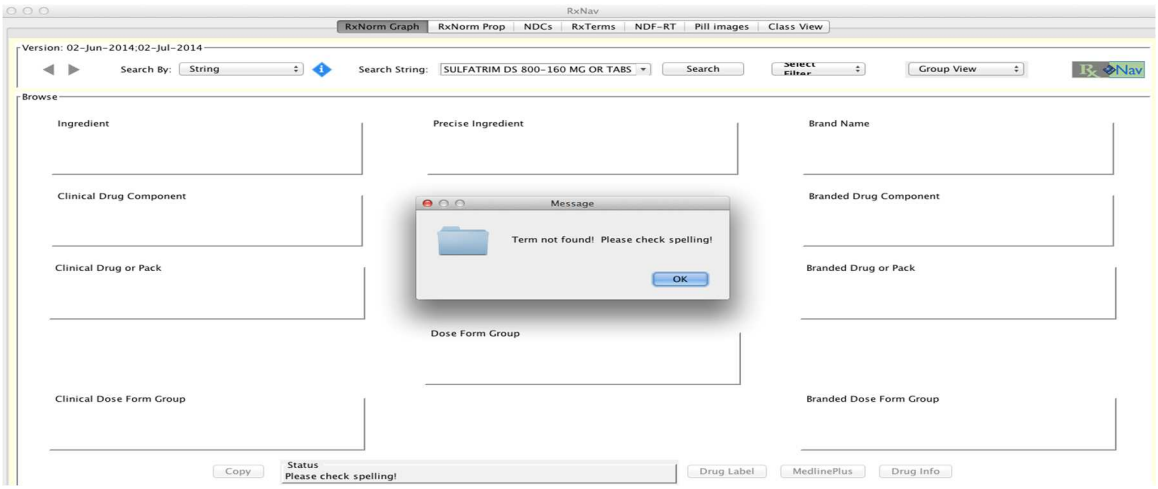

Figure 6 - RxNAV may not recognize original medication strings

The medication Information was adjusted to SULFATRIM only for the medication name.

The remaining medication information was manually entered by an expert (Figure 7).

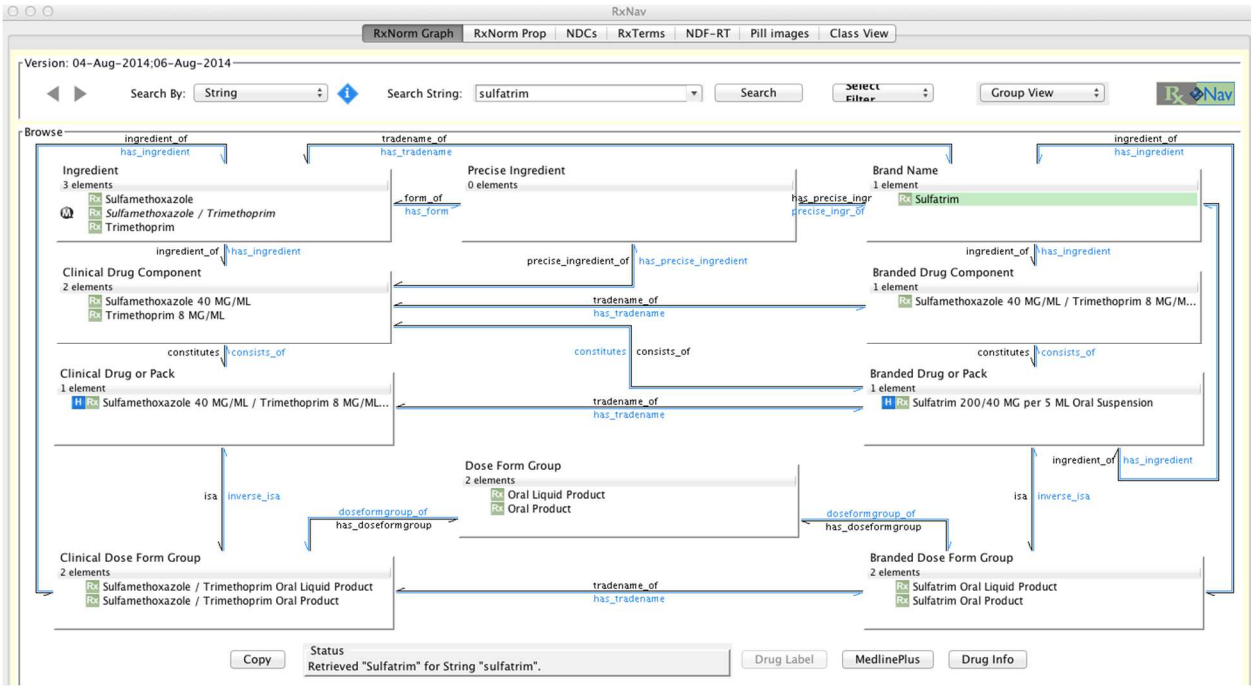

Figure 7 - Editing of original medication information needed to find a mapping result.

2.3.4.2 Automated Mapping using RxNav API

To speed up the process, we moved toward a semi-automated process in which a Java program has been created to utilize the RxNorm API web services for accessing the RxNorm data set [46]. The Java code workflow is illustrated in Figure (8).

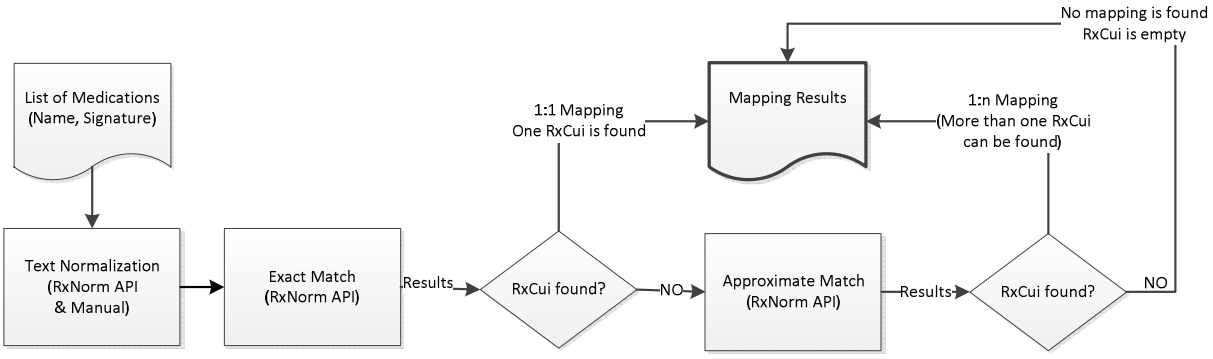

Figure 8 - Automated mapping workflow using RxNorm API

Text Normalization Using RxNorm API:

Using the lexical variant generation Norm (referred to lvg-norm) module of the RxNorm API tool, the normalization process involves stripping genitive marks, transforming plural forms into singular, replacing punctuation (including dashes) with spaces, removing stop words, lower-casing each word, breaking a string into its constituent words, and sorting the words in alphabetic order [47]. In addition, some standard abbreviations and acronyms are expanded into full names [48].

Exact match:

Using findRxCUIByString functionality of the RxNorm API tool, (http://rxnav.nlm.nih.gov/RxNormAPIs.html#uLink=RxNorm_SOAP_findRxCUIByString), the Exact match process finds a matching record in RxNorm database for the exact given post-normalization drug name passed to it.

Approximate match:

If the exact match process fails to find a matching RxCUI for the passed drug information, then the process will route the medication name and other information to the approximate match functionality of the RxNav API; 'getApproximateMatch' which does an approximate match search to determine the strings in the RxNorm data set that most closely match the search string. This function finds the closest matches of the string with drug name concepts in RxNorm. Strings containing other medical terms such as drug classes or diseases will yield little or no results since the algorithm looks for drug names in the string. (http://rxnav.nlm.nih.gov/RxNormApproxMatch.html). An approximate match group structure also contains RxCUI, a score; calculated by Jaccard's coefficient and ranging from 0 to 100 indicating the closeness of the match, and a rank of the string to all possible candidates [47].

2.3.4.3 Manual Review of Automated Mapping Results

The results of the automated mapping need to be reviewed by an expert to identify the most appropriate RxCUI to be used for each drug. Domain experts need to interactively monitor the process and help facilitate text processing steps, especially in critical domains like healthcare [49]. In addition, during this process, we tag each mapping result with a justification of why a specific RxCUI was selected.

2.3.5 Mapping Rules Identification

During the manual review of automated mapping results and the selection of the most appropriate target RxCUI concepts, clinical domain experts identified a set of business rules that can help enhance coding automation.

Some of these rules were based on observations while others were based on clinical knowledge. The following summarize the most important rules agreed upon:

- Map to the most granular RxNorm code available that matches the information provided.
- Map to the brand name if it was mentioned since brand names give more specific information about the medication.
- If no brand name is mentioned then map to the generic name. Don’t add brand names even if there is no generic product available in the market.

- When the strength is not known, map to the form and route (ex. Oral tablet). Don't map to any strength even if there is only one strength savialbe in the market or if it is the most commonly used one unless its impeded within the name of the medication.
- If the form is not known, then map to the form group. (ex. oral product).
- When the form and the strength are not mentioned, map to the active ingredient.

2.3.6 Rules Validation

The goal of this step is to validate the correctness of the identified business rules. The ultimate objective is to validate the accuracy of selected RxCUI and minimize manual intervention by an expert.

To perform the validation process, the business rules identified in the previous step were applied and the automated mapping was executed on the remaining EHR medication orders. The output was then compared to the output of the Medi-Span mapping tool. Medi-Span is a tool used by Fairview (the provider of the EHR medication orders) to map Medication orders to the appropriate RxNorm concept**.**

We used medication GPIs (Generic Product Identifiers) to match Medi-Span mappings with mappings produced by RxNav (automated mapping with business rules applied).

GPI is a common medication identifier for both Medispan and Clarity medications, which is considered a more accurate info than the medication name. Medication Orders were

then matched on the GPI and then imported along with their accompanied RxCUI from RxNAV and Medispan into table RZ_CODE_MAP_OPTION.

Any medication with the same GPI and different RxCUI is considered a mismatch between the Medi-Span mapping output and the automated RxNav mapping (with business rules applied). The discrepencies were then analyzed and justified. Finally, one RxCUI per EHR medication is selected.

Chapter 3: Results

3.1. Data Quality

3.1.1 Data Completeness Analysis

An export was generated of the medications that were missing NDC codes. Each medication ID was exported with its accompanying record identifier; generic name, form, route, and strength. When analyzing the data, it was found out that many records are missing one or more of its specifications; strength, route, form. This, in many cases, might direct the mapping to a more general term. Table (3) shows the difference in mapping two different records with complete and missing information, correspondingly. Mapping to a more general term will result in getting incomplete information about the medication the patient is taking and hence this does not satisfy study objectives for researchers. More about this challenge will be deliberated in the discussion section.

Table 3 -RxNorm mapping results for a medication with complete VS missing information

Medication specifications	RxCUI	RxNorm name
Metoprolol succinate ER 25 mg tab	866427	24 HR metoprolol succinate 25 MG Extended Release Oral Tablet
Metoprolol po	1163523	Metoprolol Oral Product

Table (4) shows the completeness of the EHR medication records that were used in this research. The results in the table below reveal that records with missing form and strength represents 57.9 % of the records which have the highest occurrence rate. On the other hand, only 34.1% of the medication orders records have all four specifications available filled in their designated fields. This means that 65.9 % of the records are missing one or more of the product identifiers. Another observation from this table is that the form is usually available with a rate of more than 99%.

Table 4 - Analysis of the completeness of EHR medication records in Clarity tables

Missing Information	Count	Percentage
Form only	2	0.05
Route only	1	0.03
Strength only	239	6.44
Form and Strength	2150	57.90
Form and Route	0	0.00
Route and Strength	0	0.00

Missing none	1266	34.10
Missing All	55	1.48
Total	3713	100

3.1.2 Pharmacological Drug Groups

A drug is usually classified either by the chemical type of the active ingredient or by the way it is used to treat a particular condition. Each drug can be classified into one or more drug classes. Table (5) clarifies different drug classes covered by the database.

Table 5 - Major pharmacological drug groups covered by the dataset

Pharmacological group	Examples
Gastrointestinal tract/ metabolism	Antacids, Laxatives/Antidiarrhoeals, Anti-diabetics, Vitamins/Dietary minerals
Blood and blood forming organs	Antithrombotics, Antiplatelets, Anticoagulants, Antifibrinolytics, thrombolytics/fibrinolytics,
Cardiovascular system	Cardiac therapy/antianginals, Antihypertensives, Antihyperlipidemics
Skin	Emollients, Antipruritics, Antipsoriatics, Medicated dressings
Genitourinary system	Hormonal contraception, Fertility agents, SERMs, Sex hormones
Endocrine system	Hypothalamic-pituitary hormones, Sex hormone, Thyroid hormones/Antithyroid agents
Infections and infestations	Antimicrobials, Vaccines

Malignant disease	Anticancer agents, Antineoplastic
Immune disease	Immunomodulators, Immunostimulants, Immunosuppressants
Muscles, bones, and joints	Anabolic steroids, Anti-inflammatories, NSAIDs, Antirheumatics, Corticosteroids, Muscle relaxants, bisphosphonates
Brain and nervous system	Analgesics, Anorectics, Anti-ADHD , Anticonvulsants, Antidementia Antimigraine , Antiparkinson's , Antipsychotics, Anxiolytics, Depressants Euphoriants, Hallucinogens, Psychedelics
Respiratory system (R)	Decongestants, Bronchodilators, Cough medicines, H1 antagonists
Sensory organs (S)	Ophthalmologicals, Otologicals
Not covered	Antidotes, Anesthetics ,Contrast media, Radiopharmaceuticals Dressings

3.2 Mapping Results

Table (6) demonstrates the ability of the automated mapping approach (using Exact or approximate match) followed by the manual review by an expert for the list of retrieved mapping results. The results show a 72.6% (Category 1- table 9) right map for Clarity medications to RxNorm. On the other hand, 2.9% of the medication records are missing some specifications; hence we had to map the medication record to RxNorm concept that has the same level of granularity to avoid adding extra information than given.

Table 6 : Mapping results according to the closeness of the RxCUI to the medication information offered from EHR.

Category	Implication	Medication records		
		N	%	
1	Exact match was found	95	3.2	72.6
	The RIGHT RxCUI was selected from the retrieved list of APPROXIMATE match options	2134	69.4	
2	The RIGHT RxCUI was not found by exact and approximate mapping. The RIGHT RxCUI was found MANUALLY	262	8.9	
3	The CLOSEST* RxCUI was selected from the retrieved list of APPROXIMATE match options	75	2.6	
4	The CLOSEST* RxCUI was not listed within the retrieved option and was found MANUALLY	8	0.3	
5	No RxCUI was found	233	7.9	
6	RxCUI selected represents a very general term.	227	7.7	
	Total	2939	100.00	

*Closest match means that there is a slight difference between the RxCUI description and the medication record information. For example, Vitamin D3 2000 oral tablet and oral capsule.

Table (7) presents the reasons for incomplete or missing match (category 5 and 6). This can be attributed to differences in dosage, strength, and route form. Various issues regarding deficient mappings for drug concepts were identified.

Table 7 : Examples of reasons that some records were not mapped to a valid RxCUI or mapped to very general term.

Reason	Example
Obsolete term, term is no longer used.	ZICAM COLD REMEDY NA; recalled product

Cosmetic	VANICREAM EX; cosmetic product are not covered in RxNorm even if they are used for a medical condition Ex dry skin or eczema
OTC not in RxNorm	HEALTHY COLON PO, LIPOFLAVONOID PO
Too complex name	GLUCOS-CHONDROIT-CA-MG-C-D PO
Unclear ingredient / unclear name	JOINT FORMULA OR
Unavailable strength or name	SENNA HERBAL LAXATIVE 50-12 MG OR CAPS
Compounds	MAGIC MOUTHWASH (DUKE) (FV COMPOUNDED) SUSPENSION, GABAPENTIN 8%/ VANICREAM (FV COMPOUNDED) CREAM
Unavailable form-ingredient combination	VITAMINS B1 B6 B12 IJ INJ; This combination is only available as oral product not injectable
Missing one or more of the active ingredients	NEOMYCIN-POLYMYXIN (OTIC) SUSP OT; this product does not exist, it should be Hydrocortisone / Neomycin / Polymyxin B Otic Suspension. Hence it's missing the hydrocortisone part.
Some herbal products	GLUCOS-MSM-C-MN-GINGER-WILLOW PO, HERBAL ENERGY COMPLEX PO
Miscellaneous	Mouth wash, NONI JUICE PO,CHOLEST OFF COMPLETE PO

3.4 Validation Using Medispan

After comparing RxCUIs generated by Medispan and the automated RxNav mapping (with business rules identified), a spreadsheet was then generated 6,809 medication records that had been mapped to codes that the Medi-Span file disagrees with. Most of these are simple things like Medi-Span file chose the generic over the brand name See table (8) for a List of discrepancy reasons. A researcher (SA, RPh) reviewed this

spreadsheet and selected the more accurate choice among the results generated from the semi-automated process and Medi-span mapping. Table (9) shows the results of this validations process. More than three thirds (75.9%) of the preferred RxCUIs were generated using the auto mapping approach.

Table 8 : List of reasons for the discrepancies between RxCUI generated by Medisapn and RxNAV API.

Reason	Medication Name	Auto-mapping Result	Medi-Span Mapping Result
Medispan added more information than given	CALCIUM + D PO	Calcium Carbonate / Cholecalciferol Oral Product	308896 - Calcium Carbonate 250 MG / Vitamin D 125 UNT Oral Tablet
Medispan chose generic name while auto-mapping chose brand	NASONEX 50 MCG/ACT NA SUSP	746201-120 ACTUAT mometasone furoate 0.05 MG/ACTUAT Nasal Inhaler [Nasonex]	746199 - 120 ACTUAT mometasone furoate 0.05 MG/ACTUAT Nasal Inhaler
Medispan Specifies quantity that is not mentioned	FLUTICASONE-SALMETEROL 250-50 MCG/DOSE IN AEPB	Fluticasone propionate 0.25 MG/ACTUAT / salmeterol 0.05 MG/ACTUAT Dry Powder Inhaler	896186 - 14 ACTUAT Fluticasone propionate 0.25 MG/ACTUAT / salmeterol 0.05 MG/ACTUAT Dry Powder Inhaler
Auto added brand name when not mentioned	FLUOXETINE HCL 20 MG PO CAPS	Fluoxetine 20 MG Oral Capsule [Prozac]	310385 - Fluoxetine 20 MG Oral Capsule
Different form	MECLIZINE HCL 25 MG PO	Meclizine Hydrochloride	995666 - Meclizine

	TABS	25 MG Chewable Tablet	Hydrochloride 25 MG Oral Tablet
Auto is missing one component	NYSTATIN 100000 UNIT/GM EX POWD	Nystatin Topical Powder	Nystatin 100 UNT/MG Topical Powder

Table 9 : Validation findings when comparing results generated by Medisapn and RxNAV

RxCUI selected Source	Number of records from the selected source	Percentage of records that are linked to the correct RxCUI
Auto Mapping	5170	75.9 %
Medispan	1518	22.3 %
Not found, needed manual mapping	121	1.8 %
Total	6809	100 %

In this study, the most granular RxCUI available was chosen when mapping the medication record. For example, brand name is a more precise approach of naming any medication because it limits the number of NDCs and thus the number of RxCUIs to choose from. Hence, choosing an RxCUI that specifies the brand in addition to the generic name was considered to be an accurate selection when mapping branded name medications. Nevertheless, integrating a brand name to the concept would be an unnecessary addition to a medication name that does not specify a brand, even when

there is no generic product available for that medication during the time the study was conducted.

Chapter 4: Discussion & Conclusion

4.1 Discussion

This study involved enhancements to automated mapping approach to map 3713 unique medications with incomplete medication information (used by about 19,560,024 medication orders) to RxNorm using RxNorm API. The automated mapping was enhanced by the addition of mapping rules that were identified by a researcher (SA, RPh) and then validated by comparing RxNav results to mapping results created by another mapping tool (Medi-Span). In this study, it was found that the total success rate of mapping medications with incomplete information to the correct RxNorm concept equals to 72.6%. The matching rate becomes 87% when mapping results of 53,000 medications using the automated approach were compared to the mapping results of Medi-Span tool; which was used to validate the study approach. It is anticipated that the enhanced automated mapping approach would support health information exchange and medication reconciliation process.

For the extracted medication orders with missing medication NDC, it was found that 65.9% of these records were also missing one or more essential medication information. In any EHR system, the medication information completeness varies at the point of

medication order entry to meet various purposes, including but not limited to, inpatient medication order entry, pharmacy inventory, and checking duplication of therapy and/or drug allergy [63#50]. For example, when checking drug allergies it is adequate to record the medication name without the strength or form since a patient will be allergic to the active ingredient regardless of the form or strength i.e. penicillin allergy. However, when stocking up a medication for a pharmacy inventory, it's important to record all specifications of the medication including form, strength, and package size.

Previous studies have shown that the incompleteness of drug signature elements yields a number of possible RxNorm mappings for the same medication. The mapping effort in this case becomes costly and time consuming since it relies heavily on manual expert interactions. The studies state that an automated method would make the process manageable especially for the mapping of large number of medications [23, 40, 41].

When using the automated mapping approach, it was found that only 3.2% of Clarity records included in this study can be directly mapped to one exact match in RxNorm that can be identified using the exact match approach. These medication orders already came with enough information that can be used to identify one RxNorm concept. In addition, about 69.4% of the medications in this study utilized the approximate matching approach to identify the correct RxNorm concept utilizing the identified business rules. This made the total success rate of the medications that were able to correctly identify a matching

RxNorm concept equal to 72.6%. Prior to the development of RxNorm, several pharmacy system knowledge base vendors embarked on an attempt to match terms across different terminologies and had a 53% success rate [51]. Better success rate was reported by other studies when medication orders with data issues were excluded for the mapping effort [62#36, 58#35]. These studies addressed challenges such as the inclusion/exclusion of salts, different representations of strength, and naming conventions. The studies also reported that further analysis is needed to understand the underlying data issues that impact the mapping rate and to identify recommendations to enhance mapping rates. Such studies are critical for the future adoption and integration of RxNorm in EHRs [42, 36].

Compared to our approach of using mapping rules to enhance automated mapping, previous studies focused on using a completely manual approach for the mapping of medications with missing information. One of these studies provided by Saitwal et al. [5] provided a methodological review of twelve terminological systems highlighting how medication information is distributed across these systems and how they are linked through common codes. In their study, only 62.5% of source medication codes could be mapped automatically. The medications with missing information were mapped using a completely manual approach, which as mentioned above is very time and effort consuming.

The majority of the medication orders had several issues that required going through the approximate match process. Some of these issues include, not specifying the active ingredient of the product, for example, COLON HEALTH PO. In addition, the medication name may contain multiple ingredients that are too complex to be mapped to a specific RxCUI such as different types of multivitamins. In such a situation, manual intervention was required to map the medication term to a more general term to minimize the number of records that do not map to any precise RxCUI. When the medication record, COLON HEALTH PO was further inspected, it was found that the active ingredient is lactobacillus and was mapped to LACTOBACILLUS ACIDOPHILUS ORAL PRODUCT. In this way, the med record was mapped to a relevant concept and distinguished from other unmapped records, however, it would take extra time and effort from the researcher to investigate the product's active ingredient, which might be inconvenient and expensive for research purposes.

Relative contribution of each component of the automated mapping process such as name splitting, drug name expansion, spelling correction and using the identified business rues all helped in finding a match to the targeted concept. For example, RxNorm did not recognize "LEVOTHYROXINE SODIUM 88 MCG PO CAPS"; the name had to be expanded to the following term "LEVOTHYROXINE SODIUM 0.088 MG ORAL CAPSULE" so it can be read flawlessly by RxNorm.

The identified business rules were also able to address some of the medication information issues for the support of automated mapping. One of these helpful rules was mapping drug names with missing strength and/or form information to a short name. For example, when looking for "POTASSIUM GLUCONATE PO", the strength and form were overlooked and the medication record was mapped to "POTASSIUM GLUCONATE ORAL PRODUCT" which matches the level of granularity of the medication term in the dataset.

In addition, mapping to the active ingredient was one of the identified business rules for mapping medication orders that did not have a specific strength, route, or form. For example, SIMVASTATIN was mapped to RxNorm name SIMVASTATIN. This approach helped classify drugs under their therapeutic class hierarchy instead of leaving it unidentified. Consequently, it helps researches identify patients with hyperlipidemia from hypertensive or diabetic patients, identify duplication in therapy, detect drug-drug interactions in poly-pharmacy patients, or determining DEA scheduling of the recorded medication.

Another business rule that has several benefits is mapping medication orders to RxNorm concepts that are associated to a slightly different term from the original record name in the database. Those minor differences should make no significant influence on the intended medication action or use. For example POLICOSANOL TAB 10 MG was

mapped to RxNorm concept POLICOSANOL 10 MG ORAL CAPSULE that only differs in the form type. This helps identify the active ingredient, strength and route, which are the most vital pieces of information for any medication especially in inpatient order entries.

The results of the validation process illustrated that the identified business rules can be used to enhance the automated mapping process. For 53,000 medications, there was an 87% matching rate in the identified RxCUI generated by both the automated process (with business rules applied) and the Medi-Span tool. For the records that have different mapping results by the automated process and Medi-Span, the expert (SA, RPh) manual review found that about 75.6% of the times the correct mapping results were actually identified by the automated process with the business rules applied. One of the reasons behind the mismatch between the automated process and Medi-Span is that there were some invalid RxCUIs that were used by Medispan but actually are no longer used by RxNorm. In RxNorm, some dosage forms have been retired and replaced by more generic dose form names, allowing for greater flexibility in representing drug names. Other dose forms have been added. At this time, manual review was the only way to check the validity in the RxCUIs used when a discrepancy was revealed.

Another reason for mismatch between the automated approach and Medispan was that Medispan selected branded name RxCUIs over generic ones where the original name in

the dataset did not specify any brand name. Instead, the automated mapping only mapped to the brand name if it was declared within the record name itself or it's specifications.

The following points explain some cases where choosing brand name for mapping purposes over the generic is preferred if, and only if, the brand name is provided:

- Brand name provides a more specific level of granularity (just second best after NDC). For example, "Lantus Solostar" is the brand name for the only available product of the active ingredient Insulin Glargin in a flex pen injectable form. Hence from the brand name we can obtain the exact NDC and map it to the exact match RxCUI. However, this would need an expert opinion to tell if there is one ingredient/route/form/packaging combination or more with the same brand name.

- Specifying the brand name reduces the number of RxCUIs generated to select from. For Example Spiriva comes in one strength/form/Route combination, however it comes in two different package sizes; 30 caps and 90 caps. So between those 2 options it would be safe enough - most of the times- to map the source medication to either one.

- Some brands are not bioequivalent Despite having the same strength/form/route/active ingredient combination. For example "Synthroid 25 mcg po tab" is not interchangeable with " Levoxyl 25 mcg po tab" although both of them have the same active ingredient; levothyroxine sodium or LT4. [52]

Brand versus generic always play an important role in the formulary list. Many brand names are not covered, not preferred or would need prior authorization from the insurance company to approve its coverage if proved to be medically necessary.

The findings of this study not only illustrate the applicability of RxNorm standards to coding medications used in the CDR database for clinical research, but also identified multiple issues related to recording and processing of medication information that would need to be addressed in order to facilitate building and deploying interoperable electronic health record systems.

Previous studies reported that the analysis of EHR medication data issues is critical for the future adoption and integration of RxNorm in EHRs [42,36]. This study analyzes large number of EHR medication orders with the focus of identifying issues, analyzing the impact to RxNorm mapping and describing an approach that can be used to enhance the automated mapping. RxNorm concepts have been utilized for coding clinical medication data in many clinical research studies and applications. However, there are very few studies that analyze the underlying reasons of the EHR medication order data issues and the impact to RxNorm mapping. It is anticipated that this study will enrich the literature with a large EHR medication order analysis. The results of this study are expected to lead to better clinical research studies and applications that utilize RxNorm concepts. Ultimately, this will enhance health care quality and reduce healthcare cost. It is also anticipated that the results can be used to identify more business rules or refine

current ones to design a more robust rule-based system for the support of automated mapping tools with minimal human intervention.

4.2 Limitations

This section covers few limitations that were encountered during this study. Some were related to RxNorm and some related to the Clarity Database. Generalization of our findings may be impacted as a result of these limitations.

RxNorm is intended to cover all prescription medications approved for use in the United States. Prescription medications from other countries may not be correctly mapped to RxNorm concepts. The same also applies to some Over-the- counter (OTC) medications. RxNorm are currently analyzing the opportunities to add and cover OTCs when reliable information about the medications can be found and when they appear to be represented in other UMLS source terminologies. If a medication source is out the designated source vocabularies included in RxNorm then the medication will not be mapped to a valid RxCUI. In addition, medications with more than four ingredients, whether prescription or OTC, are not fully represented at the present time. In some cases (e.g., multivitamins) it may not be possible to include all of them in a reasonable time frame.

Comprehensive terminology mapping evaluation process may require significantly more time, effort, and terminology-specific expertise than can be predicted at the start of a project. Because these systems continue to evolve, ongoing use of any cross-terminology

mappings must also include a plan to accommodate these changes, such as occurs when codes are given different meanings, removed from a system, or new codes are added.

4.3 Future Work

To address these challenges in future work, we need to research better methods for automatically linking concepts across systems and maintaining these links.

As with the case study presented here, creating new connections between systems often requires utilizing rule-based systems that captures experts' knowledge and translates it into a programmable rule format. These rules based technique would be used to enhance an NLP system that is able to identify medical terms and deal with data issues as terms expansion. Another approach, as explored in the Galen project [22], is to provide users with a convenient language for logically defining the terms and relations in an ontology using a common reference model, followed by the application of algorithms that can infer conceptual mappings and class/sub- class relationships among terms from different source vocabularies. A combination of these approaches is worth exploring, because satisfactory automated approaches that do not require significant human intervention may not be possible without significant artificial intelligence breakthroughs. Finally, a highly interactive partially-automated mapping tools that are directed by human experts to automate parts of the mapping process with specific expert input are a promising alternative to fully automated methods.

4.4 Conclusion

The enhanced automated mapping approach that utilizes NLM RxNav API in addition to the identified mapping business rules was useful for mapping the majority of medication records with missing information. While expert judgment is still needed, it is minimized by creating a short list of suggested target RxNorm concepts for the expert to pick from for a given EHR medication name. The study limitations suggest that a more comprehensive study from multiple healthcare institutes may refine the mapping rules identified by this study with the goal of designing a more robust system.

References

1. Weiner, M. G., Lyman, J. A., Murphy, S., & Weiner, M. (2007). Electronic health records: high-quality electronic data for higher-quality clinical research.*Informatics in primary care*, *15*(2), 121-127.
2. Wilson, J. (n.d.). The benefit of using both claims data and electronic medical record data in health care analysis. Retrieved January 26, 2015, from http://www.optum.com/content/dam/optum/resources/whitePapers/Benefits-of-using-both-claims-and-EMR-data-in-HC-analysis-WhitePaper-ACS.pdf
3. Hsiao, C. J., Hing, E., Socey, T. C., & Cai, B. (2011). Electronic health record systems and intent to apply for meaningful use incentives among office-based physician practices: United States, 2001–2011. *system*, *18*(17.3), 17-3.
4. Teich, J. M., Osheroff, J. A., Pifer, E. A., Sittig, D. F., & Jenders, R. A. (2005). Clinical decision support in electronic prescribing: recommendations and an action plan: report of the joint clinical decision support workgroup. *Journal of the American Medical Informatics Association*, *12*(4), 365-376
5. Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., & Taylor, R. (2005). Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Affairs*, *24*(5), 1103-1117
6. Xu, H., Stenner, S. P., Doan, S., Johnson, K. B., Waitman, L. R., & Denny, J. C. (2010). MedEx: a medication information extraction system for clinical narratives. *Journal of the American Medical Informatics Association*, *17*(1), 19-24.
7. Doan, S., Bastarache, L., Klimkowski, S., Denny, J. C., & Xu, H. (2010). Integrating existing natural language processing tools for medication extraction from discharge summaries. *Journal of the American Medical Informatics Association*, *17*(5), 528-531.
8. Boockvar, K. S., Blum, S., Kugler, A., Livote, E., Mergenhagen, K. A., Nebeker, J. R., ... & Yeh, J. (2011). Effect of admission medication reconciliation on adverse drug events from admission medication changes. *Archives of internal medicine*, *171*(9), 860-861.
9. Preventable Adverse Drug Reactions: A Focus on Drug Interactions. (n.d.). Retrieved January 26, 2014, from http://www.fda.gov/drugs/developmentapprovalprocess/developmentresources/druginteractionslabeling/ucm110632.htm
10. Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.). (2000). *To Err Is Human:: Building a Safer Health System* (Vol. 627). National Academies Press.
11. Liu, S., Ma, W., Moore, R., Ganesan, V., & Nelson, S. (2005). RxNorm: prescription for electronic drug information exchange. *IT professional*, *7*(5), 17-23.

12. Nelson, S. J., Zeng, K., Kilbourne, J., Powell, T., & Moore, R. (2011). Normalized names for clinical drugs: RxNorm at 6 years. *Journal of the American Medical Informatics Association*, *18*(4), 441-448.
13. Pathak, J., Murphy, S. P., Willaert, B. N., Kremers, H. M., Yawn, B. P., Rocca, W. A., & Chute, C. G. (2011). Using RxNorm and NDF-RT to classify medication data extracted from electronic health records: experiences from the Rochester Epidemiology Project. In *AMIA Annual Symposium Proceedings*(Vol. 2011, p. 1089). American Medical Informatics Association.
14. Mork JG, Bodenreider O, Demner-Fushman D, Doğan RI, Lang FM, Lu Z, Névéol A, Peters L, Shooshan SE, Aronson AR:Extracting Rx information from clinical narrative. Journal of the American Medical Informatics Association 2010,17:536-539.
15. RxNorm Technical Documentation. (2012, April 2). Retrieved May 12, 2014, from http://www.nlm.nih.gov/research/umls/rxnorm/docs/2012/rxnorm_doco_full_2012-2.html
16. Pahor, M., Chrischilles, E. A., Guralnik, J. M., Brown, S. L., Wallace, R. B., & Carbonin, P. (1994). Drug data coding and analysis in epidemiologic studies.*European journal of epidemiology*, *10*(4), 405-411.
17. O'Neill, S. M., & Bell, D. S. (2010). Evaluation of RxNorm for representing ambulatory prescriptions. In *AMIA Annual Symposium Proceedings* (Vol. 2010, p. 562). American Medical Informatics Association.
18. Levinson DR. The Food and Drug Administration's National Drug Code Directory. Department of Health and Human Services, Office of Inspector General; Available at: http://oig.hhs.gov/oei/reports/oei-06-05-00060.pdf (accessed May 11, 2014)
19. Patrick, J., & Li, M. (2010). High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge.*Journal of the American Medical Informatics Association*, *17*(5), 524-527.
20. Estabrooks, P. A., Boyle, M., Emmons, K. M., Glasgow, R. E., Hesse, B. W., Kaplan, R. M., ... & Taylor, M. V. (2012). Harmonized patient-reported data elements in the electronic health record: supporting meaningful use by primary care action on health behaviors and key psychosocial factors. *Journal of the American Medical Informatics Association*, *19*(4), 575-582.
21. Botsis, T., Hartvigsen, G., Chen, F., & Weng, C. (2010). Secondary use of EHR: data quality issues and informatics opportunities. *AMIA summits on translational science proceedings*, *2010*, 1.
22. Hernandez, P., Podchiyska, T., Weber, S., Ferris, T., & Lowe, H. (2009). Automated mapping of pharmacy orders from two electronic health record systems to RxNorm within the STRIDE clinical data warehouse. In *AMIA Annual Symposium Proceedings* (Vol. 2009, p. 244). American Medical Informatics Association.
23. Adamusiak, T., Shimoyama, N., & Shimoyama, M. (2014). Next Generation Phenotyping Using the Unified Medical Language System. *JMIR medical informatics*, *2*(1).
24. Zeng, K., Bodenreider, O., Kilbourne, J., & Nelson, S. (2007). RxNav: towards an integrated view on drug information. In *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems* (p. 2400). IOS Press.
25. Peters, L., & Bodenreider, O. (2008). Using the RxNorm web services API for quality assurance purposes. In *AMIA Annual Symposium Proceedings* (Vol. 2008, p. 591). American Medical Informatics Association.
26. Kuperman, G. J., Teich, J. M., Gandhi, T. K., & Bates, D. W. (2001). Patient safety and computerized medication ordering at Brigham and Women's Hospital. *Joint Commission Journal on Quality and Patient Safety*, *27*(10), 509-521.

27. Poon, E. G., Keohane, C. A., Yoon, C. S., Ditmore, M., Bane, A., Levtzion-Korach, O., ... & Gandhi, T. K. (2010). Effect of bar-code technology on the safety of medication administration. *New England Journal of Medicine*, *362*(18), 1698-1707.

28. Nelson, S. J., Zeng, K., Kilbourne, J., Powell, T., & Moore, R. (2011). Normalized names for clinical drugs: RxNorm at 6 years. *Journal of the American Medical Informatics Association*, *18*(4), 441-448.
29. RxNorm. (n.d.). Retrieved January 26, 2014, from http://www.nlm.nih.gov/research/umls/rxnorm/
30. Kannry, J. L., Wright, L., Shifman, M., Silverstein, S., & Miller, P. L. (1996). Portability issues for a structured clinical vocabulary: mapping from Yale to the Columbia Medical Entities Dictionary. *Journal of the American Medical Informatics Association*, *3*(1), 66-78.
31. Sherertz, D. D., Tuttle, M. S., Blois, M. S., & Erlbaum, M. S. (1988, November). Intervocabulary Mapping Within the UMLS: The Role of Lexical Matching*. In*Proceedings/the... Annual Symposium on Computer Application [sic] in Medical Care. Symposium on Computer Applications in Medical Care* (pp. 201-206). American Medical Informatics Association.
32. Kannry, J. L., Wright, L., Shifman, M., Silverstein, S., & Miller, P. L. (1996). Portability issues for a structured clinical vocabulary: mapping from Yale to the Columbia Medical Entities Dictionary. *Journal of the American Medical Informatics Association*, *3*(1), 66-78.
33. Evans, D. A., Rothwell, D. J., Monarch, I. A., Lefferts, R. G., & Cote, R. A. (1990). Toward representations for medical concepts. *Medical decision making: an international journal of the Society for Medical Decision Making*, *11*(4 Suppl), S102-8.
34. Cimino, J. J., Hripcsak, G., Johnson, S. B., & Clayton, P. D. (1989, November). Designing an introspective, multipurpose, controlled medical vocabulary. In*Proc 13th Annu Symp Comput Appl Med Care* (pp. 513-8).
35. Hernandez, P., Podchiyska, T., Weber, S., Ferris, T., & Lowe, H. (2009). Automated mapping of pharmacy orders from two electronic health record systems to RxNorm within the STRIDE clinical data warehouse. In *AMIA Annual Symposium Proceedings* (Vol. 2009, p. 244). American Medical Informatics Association.
36. Richesson, R. L. (2014). An informatics framework for the standardized collection and analysis of medication data in networked research. *Journal of biomedical informatics*.

37. O'Neill, S. M., & Bell, D. S. (2010). Evaluation of RxNorm for representing ambulatory prescriptions. In *AMIA Annual Symposium Proceedings* (Vol. 2010, p. 562). American Medical Informatics Association.
38. Bennett, C. C. (2012). Utilizing RxNorm to support practical computing applications: Capturing medication history in live electronic health records.*Journal of biomedical informatics*, *45*(4), 634-641.
39. Fung, K. W., McDonald, C., & Bray, B. E. (2008). RxTerms–a drug interface terminology derived from RxNorm. In *AMIA Annual Symposium Proceedings*(Vol. 2008, p. 227). American Medical Informatics Association.
40. Peters, L., Kapusnik-Uner, J. E., & Bodenreider, O. (2010). Methods for managing variation in clinical drug names. In *AMIA Annual Symposium Proceedings* (Vol. 2010, p. 637). American Medical Informatics Association.
41. Cimino, J. J. (1996). Formal descriptions and adaptive mechanisms for changes in controlled medical vocabularies. *Methods of information in medicine*, *35*, 202-210.
42. Zhou, L., Plasek, J. M., Mahoney, L. M., Chang, F. Y., DiMaggio, D., & Rocha, R. A. (2012). Mapping partners master drug dictionary to RxNorm using an NLP-based approach. *Journal of biomedical informatics*, *45*(4), 626-633.

43. Tools & Services. (n.d.). Retrieved January 26, 2014, from https://healthinformatics.umn.edu/research/ctsi-bmi-university-minnesota/tools-services
44. Bodenreider, O., & Peters, L. (2010). RxNav: browser and application programming interfaces for RxNorm. In *AMIA annual symposium, Washington*(p. 1330).
45. RxNav User Guide. (n.d.). Retrieved November 26, 2013, from http://rxnav.nlm.nih.gov/RxNavDoc.html
46. RxNorm API. (n.d.). Retrieved September 20, 2014, from http://rxnav.nlm.nih.gov/RxNormAPIs.html
47. Approximate Matching in the RxNorm API. (n.d.). Retrieved January 26, 2014, from http://rxnav.nlm.nih.gov/RxNormApproxMatch.html
48. Abbreviations and Acronyms. (n.d.). Retrieved January 26, 2014, from http://rxnav.nlm.nih.gov/Abbreviations.html
49. Galhardas, H., Lopes, A., & Santos, E. (2011). Support for user involvement in data cleaning. In *Data Warehousing and Knowledge Discovery* (pp. 136-151). Springer Berlin Heidelberg.
50. Arimilli, R., Farooki, M. U., & Mohapatra, O. (2010). *U.S. Patent No. 7,706,915*. Washington, DC: U.S. Patent and Trademark Office.
51. Centers for Medicare & Medicaid Services, and HHS. (2012). Medicare and medicaid programs; electronic health record incentive program – stage 2. Final rule. *Fed. Regist.* 77, 53967–54162.
52. Dong, B. J., Hauck, W. W., Gambertoglio, J. G., Gee, L., White, J. R., Bubp, J. L., & Greenspan, F. S. (1997). Bioequivalence of generic and brand-name levothyroxine products in the treatment of hypothyroidism. *JaMa*, *277*(15), 1205-1213.