RxTerms – a drug interface terminology derived from RxNorm

Kin Wah Fung MD MS MA, Clement McDonald MD, Bruce E Bray MD
National Library of Medicine, Bethesda, Maryland
kwfung@nlm.nih.gov|clemmcdonald@nlm.nih.gov|brayb@mail.nih.gov

Abstract
A good interface terminology is an essential component of any Computerized Provider Order Entry system. RxTerms is a drug interface terminology derived from RxNorm. By reorganizing the drug information into two dimensions as prescribers do when writing prescriptions and by eliminating drug names that are less likely to be needed in a prescribing environment, RxTerms helps the user to efficiently enter complete prescription orders. Preliminary evaluation of RxTerms using a list of most commonly prescribed drugs showed that its coverage was very good (99% for both generic and branded drug names). There was significant efficiency gain compared to using the unprocessed RxNorm names. RxTerms fills the gap for a free, up-to-date drug interface terminology that is directly linked to RxNorm, a U.S designated drug terminology standard.

INTRODUCTION

Computerized Provider Order Entry (CPOE) has been shown to be effective in preventing medication errors.1 Despite the widespread attention drawn to medication errors by the reports from the Institute of Medicine, and the effort by groups like the Leapfrog Group pushing for CPOE systems, care providers have been slow to accept them.2 One key to user acceptance is good interface design,3 and an important element of a well-designed interface is good interface terminology.4 In the case of medications ordering, a good drug interface terminology allows the user to complete prescriptions in a natural and expedient manner. Another important function of a good interface terminology is to link the information entered by users to standard coding systems to facilitate unambiguous transmission of information and data interoperability.

Most CPOE systems employ drug interface terminologies that are either home-grown or purchased from drug knowledgebase vendors. The home-grown option entails a significant investment of effort in the creation and maintenance of the terminology. The proprietary option requires extra and recurrent expenditure. In both cases, data interoperability may be an issue if the drug information is not linked to a standard terminology.

RxTerms was created as a drug interface terminology for a CMS (Centers for Medicare and Medicaid Services) assessment tool to be used in the post-acute care environment.5 It will also be used in the NLM (National Library of Medicine) Personal Health Record project. We think RxTerms can fill the gap for a publicly available and up-to-date drug interface terminology that is directly linked to RxNorm, a U.S designated drug terminology standard.

RXNORM

RxNorm is a standardized nomenclature for clinical drugs created by NLM.6 A clinical drug is a pharmaceutical product given to (or taken by) a patient with a therapeutic or diagnostic intent. RxNorm serves as the bridge between different naming conventions used in disparate drug information systems. By creating a standardized set of drug names and linking them to equivalent names from various sources, RxNorm allows systems using different drug nomenclatures to share data efficiently at the appropriate level of abstraction.

RxNorm is organized around concepts with normalized names for clinical drugs. The two most important classes of concepts for prescription ordering are Semantic Clinical Drug (SCD) and Semantic Branded Drug (SBD) which are assigned RxNorm unique identifiers (RXCUI). The name of these concepts contains ingredient(s), strength, route and dose form information. The brand name is appended to the name of a SBD. For example:

198440 Acetaminophen 500 MG Oral Tablet (SCD)
209459 Acetaminophen 500 MG Oral Tablet [Tylenol] (SBD)

Within RxNorm, generic and branded normalized forms are linked to each other and to the names of their individual components by a well-defined set of named relationships. RxNorm is available as one of the source terminologies in the UMLS (Unified Medical Language System) and a monthly standalone release. In the U.S., RxNorm is the designated Consolidated Health Informatics (CHI) vocabulary for clinical drugs.7 RxTerms is a subset of RxNorm which carries the RxNorm codes and is tailored for prescription order entry in the U.S.
PRINCIPLES
The creation of RxTerms follows the following principles:

1. Unambiguity – the use of the interface terminology should lead to the unambiguous identification of the clinical drug being prescribed. In practice, this means it should be able to produce the RXCUI of either the SCD or SBD from what the user enters.

2. Naturalness – the names it uses and the way they are organized should conform to the usual prescribers’ vocabulary and prescription writing habits.

3. Efficiency – the user should be able to identify the drugs they are prescribing quickly with a small number of mouse clicks and keystrokes.

4. Coverage – there should be adequate coverage of drugs used in the prescribing environment. At present, we have focused on currently available drugs in the U.S.

METHODS
A. Creation of RxTerms
Conceptually, the creation of RxTerms involves three steps:

1. Data extraction from RxNorm
Based on the latest release of RxNorm (the January 2008 release was used in this paper), we extracted a subset from the table RXNCONSO including all SCD and SBD that were not obsolete (SUPPRESS = ‘N’). Since SCD and SBD belong to RxNorm and are non-proprietary, the resulting subset is free of additional licensing requirements. An additional process ensured that the SCD related to any of the SBD in the subset was also included. Supplementary information was then gathered by following the relationships in the RXNREL table. This included: the ingredient(s) and dose form of each SCD/SBD and the corresponding RXCUI of the SCD for each SBD. A record in the subset data file included the following:

RXCUI: 205766
GENERIC_RXCUI: 562251
TTY: SBD
FULL_NAME: Amoxicillin 250 MG / Clavulanate 125 MG Oral Tablet [Augmentin]
INGREDIENT: Amoxicillin; Clavulanate
DOSE_FORM: Oral Tablet

FULL_NAME is the name of the clinical drug in RxNorm. GENERIC_RXCUI is the RXCUI of the SCD corresponding to each SBD. This field is empty for SCDs.

2. Parsing and reorganization of drug names
Initially we experimented by using the RxNorm full names as the interface terminology. The result was unsatisfactory because the names were long and difficult to display. Often there were too many matches and long lists were returned. Cognitively it was difficult to find the correct choice because of the long names and the large number of choices. This prompted us to think that some segmentation and recombination of the prescribing information was necessary.

In the clinical setting, a complete written prescription usually contains 6 components:

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Route</th>
<th>Strength</th>
<th>Form</th>
<th>Frequency</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amoxicillin (oral) 500 mg tabs q.i.d. x 1 wk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The oral route is sometimes omitted, but in a CPOE environment we think it is better to make it explicit to avoid ambiguity. The first 4 components were provided by RxNorm. The full name from RxNorm was parsed as follows to extract the components:

Ingredient – For a SCD, this was the name of the active ingredient(s) e.g. Amoxicillin. In case of multi-ingredient generic drugs the ingredients were separated by a slash e.g. Amoxicillin/Clavulanate. For a SBD, the brand name alone was used e.g. AUGMENTIN. In RxTerms we converted all brand names to upper case to distinguish them from generic drugs in a pick list. This is a standard convention in many drug formularies.

Route and form – Route was derived from DOSE_FORM in RxNorm. This was sometimes explicitly stated (e.g. Rectal Suppository) and sometimes implied (e.g. Extended Release Tablet – implying an oral route). Note that the RxNorm dose form only reflects the presumed route that the drug is formulated to be administered. In unusual situations drugs may be used in a route different from the presumed route (e.g. oral use of injectable vancomycin to treat pseudo-membranous colitis). These exceptions need to be managed at the application level. A table was used to translate DOSE_FORM into route and form. To save on screen space utilization, we made use of some common abbreviations (e.g. XR Tabs, EC Caps, and Susp for extended release tablet, enteric coated capsule, and suspension).

Strength – Strength was obtained by parsing the RxNorm full name. In case of multi-ingredient drugs...
the strengths were separated by a dash (e.g. 250-125 MG for the AUGMENTIN example)

Our next question was how to present the information to users. We think that by breaking up the full prescription into segments the pick lists will be shorter and this will reduce the cognitive load. Generally, there is a trade-off between the number of steps the user has to go through and the number of items the user has to choose from in each step. We think a good compromise is to organize the information into 2 segments: ‘ingredient + route’ (e.g. Amoxicillin (Oral-pill)) and ‘strength + form’ (500 MG Tabs). Note that we have further divided the route ‘Oral’ into 3 sub-categories (Oral-pill, Oral-liquid and Oral-other) because of the large number of oral drugs.

We made two other modifications to the drug names. The first was related to generic names of insulin. There were 41 SCDs for insulins. The list was difficult to navigate because it was long and some names were not familiar to prescribers. We first removed insulins from animal sources (e.g. beef or pork insulins) which are no longer available in the U.S. For the others, we added more uniform and familiar names (e.g. Insulin, human Mixed 50/50 (Injectable) instead of NPH Insulin, Human 50 UNT/ML / Regular Insulin, Human 50 UNT/ML Injectable Suspension). This resulted in a concise list of 15 names:

- Insulin analog, Aspart (Injectable)
- Insulin analog, Aspart Mixed 70/30 (Injectable)
- Insulin analog, Detemir (Injectable)
- Insulin analog, Glargine (Injectable)
- Insulin analog, Glulisine (Injectable)
- Insulin analog, Lispro (Injectable)
- Insulin analog, Lispro Mixed 50/50 (Injectable)
- Insulin analog, Lispro Mixed 75/25 (Injectable)
- Insulin, human L (Injectable)
- Insulin, human Mixed 50/50 (Injectable)
- Insulin, human Mixed 70/30 (Injectable)
- Insulin, human N (Injectable)
- Insulin, human R (Injectable)
- Insulin, human U 500 (Injectable)
- Insulin, human U (Injectable)

The second modification of drug names was in accordance with the FDA recommendation to use “Tall Man” lettering to avoid medication errors in 16 look-alike pairs of drug names (e.g. ChlorproMAZINE and ChlorproPAMIDE).8

### 3. Pruning

We employed several strategies to prune RxTerms of names that were less likely to be useful in a prescribing environment. Firstly, we excluded RxNorm concepts which had no links to NDC code(s) in the RXNSAT table (we included all sources for NDC codes in RxNorm, not only those with SAB=RXNORM). The assumption was that all available U.S. drugs should be linked to one or more NDC codes.

Some drug names were particularly problematic because they returned large numbers of hits. These were mostly over-the-counter pain and cold medications. For example, ‘acetaminophen’ returned a list of 405 SCDs while chlorpheniramine returned 473. We believe that most of the multi-ingredient pain and cold medicines will be referred to by their brand names (e.g. VICODIN) rather than by their full generic names (e.g. acetaminophen/hydrocodone). Therefore, for this class of drugs, if there were multiple ingredients and if there was already a branded equivalent, the generic name was suppressed. This dramatically reduced the number of hits for drugs like acetaminophen or aspirin.

Another problematic area was that some common generic ingredient names were modified into brand names, e.g. ST. JOSEPH ASPIRIN, ASPIRIN-FREE, and ACETAMINOPHEN CHILD. Not only were these names rarely used in prescribing (they were mostly over-the-counter drugs), they could actually be confusing (e.g. ASPIRIN-FREE was the name for an acetaminophen rather than aspirin preparation). Because of this, all brand names containing ‘aspirin’ or ‘acetaminophen’ were suppressed.

Two more groups of drug names were suppressed because they were not expected to be useful in prescribing. They were the allergenic extracts (e.g. Hymenoptera allergic extract) and the SBDs that contained the phrase ‘brand of’ in the brand name component (e.g. [Nasal Decongestant brand of pseudoephedrine]).

### B. Evaluation of RxTerms

To evaluate RxTerms we used a list derived from the RxList – the Top 200 Prescriptions for 2003 by Number of U.S. Prescriptions Dispensed: Generic Name. The derived list contained both branded and generic names.9 The coverage of RxTerms for these drug names was evaluated.

To evaluate the efficiency of RxTerms we simulated a CPOE environment in which the user typed in the drug name which was matched against the drug display names (ingredient + route) in RxTerms. We assumed that ‘auto-completion’, a popular feature in many form-filling applications, was used. This meant that stepwise matching occurred as the user typed in each additional letter. For example, when the user typed in ‘dil’ the returned list contained ‘dilantin,
dilaudid, dilocaine, dilor…’ but as she typed in ‘dila’ the list shortened to ‘dilantin, dilaudid, dilacor…’; and on typing ‘dilan’ only ‘dilantin’ remained. We only used the generic and brand names that had exact matches in RxTerms for this part of the evaluation because our evaluation tool could not identify semantic matches. For each drug, we recorded the size of the lists as each additional letter was typed. We also noted the minimal number of keystrokes needed to return a list of less than 15 items. This was somewhat arbitrary but 15 items could usually be displayed without scroll-bars. For comparison, we repeated the matching against the full names of all non-obsolete SCD and SBD in RxNorm.

RESULTS

There were 14,609 branded (SBD) and 18,604 generic (SCD) clinical drug names in RxNorm that were not obsolete. As a result of the pruning processes these were reduced to 13,933 and 9,333 (branded and generic) respectively in RxTerms. The NDC code requirement accounted for the bulk of this reduction. 90% of excluded branded names and 75% of excluded generic names were removed because they were not connected to NDC codes.

Of the 165 generic drug names on the most-prescribed drugs list, all but one could be found in RxTerms. Among the ones that were found, 33 generic names did not have exact matches in RxTerms either because of minor naming variations (e.g. Folate vs. Folic acid) or different ordering of ingredients (e.g. promethazine / codeine vs. codeine / promethazine). The only missing generic drug was Divalproex. Divalproex sodium is the USAN (United States Adopted Name) equivalent of the INN (International Nonproprietary Name) name Valproate semisodium, and Valproate was present in RxTerms.

Among the 222 branded drug names on the most-prescribed drugs list, 206 were found in RxTerms. Among them, 7 were not exact lexical matches because of minor naming differences (e.g. TRI-LEVLEN vs. TRI LEVLEN). Among the 16 that were not found in RxTerms, 13 were oral contraceptive pill packs (see Discussion). Of the 3 missing brand names that were not oral contraceptives, 2 were present in RxNorm but labeled as obsolete (Duragesic and Esidrix), while Actonel was excluded by one of the pruning steps described above because the RxNorm name was [Actonel brand of risedronate sodium]. These misses might represent data anomalies in RxNorm.

In summary, the coverage of RxTerms was 99% (164/165) for generic drug names and 99% (206/209) for branded drug names that were not oral contraceptives.

For the efficiency evaluation, Figure 1 shows the relation of the median pick list size for all drugs to the number of letters typed in (keystroke). List sizes were generally big when less than 4 letters were typed (e.g. median list size for generic drugs at 3 keystrokes: RxNorm 361, RxTerms 82) so they were not included in the graph. The median list size for generic drugs was considerably smaller at every keystroke for RxTerms compared to RxNorm. The biggest difference was seen when 4 and 5 letters were typed (4 letters: RxNorm 48, RxTerms 9; 5 letters: RxNorm 28, RxTerms 4). For branded drugs, the difference was still seen but less pronounced than for generic drugs.

Table 1 shows the minimal number of keystrokes needed to return a list of less than 15 items. There were some drugs for which even when the full name was entered the returned list was still 15 or above (‘never’). Overall, using RxTerms for generic drugs, one could get a list of less than 15 within 5 keystrokes for 88% of drugs. This was considerably better than the unprocessed strings in RxNorm (27%). The percentage of generic drugs that failed to get below 15 was much lower for RxTerms (2%) compared to RxNorm (66%). The three drugs that failed to get below 15 were niacin, hydrochlorothiazide and potassium chloride which occurred as ingredients in many multi-ingredient drugs. For branded drugs, the efficiency gain of using RxTerms was smaller. Using RxTerms, one could get to below 15 in 5 keystrokes for 92% of branded drugs (vs. 80% for RxNorm).
There were no branded drugs that failed to get below 15 for RxTerms (vs. 3% for RxNorm).

Table 1: Efficiency evaluation for commonly prescribed drugs (KS: keystroke, RxT: RxTerms, RxN: RxNorm, G: generic drugs B: branded drugs)

<table>
<thead>
<tr>
<th>KS</th>
<th>RxT-G</th>
<th>RxN-G</th>
<th>RxT-B</th>
<th>RxN-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 3</td>
<td>15（11%）</td>
<td>2（2%）</td>
<td>38（19%）</td>
<td>22（11%）</td>
</tr>
<tr>
<td>&lt; 4</td>
<td>86（66%）</td>
<td>20（15%）</td>
<td>133（67%）</td>
<td>89（45%）</td>
</tr>
<tr>
<td>&lt; 5</td>
<td>115（88%）</td>
<td>36（27%）</td>
<td>184（92%）</td>
<td>160（80%）</td>
</tr>
<tr>
<td>&gt; 6</td>
<td>128（98%）</td>
<td>44（34%）</td>
<td>199（100%）</td>
<td>194（97%）</td>
</tr>
<tr>
<td>never</td>
<td>3（2%）</td>
<td>87（66%）</td>
<td>0（0%）</td>
<td>5（3%）</td>
</tr>
<tr>
<td>Total</td>
<td>131</td>
<td>131</td>
<td>199</td>
<td>199</td>
</tr>
</tbody>
</table>

DISCUSSION

The purpose of RxTerms is not to replace RxNorm but to complement it. RxTerms is created for the specific use case of a prescribing application in the U.S. and does not cover other uses of RxNorm e.g. to identify drugs in an archival database. Efficiency is a high priority in RxTerms and aggressive pruning is employed, which may result in the removal of content useful in other circumstances.

Our preliminary evaluation showed that there was good coverage of both generic and branded drug names. The problem of missing a match because of minor naming differences (e.g. folate and folic acid) and variation in the order of ingredients can be solved by adding synonyms and indexing based on individual ingredients. The missing of pill packs (e.g. oral contraceptives) was a different problem. Prior to 2008, RxNorm did not have a mechanism to represent products that comprise multiple individually prescribable components (e.g. an oral contraceptive pill pack containing 3 different strengths of ethinyl estradiol/norethindrone and iron tablets). To address this problem, two new classes were introduced in RxNorm to represent these products (called Drug Delivery Devices in RxNorm parlance): Generic Pack (GPCK) and Branded Pack (BPCK). RxTerms has not yet incorporated these new classes.

In our study, the length of the returned list was used as a proxy for data entry efficiency. Whether this would actually result in a shortening of prescription time requires further study. The efficiency gain by using RxTerms rather than RxNorm was considerable. This was a result of a combination of pruning and information segmentation. One caveat of pruning by NDC codes is that the list of NDC codes in RxNorm may not be complete and some drugs might be left out unintentionally. Further evaluation with a larger list of drugs will be necessary.

The real test of RxTerms will come with its use in real systems. Even though RxTerms was originally created for use by prescribers, we believe that consumers will also be able to use it to record their medications. This assumption will be tested in our soon to be piloted Personal Health Record project. In the coming few months we will enhance RxTerms by incorporating the new pill pack classes in RxNorm and adding common abbreviations and synonyms (e.g. ASA for aspirin, INH for isoniazid). RxTerms will be made available for beta testing. Interested testers can contact the lead author (kwf). RxTerms is still work in progress and we definitely welcome all comments and suggestions.

Acknowledgements

The authors would like to thank Dr Stuart Nelson for his insightful comments and suggestions.

References