An Examination of OMOP Common Data Model Vocabulary for Completeness with respect to Drugs, Conditions and Procedures

Hugh Kawabata
Xiao Shao
Christopher Davidson
Perfecto Gayan
Teresa Simon
Bristol-Myers Squibb

Bristol-Myers Squibb
Global Health Economics
and Outcomes Research

CORDS
GHEOR Center for Observational Research & Data Sciences

Background

As part of OMOP's Common Data Model (CDM), a database of standard vocabularies is provided. This vocabulary provides mappings of commonly used terminologies and vocabularies both across them as well as a set of hierarchies. While there is tremendous value of the mappings and hierarchies, the completeness of this database need to be assessed for use with real world data.

Objectives

Assess the completeness of the OMOP supplied CDM Vocabulary (version 4.2 Q1-2013) with respect to NDC codes to increase our understanding and confidence in it.

Methods

- We created a table of unique NDC codes from four claims databases:
  - A - Claims database from employers and insurers
  - B - Payor claims database (large payor)
  - C - Collection of payors claims data
  - D - Payor claims database (large proportion Medicare)
- We searched the CDM Vocabulary 4.2 Q1-2013 and for those that we were unable to find, we put them into 4 categories:
  - Medicines
  - Prescription vitamins/supplements
  - Medical supplies
  - Other, unclassified (no names/descriptions associated with the codes)
- For conditions and procedures we calculated unmapped proportions

Results - Overall

- We found 78,459 unique NDC's from the 4 claims databases.

Database	# Claims	# of NDC codes	# ,% of NDC codes not in Vocabulary	#, % Claims for the codes not found
A - Claims database from employers and insurers 2012 Q2	135,216,769	39,827	7,603 19%	2,648,200 2%
B - Payor claims database (large payor) 2012 Q4	26,231,641	25,866	4,035 16%	489,195 2%
C - Collection of payors claims data 2006-2012	866,279,955	44,278	8,995 20%	4,191,039 0.5%
D -Payor claims database (large prop Medicare) 2007-2012	922,138,030	65,617	12,328 19%	11,302,742 1%

Results – Claims database from employers and insurers

- For the 7,603 unique NDC codes that were not in the CDM Vocabulary, they were distributed as follows:

Category	# Codes	# Claims, %	Examples
Medicines	1,921	154,089 5.8%	Nystatin, Progesterone
Prescription vitamins/supplements	500	146,533 5.5%	Strovite One Caplets
Medical supplies	1,241	1,272,413 48.0%	Syringe, Sure Comfort
Other	3,941	1,075,165 40.7%	Unknown NDC code

CORDS
GHEOR Center for Observational Research & Data Sciences

Results – Payor claims database (large payor)

- For the 4,035 unique NDC codes that were not in the CDM Vocabulary, they were distributed as follows:

Category	# Codes	# Claims, %	Examples
Medicines	1,211	32,378 6.6%	Progesterone, Testosterone
Prescription vitamins/supplements	230	23,195 4.7%	Prenatabs Rx
Medical supplies	944	261,546 53.5%	Wound Dressings, Dermabase
Other	1650	172,076 35.2%	Unknown NDC code

Results – Collection of payors claims data

- For the 8,995 unique NDC codes that were not in the CDM Vocabulary, they were distributed as follows:

Category	# Codes	# Claims, %	Examples
Medicines	1,549	118,683 2.8%	Hydrocortisone, Methyltestosterone
Prescription vitamins/supplements	703	234,675 5.6%	Prenatabs Rx, Strovite One Caplets
Medical supplies	1,929	1,962,111 46.8%	Cream Base, Lancets
Other	4,814	1,875,570 44.8%	Unknown NDC code

Results – Payor claims database (large proportion Medicare)

- For the 12,328 unique NDC codes that were not in the CDM Vocabulary, they were distributed as follows:

Category	# Codes	# Claims, %	Examples
Medicines	2,913	328,866 3.0%	Triamcinolone Acetonide (TOPICAL), Estradiol
Prescription vitamins/supplements	2,205	690,873 6.1%	Multivitamin, Vitaplex
Medical supplies	4,330	8,344,834 73.8%	Atopiclair
Other	2,880	1,938,169 17.1%	Unknown NDC code

Results – Overall

- For the 18,181 unique NDC codes that were not in the CDM Vocabulary, they were distributed as follows:

Category	# Codes	% Code
Medicines	3,687	20.3%
Prescription vitamins/supplements	2,464	13.6%
Medical supplies	4,923	27.1%
Other	7,107	39.0%

Results - ICD-9 Diagnosis (22,555 distinct ICD-9-CM diagnosis codes)

Source/Vendor	Total number of distinct codes from source	Number & % Unmapped distinct source codes in V4
A - Claims database from employers and insurers (2012 Q2)	14,117	0 0%
B - Payor claims database (large payor)(2012 Q4)	17,925	2,933 16%
C - Collection of payors claims data (2006-2012)	18,849	1,582 8%
D - Payor claims database (large proportion Medicare) (2007-2012)	17,396	227 1%

CORDS
GHEOR Center for Observational Research & Data Sciences

Results - CPT Procedures (10,453 distinct CPT codes)

Source/Vendor	Total number of distinct codes from source	Number & % Unmapped distinct source codes in V4
A - Claims database from employers and insurers (2012 Q2)	9,227	19 0.2%
B - Payor claims database (large payor)(2012 Q4)	8,648	161 2%
C - Collection of payors claims data (2006-2012)	9,966	253 2%
D - Payor claims database (large proportion Medicare)(2007-2012)	11,056	1,176 11%

Results – HCPCS Procedures (6,752 distinct codes)

Source/Vendor	Total number of distinct codes from source	Number & % Unmapped distinct source codes in V4
A - Claims database from employers and insurers (2012 Q2)	4,808	207 4%
B - Payor claims database (large payor)(2012 Q4)	3,729	109 3%
C - Collection of payors claims data (2006-2012)	5,374	256 5%
D - Payor claims database (large proportion Medicare)(2007-2012)	---	---

Conclusions

- For NDC codes, findings:
  1. The majority of the missing NDC codes are non-prescription drugs
  2. But there are significant number of medicines (2.8%-6.6%) that are missing, such as testosterone, triamcinolone and metoprolol tartrate.
  3. While the non-prescription missing NDC codes will be present in the CDM formatted data (in the *source_to_concept_map* table), they are not mapped. Hence, we intend to keep a table of these NDC codes.
- For the condition and procedure codes, the rate of missing codes appear to be mostly errors in data entry.
- The missing NDC codes are important even if they are not prescription medications for two main reasons: (1) in costs studies, these do contribute to the overall costs; and, (2) exposure to these may be important in some studies.

CORDS
GHEOR Center for Observational Research & Data Sciences