Published in final edited form as:

Am J Health Syst Pharm. 2009 October 1; 66(19): 1743–1753. doi:10.2146/ajhp080221.

Using National Drug Codes and Drug Knowledge Bases to Organize Prescription Records from Multiple Sources

Linus Simonaitis, MD¹ and Clement J McDonald, MD²

¹ Assistant Professor of Medicine, Indiana University School of Medicine, Regenstrief Institute, Inc. Indianapolis, IN

² Director, Lister Hill Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

Abstract

Purpose—Pharmacy systems contain electronic prescription information needed for clinical care, decision support, performance measurements and research. The master files of most pharmacy systems include National Drug Codes (NDCs) as well as the local codes they use within their systems to identify the products they dispense. We sought to assess how well one could map the products dispensed by many pharmacies to clinically oriented codes via the mapping tables provided by Drug Knowledge Base (DKB) producers.

Methods—We obtained a large sample of prescription records from seven different sources. These records either carried a national product code or a local code that could be translated into a national product code via their formulary master. We obtained mapping tables from five DKBs. We measured the degree to which the DKB mapping tables covered the national product codes carried in, or associated with, our sample of prescription records.

Results—Considering the total prescription volume, DKBs covered 93.0% to 99.8% of the product codes (15 comparisons) from three outpatient, and 77.4% to 97.0% (20 comparisons) from four inpatient, sources. Among the inpatient sources, invented codes explained much – from 36% to 94% (3 of 4 sources) – of the non coverage. Outpatient pharmacy sources invented codes rarely – in 0.11% to 0.21% of their total prescription volume, and inpatient sources, more commonly – in 1.7% to 7.4% of their prescription volume. The distribution of prescribed products is highly skewed: from 1.4% to 4.4% of codes account for 50% of the message volume; from 10.7% to 34.5% of codes account for 90% of the volume.

Conclusion—DKBs cover the product codes used by outpatient sources sufficiently well to permit automatic mapping. Changes in policies and standards could increase coverage of product codes used by inpatient sources.

INTRODUCTION

Caregivers need an accurate list of their patients' medications to avoid prescribing errors and provide optimal care.¹ Over time, a patient's medications will be prescribed by many different providers and dispensed by many different pharmacies; and so that patient's medication records will tend to be scattered. Consequently, clinicians today must gather a medication history, including both active and inactive medications, directly from their patients. Hospitals must do the same as part of the medication reconciliation required by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO).² These processes are both time-intensive and error-prone, and beg to be automated.³

Almost all inpatient and outpatient pharmacies use computers to process and fill prescriptions/drug orders. So – in theory – care providers should be able to obtain a complete record of all of a patient's medications by pulling the prescription records from all of the pharmacies that served their patient. Indeed, this idea inspired a consortium of pharmacy benefit managers (PBMs) to create RxHub, in order to coordinate the aggregation of pharmacy dispensing records into a unified medication history.⁴ Extended to all medical information, the same idea motivated the creation of Regional Health Information Organizations (RHIOs) which aggregate clinical information of many kinds from regional sources.⁵

Pharmacy systems, from both the inpatient and outpatient environments, employ levels of standardization that could enable the delivery of prescription records to care providers in a computer understandable manner. Most community pharmacies use the National Council of Prescription Drug Programs (NCPDP) Telecommunication Standard,⁶ and most hospital pharmacies use Health Level Seven (HL7) pharmacy order messages.⁷ Based on our Indiana experience we believe that most pharmacy systems either do, or could, include a universal product identifier, i.e., the National Drug Code (NDC)⁸, in these prescription messages because they carry this identifier in their master formulary for most of the medications they dispense.

However, this universal product identifier can not directly enable decision support or performance measurements and can not be used to organize medication profiles and flowcharts. NDCs were designed for inventory management and reimbursement. Each product labeler assigns its own NDC for every product that it markets. So every distinct combination of brand name, dosage form, strength, and package size gets many different NDCs. As a result products that clinicians might consider as a single medication are represented by hoards of different NDC codes. Amoxicillin 500 mg capsule, for example, has at least 227 distinct NDCs, and there is nothing intrinsic to these codes that ties them together, as you can see from two randomly selected NDC codes for Amoxicillin 500 mg capsule: 00003-0109-55 and 52959-0020-60. As argued by others, clinically equivalent NDCs should be mapped to a higher level code that identifies the clinically relevant concept- the generic drug, dosage form and strength.^{9,10,11} Following common usage, we call this the “clinical drug” code.

The commercial Drug Knowledge Base (DKB) vendors provide exactly the clinical drug codes needed for clinical use. They also provide tables for mapping NDC codes to their clinical drug codes. These commercial DKBs have been adopted by pharmacies and clinical care systems to assist the prescribing and dispensing process. The National Library of Medicine (NLM) also provides a public use DKB – RxNorm – with a table for mapping NDC codes to their “Semantic Clinical Drug” (SCD) code.¹² So, in theory, with what is now in place, hospitals and office practices could automatically capture these prescription records from all relevant pharmacy systems, inpatient and outpatient, map the NDC codes to the clinical drug code from one DKB and file the prescription records from all prescription sources under the appropriate clinical drug within their medical record system.

However, the success of such an effort will depend upon the degree to which the drug identifiers used in pharmacy messages are also carried in the DKB mapping tables. To the degree that the DKBs fail to include NDCs in common use, or that pharmacy systems use locally invented product codes in these messages, this automated mapping process fails.

In order to assess the degree to which DKB mapping tables *can* translate the codes that come in pharmacy messages, we obtained mapping tables from five DKBs. We also obtained a large sample of production NCPDP and HL7 pharmacy messages, and their associated master formulary files. We then assessed the degree to which the codes in the prescription records and formularies could be found in the DKB mapping tables, and if not, why not. Here we report the results of that assessment.

METHODS

Background on the NDC

An NDC is a 10 digit code, consisting of 3 parts, delimited by dashes: (1) the Labeler segment assigned by the FDA to the distributor, manufacturer, or repackager of the product; (2) the Product segment, which identifies a specific drug product (e.g., Zocor® 20 mg tablets^a); (3) a Package segment that is supposed to distinguish different package sizes produced by one labeler (e.g., bottle of 100 tablets). It is important to note that the FDA does not control assignment of the entire code, but only its first segment. This assignment process is akin to the assignment of Internet addresses: a root code is assigned to an organization, and the organization assigns more specific codes by adding digits to that root.

The NDC was introduced in 1972 as a 10-character code with a “4-4-2” configuration to identify the Labeler, Product and Package segment, respectively. Later, the FDA expanded the Labeler segment to 5 digits, with 2 configurations (“5-4-1” and “5-3-2”). All of these configurations used dashes as delimiters to distinguish the three segments. Today, most users convert the historic codes to an 11-digit format. In this format the first 5 digits represent the Labeler segment, the next 4 digits, the Product and the last 2 digits the Package. To convert a historic NDC to this newer 5-4-2 configuration, one must add a leading zero either to a 3-digit Product segment, or to a 1-digit Package segment. The 11-digit format includes no dashes. The 11-digit format is the only one permitted in NCPDP messages, and HIPAA regulations mandate this 11-digit format for all HIPAA transactions.¹³

The primary focus and starting point for the processes described in this paper are the industry accepted product codes found in NCPDP messages, HL7 messages, pharmacy master files and DKB mapping tables. We are interested in these industry codes because, today, they are used widely by pharmacies to identify prescribed products and thus provide the key to aggregating prescription records from many sources. There is no accepted name for the universe of codes we are describing. Many refer to them casually as NDC codes. However, this universe of codes is actually a collection of: (a) registered NDC codes – i.e., codes for medications (including therapeutic biological products) that have been officially registered in the FDA’s National Drug Code file; (b) semi-official NDC codes – i.e., codes for medications that have been properly assigned according to FDA rules, but not yet registered in the FDA’s database; (c) device/supply codes.

The two types of codes used for identifying devices and supplies are the National Health Related Item Code (HRI)¹⁴ system and the Universal Product Code (UPC)¹⁵ system. The HRI system was developed in the 1970s by the FDA’s Center for Devices and Radiological Health. The FDA set aside a block of 4-digit Labeler Codes for the HRI system, designed not to collide with the Labeler Codes for the NDC system. Although HRI codes consist of 2 segments and 10 digits, they can be converted to a string of 11 digits, in accordance to NCPDP standards, in a way analogous to that for NDC codes.

UPC codes derive from an even more complex system, beyond the authority of the FDA. These codes are coordinated by the global standards development organization GS1, and its U.S. member organization GS1 US (formerly known as the Uniform Code Council). Different versions of these codes exist, with different lengths. However, the most common version of UPC in this country appears to be the 10-digit pattern found on bar codes (typically accompanied by an 11th digit on the left to indicate product type, and a 12th modulo check digit on the right). This 10-digit UPC is the one which is often converted to an 11-digit string by adding an additional zero between its two 5-digit segments.

^aZocor® (Simvastatin) is a registered trademark of Merck & Co., Inc., Whitehouse Station, NJ.

In the DKBs which we examined, the UPC codes and HRI codes are stored in the same database columns as the NDC codes. However, the commercial knowledge bases do provide some clues to the different origin of these codes, such as an additional indicator field.

All of these categories of codes (registered NDC codes, semi-official NDC codes, HRI codes and UPC codes) are delivered in the same slot of HL7 and NCPDP messages, and they all appear in the same slot of Pharmacy master files. Operationally, this universe is defined by the DKB vendors, who actively gather these codes and information about the products these codes represent. For convenience, in this paper we will call them “NDCs”, using quotes to remind the reader that we are referring to the above defined universe of NDC-like codes. We also include industry-assigned supply codes in this rubric for the sake of simplicity. When speaking of NDCs officially registered with the FDA, we will not use quotes and will always precede NDC with the words “official”.

Another population of prescribed codes flow in messages in the same place as the “NDCs”. We will call them “invented codes” because they are invented locally by pharmacies and other organizations without FDA-assigned labeler codes and are not part of any national coding system. Invented codes may appear in the “NDC” slots of local formularies and electronic prescriptions messages where “NDCs” are usually found. Such codes are sometimes – but not always – easily distinguishable from “NDCs” by their format.

The FDA does provide an enumeration of the official NDCs in its database. But the FDA’s enumeration does not include many of the “NDCs” one finds in electronic messages. Importantly, each DKB provider has its own enumeration of “NDCs” embodied in its mapping table. But each of these is also incomplete. Indeed, no complete enumeration of assigned “NDCs” exists. Because of this lack we had to develop the operational definition of invented codes described below.

Sources of Prescription Data

a) Inpatient Samples—We obtained sample inpatient data from the five Indianapolis hospital systems participating in the Indiana Network for Patient Care (INPC).¹⁶ Each of the five provided us with their inpatient master formulary file (which defines orderable medications). Four also provided HL7 messages from their inpatient pharmacy system. One of these provided us with HL7 “Detailed Financial Transaction” (DFT) messages which used “NDCs” to identify the prescribed medication. The other three hospital systems provided us with HL7 “Pharmacy/Treatment Encoded Order” (RDE) messages – one using “NDCs”, and the other two using local service codes to identify the prescribed entity. (In these two cases, we mapped these service codes to the “NDCs” in the respective hospitals’ formulary master file.) Two of the hospitals provided a one week sample, and two provided a one month sample of pharmacy prescription messages.

b) Outpatient Samples—Our outpatient prescription samples came from three sources. (1) We obtained a 6-month collection of NCPDP messages from a large stand-alone outpatient pharmacy system associated with one of the Indianapolis hospitals. (2) We also obtained a 6-month collection of HL7 “Pharmacy/Treatment Dispense” (RDS) outpatient prescription records delivered to the INPC by RxHub. RxHub provides the INPC with medication histories for patients who visit any Emergency Department in Indianapolis. (3) RxHub also provided us with 24 different outpatient formularies for health insurance companies from across the country. We aggregated the “NDCs” from these 24 formularies into one database and treated this aggregate as a single source of “NDCs”.

c) Archival Outpatient Sample—All of the above sources represent real content from production systems. With the data listed above, we could assess the problems of mapping the

identifiers in *current* prescription records to a clinical drug code. We also assessed the problems which an organization would face, if it wanted to extract long-term medication histories from archival sources for care or research. We obtained a list of all of the “NDCs” contained in a 12-year period in the Indiana State Medicaid prescription database for this purpose.

Of these 13 samples four were obtained in the second half of 2005 and nine in the first half of 2006. The smallest of these sample contained more than 40,000 prescriptions, and the largest 60 million.

DKB Mapping Tables

We obtained mapping tables from each of four commercial DKBs: (1) First DataBank National Drug Data File Plus™ (First DataBank, Inc., San Bruno, CA); (2) Medi-Span Master Drug Database® (Wolters Kluwer Health, Inc., Conshohocken, PA); (3) Multum® Lexicon (Cerner Multum, Inc., Denver, CO); and (4) Thomson Micromedex Red Book™ (The Thomson Corporation, Greenwood Village, CO). These four DKBs are used widely in pharmacy information systems, and are listed as knowledge base options in the NCPDP Database Indicator field.¹⁷ Each of these vendors kindly provided us with their core database content at no cost for this study. Each DKB supplier included their mappings from well over 100,000 current “NDCs” to their own clinical drug codes. We also obtained the RxNorm knowledge base, which the NLM makes freely available to the public.

Unlike these five DKBs, the FDA National Drug Code directory currently does not include a clinical drug coding system, and is not directly comparable to the other five. We included the National Drug Code directory (obtained from the FDA website⁸) in some of the analyses, in order to assess its coverage of “NDCs” in common use. (See Table 1 for a listing of the five DKBs and the FDA directory.)

Prescribed Supplies

Pharmacies fill prescriptions for supplies – such as insulin syringes, glucose test strips and gauze sponges – as well as medications. In pharmacy messages, the codes for these supplies are treated in the same way as the codes for medications. In NCPDP messages, these prescribed supplies are identified by a UPC or HRI code. The UPC or HRI is recorded in the same slot used to record “NDCs”. An additional field in the NCPDP message names the coding system used, and a field is available in HL7 to do the same. The FDA has reserved a separate block of labeler segments for HRI codes; so these will not collide with NDC codes. Neither the official NDC table nor RxNorm carries such supply codes.

Within hospital-delivered HL7 messages, the prescribed product identifier can be an “NDC” or a local “service code”. Each hospital pharmacy has a master formulary, which includes a record for each service code that usually also carries the associated “NDC”. Thus in the case that pharmacies send only local service codes in their messages, these codes can usually be translated into NDCs, HRIs or UPCs – as the case may be – via their formulary table.

Preprocessing of Source Prescriptions

In the case of NCPDP messages, we extracted the “NDC” from the “Item Number” field of the “Drug Segment”. In the case of HL7 RDE messages and RDS messages, we extracted the “NDC” from the first component of Field 2 in the RXE segment and RXD segment, respectively. In the case of the HL7 DFT messages, we extracted the “NDC” from the first component of Field 9 of the FT1 segment. We converted all “NDCs” to the 11-digit format before doing any matching.

All but one of the prescription sources supplied a large sample of prescription messages and their master formulary table. The one inpatient service that provided us only with a master formulary table does not currently send HL7 messages. We used the same preprocessing to analyze the contents of messages and master formulary files.

We examined a sample of messages from each source by hand. Doing this, we found that one of our message sources included a null value in the HL7 field that carries the prescribed entity ID. The messages with null IDs carried clinical data (mostly values of creatinine clearances as free text), a misuse of the HL7 intended to carry information about prescriptions. We excluded these messages from all of our calculations because they were easy to distinguish and did not represent prescriptions. This experience emphasizes the need for some investment in manual review of messages from each source at the start in order to discover, and adjust for, peculiarities and deviations from the relevant message standards.

Locally Invented Codes

The institutions in the INPC sometimes invent their own local prescribed entity codes and use them in the same fields as “NDCs”. We could spot many of these at initial inspection of the messages because their format was incompatible with the format of an “NDC”. Invented codes will not be present in any DKBs and will not be mappable by automatic means. We developed an operational definition for invented codes to quantify their prevalence: i.e., a code is “invented” if it:

- a. is longer than 11 or shorter than 10 digits;
- b. contains alphabetic characters;
- c. begins with a sequence of five identical digits (e.g., 11111, 22222, etc.);
- d. begins with the digits “991”. (We encountered this case at only one hospital.)

The first two rules are based on the format definitions of “NDC” codes and the last two are based on an empiric review of the source databases. We tested our working definition by searching through the four commercial DKBs, which contain a combined total of 249,098 distinct “NDCs”. We found only two records that would have been called invented codes by any of the above criteria.

Counting and Calculations

For non-invented “NDCs”, we counted all unique codes in a given message stream. For invented codes, we counted unique combinations of prescribed entity name *and* code, because a review of the product names associated with these codes showed that the same invented code could be used to identify many different products. To get the total number of unique codes from a given source, we added the number of unique non-invented codes to the number of unique invented code-name combinations.

We had originally assumed that the first 9 digits of the “NDC” would be the appropriate level for comparison, because the last 2 digits of the “NDC” are supposed to represent only the package size (e.g., bottle of 100 capsules), which is irrelevant for most clinical purposes. We assessed the degree to which a given DKB covered the “NDCs” from each of our sources by using the leading 9 digits of the “NDC”, and again by using all 11 digits. We report only the 11-digit comparison, because looking within individual databases we found several hundred pairs of NDCs which differed only by the last 2 digits but which represented quite different drugs. One such pair, selected at random, is 00686-0360-10 (Ipecac Syrup) and 00686-0360-67 (Digoxin Elixir); another such pair is 55289-0033-28 (Ampicillin Capsule) and 55289-0033-97 (Prochlorperazine Tablet). Apparently, labelers or manufacturers sometimes employ the last

2 digits for distinguishing ingredients rather than package size. Using 9 digits did not increase the coverage of any DKB by more than 1 or 2 percentage points.

De-identification and Institutional Review Board

All this work was done with non-patient data from master formularies or de-identified portions of prescription messages. We obtained approval from the Indiana University Institutional Review Board for this study.

RESULTS

Codes in the DKB Tables

The number of unique “NDCs” contained in the mapping tables from DKB providers (the 4 commercial DKBs and NLM) ranged from 113,221 to 232,111. The numbers of distinct clinical drug codes in these same tables ranged from 8,082 to 14,405. (See Table 1.) Appendix A provides the details of how we counted clinical drug codes in these DKBs. The differences in number of distinct drug codes are due to differences among the rules for distinguishing supplies from medications across DKBs, the granularity at which some drugs, e.g., multivitamins, are represented, and the inclusion of special content, e.g., allergy shots, in some DKBs and not in others. Therefore the numbers are not directly comparable. Difference in the size of these numbers did not predict success in covering the product codes we obtained from pharmacy sources.

Distribution of product codes among all of the messages

The frequency distribution of code use across the total message volume is highly skewed. Across all the seven message sources, a very small percent (from 1.4% to 4.4%) of the unique codes account for 50% of the total prescription volume, and from 10.7% to 34.5% account for 90% of the prescription volume. The thin tail of this skewed distribution includes large numbers of codes that occur just once among 40,000 to 66 million prescriptions. Across six of the seven message sources, “NDCs” that occurred just once in the prescription sets made up 9 to 18% of the unique codes, but only 0.01% to 0.47% of the total message volume (i.e., the prescription sets, ranging in size from 40,000 to 66 million prescriptions). For the seventh source, the single instance codes made up only 3% of the unique codes. The skewing of invented codes was even more severe than that of the non-invented codes.

DKB Coverage of Unique Product Codes Found in Messages and Formularies

For completeness sake we report the DKB coverage of unique product codes in formularies and prescription records across all sources and DKBs. (See Table 2.) Note that prescription records and formularies from the same institution are listed as separate data sources in the table.

The DKBs cover less than 95% of the *unique* product codes in 12 of the 13 data sets listed in Table 2. The coverage was much better for the 13th data set — the RxHub prescription messages. Two DKBs covered nearly 98%, and the other three DKBs covered at least 94%, of the unique codes in the RxHub messages. Overall DKBs coverage of unique codes from inpatient sources was not as high as that from outpatient sources. DKBs failed to cover from 8.1% to 23.7% of the unique codes in inpatient messages, and from 4.4% to 13.3% of those in inpatient formularies.

A DKB will not cover a product code for one of two reasons: either (1) the unrecognized code is one invented by a local institutional source and could not be known by the DKB, or (2) the unrecognized code is a valid “NDC”, that was not included in that DKB.

Invented codes comprised a small percentage (1.6% or less) of the unique codes from all but one *outpatient* source. (See Table 3.) One outpatient source invented a profligate 20.7% of its unique codes. Across the five inpatient formularies, invented codes comprised between 1.6% and 12.3% of the unique codes; and across the four inpatient message sources, they comprised between 7.5% to 22.9% of the unique codes. By definition DKBs will not carry invented codes in their mapping tables, so invented codes will always decrease the rate of coverage.

DKB Coverage of the Total Volume of Product Codes

Coverage of unique codes is a misleading assessment of DKB coverage because the frequency distribution of product codes in messages is highly skewed. Using unique codes as the denominator weights codes that identify one prescription in a million the same as those that identify 20,000 prescriptions in a million. It is more appropriate to base the assessment on total prescription volume, that is, by counting the number of prescriptions covered by a DKB, and dividing it by the total number of prescriptions from a given source.

In the outpatient setting, all DKBs covered at least 93.0% of the total prescription volume across all sources (See Table 4.) In the case of Hospital E's outpatient pharmacy, three DKBs covered 97.6% or more, and one, 99.6%, of its total prescription volume. As Table 5 shows, invented codes accounted for only 0.21% of their total prescription volume from this pharmacy (compared to 20.7% of the *unique codes* in this data set). In the case of the RxHub prescription messages every DKB covered 99% or more of the total prescription volume.

On the inpatient side, the coverage of the codes from any one source by any one DKB was also better when measured in terms of total prescription volume – from 77.4% to 97.0% (with a median of 93.7%) for the 20 inpatient source cells in Table 4. But the coverage did not reach the heights of the outpatient sources. The lower rate of coverage in the inpatient setting is due in part to the higher use of invented codes (from 1.3% to 7.4% of the total prescription volume) compared to the outpatient setting. For each data cell in Table 4, we divided the number of invented codes by the total number of non-covered codes. For three of the four inpatient sources, the invented codes explained between 36% and 94% of the DKBs non-coverage. For the fourth inpatient source, they accounted for a smaller, but still important, proportion of the non-coverage: 6% to 32%. Of course the remaining proportion of the non-coverage in a given comparison was due to gaps in “NDC” coverage within the DKB.

Some of the DKB vendors cover outpatient sources better, while others cover inpatient sources better. Interestingly, RxNorm provided the highest, or second-highest, coverage of most sources. The FDA's published “official” NDC Directory usually had the lowest coverage: a median of 17.4 percentage points below that of the highest. The DKB with the highest coverage rates for a given source organization was sometimes the DKB employed by that source organization. The DKB vendors get feedback from their customers about missing “NDCs” and add them to their database product. So, over time we would anticipate most vendors to provide good coverage of the “NDCs” that their customers encounter.

Why are Codes Invented?

To determine why organizations invent codes, we hand reviewed records containing the invented codes. We found that pharmacies invent codes to accommodate (1) locally compounded medications; (2) drugs used in randomized controlled trials; (3) special items (e.g., “pig skin”); and (4) pharmacy actions not associated with a dispensing event. Of course, NDCs do not cover any of these uses. Pharmacies also invent codes to dispense a product which has not yet been registered in their master formulary file, even if that product already carries a valid NDC assigned by the manufacturer.

At one institution, compounded dermatologics accounted for over 315 unique invented codes. At another, compounded mixes of intravenous fluids and medications accounted for 113. “TPN” (total parenteral nutrition) accounted for 4–8% of the invented codes at hospitals that used this code. One hospital invented 47 codes for as many clinical trials. Most invented a few codes for identifying pharmacy actions that were not associated with the dispensing of any product. Examples included: “remove Fentanyl patch”, “compounding fee”, “read PPD skin test” – the last one accounting for 8–14% of the invented codes at two hospitals. Although the number of distinct codes in this category was small, they tended to be frequently used. None of the product codes that we categorized as “invented” by our operational definition appeared in any of the DKB mapping tables, giving credence to our definition.

Coverage of Archival Sources of Prescription Information

To assess the DKB coverage of archival prescription records, we looked at one source of “old” prescription records: the Indiana Medicaid database, which contains records of 66 million dispensed prescriptions, dated from 1994 to 2006, and includes 41,727 unique “NDCs”. The standard releases of the DKBs covered from 74.5% to 80.9% of these unique “NDCs”, and from 97.3% to 98.7% of the total volume of the product codes in the Medicaid database.

The commercial vendors include only active and recently inactivated “NDCs”, to serve the requirements of their pharmacy customers. Most of them do, however, keep the inactive “NDCs” in their internal databases. We obtained a custom database from one of the DKBs that included *all* the active and inactive “NDCs” from their internal database. This more complete set of NDCs covered 95.0% of unique Medicaid “NDCs” and 99.8% of the Medicaid prescription volume.

Reuse of “NDCs”

Official NDCs can be re-used (i.e., re-assigned to a completely different product) five years after the supplier has reported their inactivation to the FDA.¹⁸ Some DKBs keep track of the discontinuation, re-activation, and re-use of these codes. Based on the content of one large DKB, labelers re-use codes infrequently in practice: 0.4% of “NDCs” were flagged as re-used.

Coverage of supplies

In the outpatient setting, physicians prescribe, and pharmacies dispense, some items such as: cotton balls, alcohol swabs, insulin syringes, home glucose test strips. These items are not medications. They are usually referred to as “supplies”. Supplies have national product codes which the DKBs and pharmacy messages carry in the same field as the drug product codes. All of the commercial DKBs contain these commonly prescribed supply items. So a system using one of these commercial databases can convert product codes for supplies to a more general DKB code, analogous to that DKB’s clinical drug code and convert all prescribed items into a more clinically useful form. However, using the definition of supplies provided by one of the DKBs, we found that only 2.1% of the unique codes and 1.8% of the prescription volume (from a combined data set consisting of all of our sample prescriptions) represented supplies. At this time, RxNorm does not include general supplies in its database, as the commercial DKBs do. But this lack did not have much influence on its coverage rate, due to the relatively low prevalence of general supplies in prescriptions.

DISCUSSION

Our goal in this work was to assess the problems of converting prescribed entity codes, as delivered in HL7 and NCPDP pharmacy messages, into clinical drug codes, as defined by DKB suppliers. This conversion is but one step in the process of aggregating prescription records

from many different sources. But if the DKBs do not include the product codes delivered in pharmacy messages, this step will be the rate limiting one.

In the case of outpatient pharmacy records, DKBs do cover the codes for the prescribed entities well. The success with RxHub and Medicaid prescription coverage is notable because both sources are comprised of data from multiple pharmacy benefit managers and hundreds of pharmacies. The high outpatient coverage rate is due to the fact that outpatient pharmacies invent product codes rarely (less than 0.21% of the message volume) and the fact that DKBs tend to cover the commonly prescribed outpatient NDCs well. The DKB coverage of inpatient medication messages was not as good as that for comparable outpatient messages, due in large part to the higher usages of invented codes – up to 7.4% of the total prescription volume at some hospitals.

Based on our numbers, a clinical system that received prescription messages from outpatient settings could map more than 99% of them automatically to a clinical drug code that would enable useful clinical displays, decision support, performance measurement and research. For the very few (less than 1%) of outpatient messages that could not be mapped automatically, the receiving system could store them under one “miscellaneous drug” code. The drug names of these few could still be displayed under that miscellaneous code for human viewing, but these non-mapped prescriptions could not be used for the analytic purposes described above. Receivers could choose to review the unmapped codes as they accumulate and manually map those that occur frequently in a few minutes per week.

This same set of strategies could also be applied to the inpatient setting. However, a greater proportion of the received messages would not be covered, so more orders (5 to 10%) would be coded as “miscellaneous drug” and proportionately more codes would have to be reviewed to reach the level of coverage available in the outpatient setting. However non-mapped medications could still be displayed to care providers by name. It may not be as important to have inpatient medications fully coded for long term care, because most of the continuing drugs would be retrieved from outpatient sources.

Non-coverage can be due to gaps in the chosen DKB’s coverage or to the delivery of invented codes by the source pharmacy. Receivers can mitigate non-coverage by picking a DKB that covers the sources of most interest to them. We understand that DKBs tend to fill in “NDC” gaps that their customers experience – so the portion of non-coverage due to DKB gaps is likely to be self-correcting. However, much of the non-coverage – especially on the inpatient side – is due to invented codes. Hospitals invent codes for items that do not have “NDCs” including locally compounded preparations, clinical trial drugs, rare supplies and non-dispensing actions (e.g., “remove Fentanyl patch”). They also invent codes as a workflow convenience to permit them to dispense medications which have “NDCs” without loading them into their master formulary file. Changes in pharmacy processes and pharmacy message standards could reduce or eliminate most invented codes. Inpatient pharmacies should send the NDC for the main active ingredient in the HL7 message – as community pharmacies now do in their NCPDP message – or use the available HL7 mechanism consistently for delivering NDCs for all ingredients. HL7 and NCPDP could mitigate the problem with invented codes for randomized trials, rare supplies, and non-dispensing actions, by defining a field that identified these special purpose codes as such.

The practice of dispensing medications without registering them in the pharmacy’s master formulary accounts for a large share of the inventing. This practice should be discouraged, because pharmacy safety checks can not be applied to drugs that have not been registered in the pharmacy’s master formulary. It is also important to note that hospital pharmacy systems tend to load the NDC code for a given clinical drug when they initially create a formulary entry.

They rarely change this code, even when they get a new supply of the same clinical drug with a different NDC. However, this practice will only have consequences when the initially stored NDC code is retired from the DKB they use – and could be corrected when that happens.

A better approach than mapping at the receiver site would be mapping at a central facility, such as a RHIO. Even better would be for prescription sources to identify the medications within their prescription message via a universal clinical drug code in addition to the “NDC” product code. Such a solution would eliminate any need for mapping and speak to the 2008 resolution from the American Society of Health-System Pharmacy (ASHP) to explore: “the potential benefits of supplementing or modifying the National Drug Code with a coding system that can be effectively used across the medication-use continuum.”¹⁹

In this context, the coverage of the “NDCs” by RxNorm, the public use knowledge base deserves comment. Considering all of the DKBs, the RxNorm mapping table provided the best or second best coverage of “NDCs” from all but one of our prescription sources. RxNorm is non-proprietary, and has been designated by the Federal Consolidated Health Information (CHI) committee as the standard for the clinical drug code in prescription messages.²⁰ It could supply the universal clinical drug codes that we need.

The FDA’s official table of NDCs covered a smaller proportion of the NDCs in our sample (range: 63.3% to 92.2%) than the DKBs. This relatively low coverage is due to the fact that labelers generate new “NDCs” as they need them, put them into use and the rest of the industry – including the DKB vendors – adopt them before they get into the FDA database. Delays in labelers submissions to the FDA²¹ and some delay in the entry of submissions into the FDA’s data base account for most of the discrepancy. In August of 2006, the FDA proposed changes to its regulations that will eliminate these discrepancies, provide a unique, chemical structure-based identifier for drug ingredients and a rich machine-readable source of information about all drug products.²² Some of these have already been adopted.²³ The comparisons in this report pre-date those changes, so it is likely that the coverage of the FDA’s data has improved since then.

Through collaboration with the FDA, NLM is deploying this drug information and cross-linking the FDA’s identifiers with RxNorm.²⁴ The full set of FDA proposals, when fully implemented, will assure that all drug products have an official NDC and will ease the whole process of mapping NDCs to clinical drug codes.

Much of our data came from Indianapolis hospitals; so we can not be certain that our results for DKB coverage and invented codes generalize to other parts of the country. But the experience does go beyond one major city. The Medicaid database included records from the entire state of Indiana and many hundreds of pharmacies, as did the RxHub prescription sample. Nearly 1.4% of the RxHub prescriptions came from outside of Indiana (personal communication with RxHub). Further, Indiana is in the middle ranges of drug usage for all but one class of medications (Macrolides)²⁵, so it may be representative of other states. Though the hospital pharmacy data came from hospitals in Indianapolis or collar counties, it represents a broad spectrum of five different hospital information, and pharmacy, systems. Nonetheless, this is the first report of its kind; so until groups from other parts of the country publish similar data, we will not know the full possible variation.

The mapping of incoming codes to the more general clinical drug code is one step in the process of integrating prescription information from many outside sources in a medical record system. Prescriptions for the same patient from different sources may have different patient identifiers. The linking of these disparate identifiers is another required step, but is beyond the scope of our paper. However, this problem can be solved through standard linkage techniques if the

prescription records come with enough patient registration information.²⁶ It has been solved by RxHub and by RHIOs – such as the INPC – for their targeted scope.

Receiving systems also need machinery for aggregating the clinical drug to a higher level of categorization, to enable decision support and statistical analysis. For example, Amoxicillin 250 mg capsules can be generalized to Amoxicillin oral preparations, or further to Penicillins, or even further to Antibiotics. All of the commercial knowledge bases have built data structures and hierarchies that permit such aggregation. Indeed they carry rich content including the ingredients for each drug (important for allergy checking) and definitions of drug-drug, drug-test, drug-diagnosis and/or drug-food interactions. Most also include human-readable information about the drug, designed for physicians, pharmacies, and/or patients (depending upon the source DKB). They also feature attributes that facilitate charging – e.g., the average wholesale price. We did not obtain or examine this rich additional information, so we cannot comment on these capabilities, except to say that they are invaluable for many drug-prescribing and clinical care purposes, and the knowledge base vendors differ in the specific tools and services they provide.

For research and historical purposes, many groups will want to incorporate archival prescription records in their medical record database along with current medications. To include such older prescriptions, one will need to ask one's DKB vendor for all of their old "NDCs" which are not part of their standard release.

The results suggest that a receiving system can automatically aggregate outpatient prescriptions from many sources and store them under the clinical drug code of an appropriate DKB. Because of the maturity of standards (for codes and messages) medication records are the most ripe for electronic sharing. RHIOs, office practices, hospitals, nursing home, payers and researchers could all benefit from shared access to all of the medication records of their patients. We should get on with it.

CONCLUSION

DKBs cover the product codes used by outpatient sources sufficiently well to permit automatic mapping. Changes in policies and standards could increase coverage of product codes used by inpatient sources.

Acknowledgments

This work was performed at the Regenstrief Institute, Indianapolis, IN. It was supported, in part, by grant T15 LM07117 from the National Library of Medicine; in part, by National Cancer Institute grant U01 CA91343, a Cooperative Agreement for the Shared Pathology Informatics Network; in part, by Biomedical Information Science and Technology Initiative grant P20 GM66402 from the National Institutes of Health; and in part by a grant from the Indiana Twenty-First Century Research and Technology Fund for proposal ID 510040784. We wish to thank Anne Belsito, John Clifford, and Larry Lemmon for their assistance processing the data. We especially thank Sandy Poremba for preparation of this manuscript. We are very grateful to these drug knowledge base vendors for allowing us to examine their products at no cost: Cerner Multum, Inc.; First DataBank, Inc.; Thomson Healthcare, Inc.; Wolters Kluwer Health, Inc. We are also grateful to the hospitals of Indianapolis for providing their formularies and de-identified prescription messages.

References

1. Institute of Medicine. Preventing medication errors. Washington, DC: National Academies Press; 2006.
2. Joint Commission website. Hospital and Critical Access Hospital National Patient Safety Goals webpage. 2007 [Accessed May 2008]. Available from: http://www.jointcommission.org/PatientSafety/NationalPatientSafetyGoals/07_hap_cah_npsgs.htm

3. Poon EG, Blumenfeld B, Hamann C, et al. Design and implementation of an application and associated services to support interdisciplinary medication reconciliation efforts at an integrated healthcare delivery network. *J Am Med Inform Assoc* 2006 Nov-Dec;13(6):581–92. [PubMed: 17114640]
4. RxHub website. Available from: <http://www.rxhub.net>. Accessed May 2008.
5. Halamka J, Aranow M, Ascenzo C, et al. Health care IT collaboration in Massachusetts: the experience of creating regional connectivity. *J Am Med Inform Assoc* 2005 Nov-Dec;12(6):596–601. [PubMed: 16049225]
6. National Council for Prescription Drug Programs. Telecommunication Standard Implementation Guide. Version 5.1. Scottsdale, AZ. September 1999.
7. Health Level Seven. HL7 Messaging Standard Version 2.5. An application protocol for electronic data exchange in healthcare environments [standard]. Ann Arbor, MI: Health Level Seven; 2003.
8. U.S. Food and Drug Administration website. National Drug Code Directory webpage. [Accessed May 2008]. Available from: <http://www.fda.gov/cder/ndc>
9. Cimino JJ, McNamara TJ, Meredith T, et al. Evaluation of a proposed method for representing drug terminology. *Proc AMIA Symp* 1999:47–51. [PubMed: 10566318]
10. Nelson SJ, Brown SH, Erlbaum MS, et al. A semantic normal form for clinical drugs in the UMLS: early experiences with the VANDF. *Proc AMIA Symp* 2002:557–61. [PubMed: 12463886]
11. Sperzel WD, Broverman CA, Kapusnik-Uner JE, et al. The need for a concept-based medication vocabulary as an enabling infrastructure in health informatics. *Proc AMIA Symp* 1998:865–9. [PubMed: 9929342]
12. National Library of Medicine website. RxNorm webpage. [Accessed May 2008]. Available from: <http://www.nlm.nih.gov/research/umls/rxnorm/index.html>
13. Health Insurance Reform: Standards for Electronic Transactions, Final Rule. 65 Fed. Reg. 50329. August 17, 2000.
14. U.S. Food and Drug Administration website. National Health Related Items Code webpage. [Accessed May 2008]. Available from: <http://www.fda.gov/cdrh/nhric/nhric.html>
15. GS1 US website. Bar Codes and eCom webpage. [Accessed May 2008]. Available from: <http://barcodes.gs1us.org>
16. McDonald CJ, Overhage JM, Barnes M, et al. INPC Management Committee. The Indiana Network for Patient Care: a working local health information infrastructure. *Health Affairs* 2005;24(5):1214–20. [PubMed: 16162565]
17. National Council for Prescription Drug Programs. External Code List. Scottsdale, AZ. January 2007.
18. Notification of registrant; drug establishment registration number and drug listing number, 21 C.F.R. Sect.207.35.
19. American Society of Health-System Pharmacists. ASHP House of Delegates takes on pressing professional issues in Seattle; June 20, 2008; [Press release]
20. Letter to Secretary of the Department of Health and Human Services Tommy Thompson from Chair of the National Committee on Vital and Health Statistics John Lumpkin [letter]. 2003 Nov 5. Available from: <http://www.ncvhs.hhs.gov/031105lt3.pdf>. Accessed May 2008.
21. Department of Health and Human Services, Office of Inspector General. The Food and Drug Administration's National Drug Code Directory. August 2006.
22. Food and Drug Administration. Requirements for Foreign and Domestic Establishment Registration and Listing for Human Drugs, Including Drugs that are Regulated Under a Biologics License Application, and Animal Drugs [proposed rule]. August 23, 2006.
23. Schadow G. Assessing the impact of HL7/FDA Structured Product Label (SPL) Content for Medication Knowledge Management. *Proc AMIA Symp* 2007:646–50. [PubMed: 18693916]
24. National Library of Medicine website. DailyMed webpage. [Accessed May 2008]. Available from: <http://dailymed.nlm.nih.gov>
25. Motheral BR, Cox ER, Mager D, et al. Express Scripts prescription drug atlas, 2001: a study of geographic variation in the use of prescription drugs. Express Scripts, Inc. January 2002.
26. Grannis SJ, Overhage JM, McDonald CJ. Analysis of identifier performance using a deterministic linkage algorithm. *Proc AMIA Symp* 2002:305–9. [PubMed: 12463836]

APPENDIX A

Methods for Counting the Unique Clinical Drugs in Each DKB

RxNorm includes 29,734 unique identifiers of type “Semantic Clinical Drug”. We removed 11,295 deprecated codes that were marked as “obsolete” in the “Suppress” field of the Concept (RXNCONSO.RRF) file, and another 10,357 codes that were not linked to any “NDC” in the Attribute (RXNSAT.RRF) file. Thus we obtained a count of 8,082 unique RxNorm “Semantic Clinical Drugs”.

The Medi-Span MDDB® includes 11,894 unique Generic Product Identifiers (GPIs). We removed the 669 GPIs in the “Medical Devices” category and the 1,389 GPIs in the “Chemicals” category, to obtain a total of the medications.

For the Thomson Micromedex Red Book™, we began with all the 25-digit codes in the UltiMedex hierarchy. These codes define a polyhierarchy, so we focused on the last 10 digits of this code, which uniquely identifies the drug name, route, dose form, and strength, and squeezed out duplicates to obtain the count of unique clinical drug codes.

For the Multum® Lexicon, we used the count of “main_multum_drug_codes” – a unique identifier for each combination of drug name, route, dose form, and strength in the Multum® drug database.

For the First DataBank NDDF Plus™, we started with the 20,301 Clinical Formulation Identifiers (“GCN_SEQNOs”). Then, we subtracted 2,165 identifiers with Therapeutic Class “Supply”, 1,060 identifiers with Therapeutic Class “Bulk Chemicals”, and 2,671 identifiers with a null Therapeutic Class.

The Drug Knowledge Bases.

Table 1

Drug Knowledge Base	Contact Information	Version	Count of Distinct NDCs	Count of Distinct Clinical Drugs ^a
FDA National Drug Code Directory	ddis@cderr.fda.gov 301-210-2840	Jan 2006	122,117	N/A
Medi-Span Master Drug Data Base (MDDB®) ^b	medispan-support@wolterskluwer.com 800-388-3884	Feb 2006	126,042	9,836
Multum® Lexicon ^c	info@multum.com 800-968-5886	Feb 2006	113,221	8,383
National Drug Data File Plus (NDDF Plus TM) ^d	inquiry@firstdatabank.com 800-633-3453	Feb 2006	117,025	14,405
RED BOOK ^{TM,e}	mdx.info@thomson.com 800-525-9083	Feb 2006	181,113	13,801
RxNorm	rxnorminfo@nlm.nih.gov 888-346-3656	Feb 2006	232,111	8,082

^aFor reasons given in the text, these numbers are not directly comparable. Please see Appendix for details of the calculation.

^bMDDB® is a registered trademark of Wolters Kluwer Health, Inc. (Conshohocken, PA)

^cMultum® is a registered trademark of Cerner Multum, Inc. (Denver, CO)

^dNDDF PlusTM is a trademark of First DataBank, Inc. (San Bruno, CA)

^eRED BOOKTM is a trademark of The Thomson Corporation (Greenwood Village, CO)

Table 2

DKB percent coverage of unique codes found in each of the listed sources.

	FDA ^a	MDDb ^b	MMSL ^c	MMX ^d	NDDF ^e	RxNorm	None of 6
Outpatient Sources							
Hospital E Pharmacy Messages (n=2834)	56.5%	72.5%	77.7%	74.7%	69.6%	76.4%	21.2%
Medicaid Archival Records (n=41727)	45.5%	75.8%	80.3%	80.9%	74.5%	79.3%	2.4%
RxHub Formularies (n=117151)	29.7%	78.4%	51.1%	64.9%	90.1%	52.5%	2.5%
RxHub Messages (n=7838)	78.5%	97.8%	94.3%	95.2%	97.9%	94.0%	1.6%
Inpatient Sources							
Hospital A Formulary (n=3598)	65.0%	83.4%	85.9%	83.1%	80.9%	86.0%	13.3%
Hospital A Messages (n=1641)	70.6%	87.4%	88.8%	86.7%	85.8%	89.2%	10.3%
Hospital B Formulary (n=2317)	59.0%	71.6%	92.3%	84.9%	67.8%	93.5%	5.9%
Hospital B Messages (n=1716)	54.3%	64.9%	80.2%	75.5%	62.2%	81.1%	18.5%
Hospital C Formulary (n=3820)	76.0%	94.7%	93.5%	93.3%	94.7%	93.6%	4.4%
Hospital C Messages (n=1206)	74.5%	91.3%	90.7%	90.3%	91.4%	90.1%	8.1%
Hospital D Formulary (n=2905)	70.2%	86.8%	91.4%	90.2%	86.2%	91.3%	7.8%
Hospital D Messages (n=3439)	59.9%	72.1%	75.6%	74.4%	71.4%	75.8%	23.7%
Hospital E Formulary (n=2543)	73.5%	90.8%	91.1%	90.1%	90.8%	90.4%	7.2%

^aFDA=Food & Drug Administration NDC Directory (The FDA source is not literally a DKB; but we included it for comparison purposes.)

^bMDDb=Medi-Span Master Drug Data Base

^cMMSL=Multum Lexicon

^dMMX=Thomson Micromedex Red Book

^eNDDF=First DataBank National Drug Data File.

Table 3

The 13 Data Sources: Distinct Identifiers.

	Unique Codes	Invented Codes	Non-invented NDCs, not in any DKB	Non-invented NDCs, found in ≥ 1 DKB
Outpatient Sources				
Hospital E Pharmacy Messages	2,834	20.7%	0.42%	78.9%
Medicaid Archival Records	41,727	0.01%	2.4%	97.6%
RxHub Formularies	117,151	0.48%	2.0%	97.5%
RxHub Messages	7,838	1.6%	0.01%	98.4%
Inpatient Sources				
Hospital A Formulary	3,598	12.3%	0.92%	86.8%
Hospital A Messages	1,641	9.6%	0.67%	89.7%
Hospital B Formulary	2,317	1.6%	4.2%	94.2%
Hospital B Messages	1,716	16.0%	2.6%	81.4%
Hospital C Formulary	3,820	4.1%	0.31%	95.6%
Hospital C Messages	1,206	7.5%	0.66%	91.8%
Hospital D Formulary	2,905	6.8%	1.0%	92.2%
Hospital D Messages	3,439	22.9%	0.73%	76.4%
Hospital E Formulary	2,543	6.2%	0.94%	92.9%

The first data column gives the number of distinct codes. The remaining three data columns give the percentage breakdown of these distinct codes.

Table 4

DKB percent coverage of total medication records.

	FDA ^a	MDDb ^b	MMSL ^c	MMX ^d	NDDF ^e	RxNorm	None of 6
Outpatient Sources							
Hospital E Pharmacy Messages (n=464404)	88.5%	97.6%	99.6%	93.6%	93.0%	98.2%	0.28%
Medicaid Archival Records (n=66291780)	81.0%	98.7%	98.7%	98.2%	97.3%	98.7%	0.17%
RxHub Messages (n=304855)	92.2%	99.8%	99.0%	99.5%	99.8%	99.0%	0.17%
Inpatient Sources							
Hospital A Messages (n=40855)	74.7%	89.0%	91.7%	87.7%	87.5%	92.1%	7.8%
Hospital B Messages (n=149675)	63.3%	80.7%	95.4%	90.9%	77.4%	96.0%	4.0%
Hospital C Messages (n=40622)	73.7%	96.9%	96.9%	95.8%	97.0%	96.7%	2.6%
Hospital D Messages (n=396689)	81.7%	93.7%	95.7%	93.9%	93.5%	95.8%	3.5%

^aFDA=Food & Drug Administration NDC Directory (The FDA source is not literally a DKB; but we included it for comparison purposes.)

^bMDDb=Medi-Span Master Drug Data Base

^cMMSL=Multum Lexicon

^dMMX=Thomson Micromedex Red Book

^eNDDF=First DataBank National Drug Data File.

Table 5

Seven Data Sources of Messages.

	Total Number of Records	Invented Codes	Non-invented NDCs, not in any DKB	Non-invented NDCs, found in ≥1 DKB
Outpatient Sources				
Hospital E Pharmacy Messages	464,404	0.21%	0.06%	99.73%
Medicaid Archival Records	66,291,780	0.11%	0.06%	99.83%
RxHub Messages	304,855	0.17%	0.001%	99.83%
Inpatient Sources				
Hospital A Messages	40,855	7.4%	0.35%	92.2%
Hospital B Messages	149,675	1.3%	2.7%	96.0%
Hospital C Messages	40,622	1.5%	1.1%	97.4%
Hospital D Messages	396,689	3.2%	0.30%	96.5%

The first data column gives the total number of prescription messages/records in each source. The remaining three data columns give the percentage breakdown of the message volume.