Medication and Indication Linkage:
A Practical Therapy for the Problem List?

Matthew M. Burton, MD[1, 2, 3]; Linas Simonaitis, MD, MSc[1, 2];

Gunther Schadow, MD, PhD[1, 2, 3]

[1]Regenstrief Institute, Inc.; [2]Indiana University School of Medicine; and [3]Indiana University School of Informatics

Background: *Establishing a relationship between medications and diagnoses within a functioning electronic medical record system (EMR) has many valuable applications, such as improving the quality and utility of the problem list to support better decisions.*

Methods: *We evaluated over 1.6 million de-identified patient records from the Regenstrief Medical Record System (RMRS) with over 90 million diagnoses and 20 million medications. Using RxNorm, the VA National Drug File Reference Terminology, and SNOMED-CT (S-CT) standard terminologies and mappings we evaluated the linkage for local concept terms for medications and problems (diagnoses & complaints).*

Results: *We were able to map 24,398 candidates as medication and indication pairs. The overall sensitivity and specificity for term pairs was 67.5% and 86% respectively and 39.5% and 97.4 when adjusted for term pair occurrence within single patient records.*

Conclusions: *Medications can be mapped by machine to a disease/ disorder using established terminology standards. This mapping may inform many knowledge management and decision support features in an EMR.*

Introduction

One of the first concerns that a Physician has about a patient is: "What diseases, symptoms, or problems does this patient have?" The need to approach patient care more systematically and scientifically using Problem Lists has long been advocated.[1,2,3,4,5] Weed argued that "Each Medical Record should have a complete list of all the patient's problems, including both clearly established diagnoses (and) abnormal findings or symptoms."[1]

The Institute of Medicine's report on The Computer-Based Patient Record (CPR) states that "a problem list that clearly delineates the patient's clinical problems and the current status of each" as the first of 12 desired attributes of a comprehensive CPR. Additionally, it clearly states the advantage of this approach.[2] Finally, it has been demonstrated that even in the context of utilizing other systematized organizations of domain knowledge for clinical practice, such as ADCVAANDIML, for Computerized Provider Order Entry (CPOE), physicians still resort to a problem-based mental model.[6]

The second question that a Physician has about the problem is "how is this being addressed?" Recent policy and technology developments have focused significant resources on the issue of medication reconciliation and improving the quality and completeness of the medication list. In 2006, the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) issued a mandate for accredited organizations to implement medication reconciliation best practices.[7]

Unfortunately, while a candidate for the medications list can be created from dispense records[8], in most clinical settings there is no well-managed problem list. While most Electronic Medical Records systems (EMRs) do have a Problem List feature, it is often poorly maintained and wrought with error.[4] It can be inaccurate, outdated, confusing, and time consuming at a critical point in the patient's care. Therefore, physicians frequently use the medications list as a proxy for the active problems as an alternative to scanning a long, unorganized list of diagnoses, symptoms, procedures and complaints often cluttered with the novelties of medical billing.

In this paper we investigate whether we can automatically assist our physicians in this thought process, by connecting the problems on the problem list with the active medications found in orders and dispense records. Our ultimate goal is to represent the problem list and medications list in a manner that allows the physician to easily make an assessment of the patients current and active issues and address them accordingly.

In the U.S., the pharmaceutical industry and the Food and Drug Administration (FDA) together are the authoritative source for drug information. The FDA drives an initiative to improve drug knowledge most notably through its Structured Product Labeling (SPL) initiative. SPL is a knowledge representation standard defined by Health Level-7 (HL7)[9] and uses its Reference Information Model (RIM)[10] to represent the chemical and physical nature of medical products as well as knowledge about their safe and effective use including diseases, symptoms, or problems for which they are indicated. However, there have been challenges in implementing this knowledge content.[11]
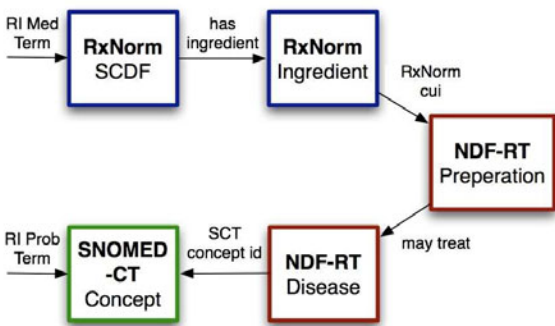
Figure 1: Mapping RI dictionary terms for medication orders via RxNorm to NDF-RT, and RI terms for problems and diagnoses through SNOMED-CT. NDF-RT indications are cross-referenced to SNOMED-CT. Finally the connection between SNOMED-CT indication and SNOMED-CT problem is connected by the SNOMED-CT subsumption hierarchy.

The FDA SPL indexing initiative has prioritized drug classifications and indications as essential knowledge to be added to SPL labels for use in decision support.[12] As part of this effort, they are utilizing the Veteran's Administration National Drug File Reference Terminology (NDFRT) [13]. In an effort of making such knowledge available earlier, we found that the NDF-RT already contains some of the necessary knowledge linkages between medications (codes) and their indications with SNOMED-CT concepts. See Figure 1.

Such mappings enable EMR's to leverage improvements in the medication list to infer relationships in the problem list. These relationships will likely improve the utility and effectiveness of the problem list for general clinical use, information display, clinical decision support as well as secondary uses of clinical data such as knowledge discovery, quality measures, and information retrieval.

Methods

Study Design

This study is a retrospective, de-identified cohort of patients with their problems and medication orders selected from the Regenstrief Medical Record System (RMRS).[14] The study was approved by the IRB as exempt research (EX0801-38). We selected 140,090 patients that had both problems (diagnoses and complaints) and medication orders in RMRS. RMRS has a central concept dictionary with terms for medications and problems (diagnoses and complaints).

Data and Analysis

We used the de-identified cohort to establish a frequency based summary of problem-medication pairs as they appear in combination in RMRS. A complete set of problem lists (diagnosis and complaints) that originate in the RI Gopher system were gathered from the Regenstrief Medical Record System (RMRS), excluding terms that didn't match to SNOMED-CT. Matching medication order records were selected across all INPC15 member institutions. All data was de-identified prior to exporting into a research database system for the experiment. The de-identification procedure replaced internal patient ids with random generated pseudo-ids and for each pa-tient, a random time offset of ± 180 days was generated and added to all dates of birth, allergy, and medication records. A fixed offset was added when necessary to hide ages beyond 90 years.

Medication Terminology Mapping

We previously mapped 2,283 RMRS medication terms to 3,692 RxNorm concept unique identifiers (CUIs). This accounts for 82.2% of terms and covers 90.1% of patients and 92.0% of term instances for medications in RMRS.

Diagnoses Terminology Mapping

We used the Unified Medical Language System (UMLS) to map 8,763 RMRS problem (diagnosis and complaint) terms to 11,561 SNOMED-CT concepts. This accounts for 43.7% of problem terms and covers 68.3% of term instances for problems in RMRS.

Linking Medications to Problems

We used NLM's RxNorm, the VA National Drug File-Reference Terminology (NDF-RT), and SNOMED-CT to map medications to potential indications as shown in Figure 1. NDF-RT contains a "may_treat" relationship from single-ingredient drug-products to NDF-RT disease concepts derived from co-occurrence data and the UMLS Metathesaurus using methods previously described.[16] NDF-RT contains mappings for most of these single ingredient products to RxNorm and most of the NDF-RT disease concepts to SNOMED-CT. Thus, NDF-RT connects ingredients specified in RxNorm to indications specified in SNOMED-CT.

We loaded all of the original SNOMED-CT distribution into the relational database and computed the full materialized transitive and reflexive closure using Warshall's algorithm implemented in SQL:

```
CREATE TABLE SCT_isa AS
  SELECT conceptid1 as spec,
         conceptid2 as gen
    FROM SCT_Relationship
    WHERE relationshipType = 116680003;

INSERT INTO SCT_isa
    SELECT c.spec, p.gen
      FROM SCT_isa c, SCT_isa p
      WHERE p.souce = c.target
```

```
   EXCEPT
     SELECT * FROM SCT_isa;
```

The above INSERT statement is repeated until nothing new is added (only 6 iterations). Finally the reflexive closure is completed by adding:

```
INSERT INTO SCT_isa
  SELECT conceptid, conceptid
    FROM Concept;
```

With this preparation, we can map indications to disorders or findings as easily as joining the problem and the indication through this transitive and reflexive relation.

Evaluation

To determine the sensitivity and specificity, we selected 1,000 distinct medication and problem pairs distributed equally and randomly from the mapped and non-mapped data sets. Some adjustments were made to reduce the occurrence of high frequency terms including limiting the maximum of any term to 10 per set. These term sets were then given to two Board Certified Internists who were blinded and asked to evaluate the problem list term as a reasonable indication for the medication. Concordance and Cohen's kappa coefficient (κ) were calculated for the raters. Finally, sensitivity and specificity analysis were performed on the term sets created with the mapping methodology.

Results

We were able to map 24,398 problem and medication term pairs as medication and indication pairs using the methodology described previously. There were an additional 2,087,412 problem and medication term combinations that co-occurred on at least one patient's record that did not map as medication and indication pairs.

Not surprisingly, some of the more common problems and their common treatments were the more frequent mapped medication and indication pairs as seen in Table 2. For this table, term pairs with duplicate terms in either problems or medications as previous pairs were omitted to demonstrate a more representative example of medica-

tion and indications. These pairs are ranked in descending order by frequency of co-occurrence in longitudinal patient records.

For our expert clinician analysis of medication to indication term mapping, the inter-rater concordance was 85.9% and $\kappa = 0.66$. For the purpose of sensitivity and specificity analysis, we decided to use the individual assessments and adjust the final analysis by disregarding term pairs where disagreement occurred. See Table 1. Of the 1,000 term pair sample, Rater 1 found 452 true positives, 247 false negatives, and only 52 false positives giving this mapping methodology an overall sensitivity of 64.7% and specificity of 82.7%. Rater 2 found 444 true positives, 259 false negatives, and only 56 false positives giving this mapping methodology an overall sensitivity of 63.2% and specificity of 81.1%.

Table 1: Sensitivity and Specificity of Map

Map	Rater 1		Rater 2		Adjusted	
	+	-	+	-	+	-
+	452	52	444	56	428	32
-	247	249	259	241	206	196
Sens	64.7%		63.2%		67.5%	
Spec	82.7%		81.1%		86.0%	

The overall sensitivity and specificity for the adjusted term pairs was 67.5% and 86% respectively. However, the sensitivity and specificity change significantly when adjusted for frequency of term pair occurrence within single patient records, 39.5% and 97.4% respectively. The lack of sensitivity is in part due to the disparity between the expert clinician's perception of "reasonably indicated" and the NDF-RT's relational concept of "may_treat". Both likely contain varying degrees of ambiguity. The increase in specificity is due to the inter-rater agreement on true positives, especially for high frequency terms. Table 3 lists the top ten non-duplicate term pairs determined to be false negatives by frequency of occurrence in single patient records. Duplicate problem or medication terms in any given pair were omitted to demonstrate a more representative set. These pairs are ranked in descending order by frequency of co-

Table 2: Top 10 Non-Duplicate Positive Maps.

Problem	Medication
otitis media	Amoxicillin Susp
UTI	Bactrim DS Eqv
asthma	Albuterol Inhl
back pain	Ibuprofen
bacterial vaginosis	Metronidazole
hypertension	Hydrochlorothiazide
abdominal pain	Promethazine
depression	Fluoxetine
chest pain	Aspirin
esophageal reflux	Esomeprazole

Table 3: Top 10 Non-Duplicate False Negatives.

Problem	Medication
URI	Robitussin DM, Susp
bronchitis	Albuterol Inhl
UTI	Metronidazole
asthma	Prednisone
sinusitis	Bactrim DS Eqv
chest pain	Nitroglycerin Sl
cellulitis	Cephalexin
diabetes mellitus	Benazepril
conjunctivitis	Sulfacet 10% Ot
rhinitis allergic	Fluticasone 0.05% Nasal Sp

occurrence in longitudinal patient records.

Discussion

Medications may be used to infer a number of actively managed problems (and diagnoses) on a problem list. While there are still significant limitations to this approach, we assert that a well managed, complete, and accurate medication list can be leveraged to organize, prioritize, and possibly infer a reasonably useful problem list.

Table 4: Errata of Interest.

Error	Problem	Medication
False +	Gastrointestinal System Disorder	Prednisone
False -	Diabetes mellitus	Insulin
False +	Low back pain	Promethazine
False +	Conjunctivitis	Cromolyn Nebulizer Sol
False -	Angina Pectoris	Aspirin
False -	Acute Bronchitis	Azithromycin

While there were clearly failings in both sensitivity and specificity in the mapping methodology, there are clearly conceptual distinctions between clinician perceptions of "reasonable indication" and reference terminology established relationships between medications and their respective indications. 12 of the top 13 false negatives by frequency involved forms of upper respiratory infections and symptomatic therapeutics such as antihistamines, decongestants, cough suppressants, and bronchodilators. See Table 3. Similar false negatives were seen with Chest Pain and Coronary Artery Disease treatments as well as those for some diabetic complications (ACE inhibitors). This may lend one to think that the NDF-RT indications relationships should be broadened to include symptom or complication relationships in its "may_treat" mapping. However, the authors suggest an approach that utilizes the semantic relationships between medication (or drug ingredient) and true indication as well as between symptom and disorder as shown in Figure 2.

Simply expanding the "may_treat" relationships was actually a significant source of false positives in our study. Our mapping indicated diabetes as an indication for both Maalox and Milk of Magnesia which are both used to treat gastric complications of diabetes. Such ambiguity in the "may_treat" relationships within the NDF-RT caused the majority of our false positives. The most frequent false positive was promethazine for pain. As most clinicians know, promethazine is an anti-emetic given with the analgesic meperidine to prevent its potential side-effects of nausea and vomiting. In this case,
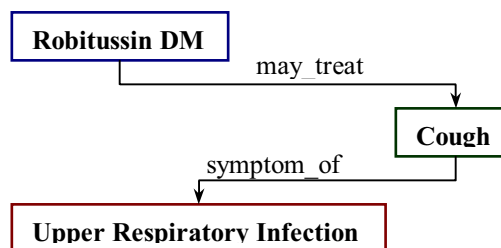

Figure 2. Ideal Linkage between the cough suppressant, Robitussin DM, and URI.

promethazine has a "may_prevent" relationship to nausea and vomiting which are "complications_of" the medication meperidine that "may_treat" the symptom pain. A further and obvious example of this ambiguity is the NDF-RT's "may_treat" relationship is "Disorder of the Gastrointestinal System" as an indication for prednisone. The remaining plurality of false positives were related to dose form (e.g. conjunctivitis as an indication for Cromolyn Nebulizer Sol) which is similarly ambiguous in the NDF-RT.

Interestingly, one of the false positives that occurred in a large number of patient records was the tricyclic antidepressant Amitriptyline as an indication for low back pain. It is interesting to note that the NDF-RT has allowed for this off-label use as an indication while our raters did not include low back pain as a reasonable indication for amitriptyline.

Another significant source of false negatives is a result of the lack of completeness of diagnoses terms in the NDF-RT indication terms and/ or the absence of linkage relationships between terms in SNOMED-CT. In many instances, NDF-RT attempts to use more generalized SNOMED-CT disorder and findings terms, but SNOMED-CT lacks the necessary linkages to the more specialized terms that map to problem list terms.

For instance, in the NDF-RT, the SNOMED-CT concepts that are indications for Aspirin include: Rheumatic Fever; Pain; Gout; Osteoarthritis; Inflammation; Fever; Arthritis, Rheumatoid. This allows Aspirin to map to Chest Pain, but not Angina Pectoris, Myocardial Infarction, Myocardial Ischemia, or Coronary Artery Disease. Similarly, the SNOMED concepts that are indications for Nitroglycerin include: Intraoperative Complications; Fissure in Ano; Heart Failure, Congestive; Esophageal Spasm, Diffuse; Eclampsia; Hypertension; Angina Pectoris; Myocardial Infarction; Hypertension, Pulmonary; and Myocardial Ischemia. Note that Chest Pain and Coronary Artery Disease are missing in these indications.

Additionally, evaluating drugs in the same class of drugs often yields similar, but not identical indication terms. For instance, erythromycin and azithromycin have four indication terms in common, but erythromycin has 10 distinct terms and azithromycin has 12. Simi-

larly, common drugs used to treat common diseases often fail to union on common indication terms. This is true for aspirin, nitroglycerin, lisinopril, and metoprolol on coronary artery disease.

A clear example of the absence of linkage relationships between terms in SNOMED-CT that caused numerous false negatives is antibiotics. For instance, in NDF-RT, azithromycin includes the SNOMED-CT concept "Respiratory Tract Infections" as an indication. However, in SNOMED-CT there is no relationship between "Acute Bronchitis" and "Respiratory Tract Infection".

A further point of interest is the notable change in both sensitivity and specificity when adjusted for the frequency of co-occurrence in patient records. The sensitivity decreases from 67.5% to 39.5% and the specificity increases from 86.0% to 97.4%. While there may be several explanations for this result, the systematic errors described above and term frequency (or prevalence) provide a framework and prioritization for focus of future work to improve the mapping.

Conclusion

Local mappings in combination with RxNorm, NDF-RT, and SNOMED-CT can be used to make inferences about the Problem List of a single patient record using the Medications List. While the sensitivity is much less than desirable, the specificity is reasonable, particularly in the context of term frequency. Furthermore, we have identified challenges and opportunities for these standards and in there utility as a knowledge base for improving the FDA's Structured Product Labels and use in clinical decision support. We believe the Problem Oriented Medical Record system is still achievable using open standards.

References

[1] Weed, LL. Medical Records that Guide and Teach. *NEJM*. 1968;278: 593-600.

[2] Institute of Medicine: The Computer-Based Patient Record: An Essential Technology for Health Care. 1997. Available from http://books.nap.edu/openbook.php?record_id=5306&page=R1

[3] AMIA: A Roadmap for National Action Plan on Clinical Decision Support. June 13, 2006. Available from: http://www.amia.org/inside/initiatives/cds/cdsroadmap.pdf

[4] Brown, SH. Empirical Derivation of an Electronic Clinically Useful Problem Statement System. *Ann Intern Med*. 1999;131:117-126

[5] Zeng, Q. Providing Concept-oriented Views for Clinical Data Using a Knowledge-based System. JAMIA. 2002;9:294-305.

[6] Johnson, CD. Task Analysis of Writing Hospital Admission Orderes: Evidence of a Problem-Based Approach. AMIA 2006 Symposium Proc. P. 389

[7] Available at http://www.jointcommission.org/NR/rdonlyres/9ECF1ED6-E04E-41DE-B7BC-174590CEDF33/0/07_NPSG_FAQs_8.pdf

[8] McDonald CJ. The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries.
Health Aff (Millwood). 2005 Sep-Oct;24(5):1214-20.

[9] Schadow G, Gitterman S, Boyer S, Dolin RH eds. HL7 v3.0 structured product labeling, release 2 [standard]. Ann Arbor, MI, Health Level Seven, 2005

[10] Schadow G, Russler DC, McDonald CJ. Conceptual Alignment of Electronic Health Record Data with Guidelines and Workflow Knowledge. Int J Med Inf. 2001 Dec;64(2-3):259-74

[11] Schadow, G. Assessing the Impact of HL7/FDA Structured Product Label (SPL) Content for Medication Knowledge Management. AMIA Annl Proc 2006: 646-650

[12] FDA: Guidance for Industry Indexing Structured Product Labeling [Draft Guidance]. March 2007. Available from http://www.fda.gov/CDER/GUIDANCE/7662dft.htm

13 Brown SH, Elkin PL, Rosenbloom ST, et al. VA National Drug File Reference Terminology: a cross-institutional content coverage study. Medinfo. 2004;11(Pt 1):477-81.

[14] McDonald CJ, Overhage JM, Tierney WM, Dexter PR, Zafar A, Schadow G, et al. The Regenstrief Medical Record System: a quarter century experience. Int J Med Inform. 1999 Jun;54(3):225-53.

[15] McDonald CJ. The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries.
Health Aff (Millwood). 2005 Sep-Oct;24(5):1214-20.

[16] Carter JS, Brown SH, Erlbaum MS, Gregg W, Elkin PL, Speroff T, Tuttle MS. Initializing the VA Medication Reference Terminology Using UMLS Metathesaurus Co-Occurrences. Proc AMIA Symp. 2002:116-20.