Conversion of the Premier database to the OMOP Common Data Model

Rupa T. Makadia, MS[1], Patrick B. Ryan, PhD[1]

[1]Janssen Research & Development, LLC, Titusville, NJ

BACKGROUND

- The Premier database is an inpatient/outpatient hospital billing database in the United States, which consists of 467 hospitals.
- Data spans from 2000-2013.
- The data is oriented as a visit centric model.

OBJECTIVES

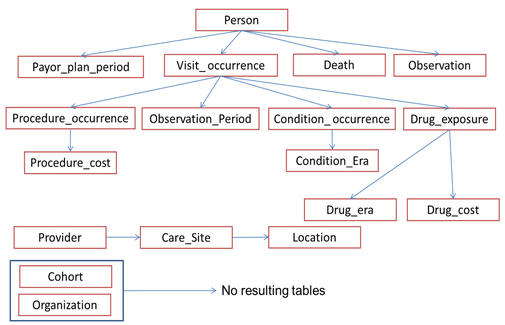
- Describe the conversion of the Premier hospital database into OMOP CDM.
- Validation exercises:
  1.) Replicating a research study and identify advantages of using the CDM over the raw data format.
  2.) Comparing conditions in claims database to CDM Premier.

Hypothesis: Transforming the data into the OMOP CDM and using it in the replication study will strengthen the use of the CDM. When comparing the inpatient data from a claims database to Premier the stratified results by age and gender will produce similar results.

METHODS

Data mapping: Data was mapped from the raw format to CDM.

Figure 1. CDM Data Mapping



Standard charge code mapping:

- Pharmacy department codes were transformed into RxNorm concept ID's, the remaining codes were transformed into SNOMED procedure concepts.
- The mapping was using fuzzy string matching using Perl. A score was given from 0 to 1000.
- Scores below 400 was mapping to a concept ID of zero.

Visit logic:

- Logic was added to incorporate the number of service days and ordinality of visits.
- For those patients that only have one visit, the start date is the first of the month and the end date is the start date plus maximum number of service days.
- For patients that have multiple visits in the same month, the logic adds one day to the previous end date, to preserve order and length of stay for visit.

Validation 1: Replication study:

- The study "Estimating pediatric inpatient medication use in the U.S." by Lasky et al. was replicated [1].
- The study looks at the demographic characteristics of the pediatric population, in an inpatient setting in 2008.
- The top ten drugs in an inpatient encounter were obtained and the number of patients that had drug utilization was captured. String search was used in original study, vocabulary used in CDM replication.

Validation 2: Claims database comparison:

- The Optum database was used to compare inpatient conditions to Premier inpatient conditions in 2011.
- Persons with a valid observation period in 2011 with inpatient stays were obtained from both database conditions rolled up to MedDRA preferred terms.
- The proportion of patients in each database were captured and stratified by gender and age decile.

RESULTS

Table 1: Standard charge code mapping:
The number codes mapped represent 91.4% of the overall data. 1,000 codes were manually reviewed for accuracy which represents 72.9% of data.

Table 2: Demographics: Demographics from both the raw Premier and CDM are the same for discharge status and region, and less than 1% different for sex, admission source, region, and hospital type.

Table 3: Drug utilization: Drug administration prevalence for inpatient stays is similar for ceftriaxone and gentamicin. Utilization for acetaminophen is higher in the CDM than raw data.

Table 1. Standard charge code mapping statistics

Statistic value	Number	Percent of codes	Number of bill records	Percent of data
Total Number of codes	55,470	100.0%	9.6 B	100.0%
Total number of codes (mappable)	44,346	79.9%	8.7 B	91.4%
Total mapped to zero	6,647	12.0%	1.2 B	12.7%
Total number of codes reviewed	1,000	1.8%	6.9 B	72.9%
Total re-mapped by hand	162	0.3%	7.7 M	8.0%

Table 2. Demographic information from replication study

Demographic	Study results 2008	Premier 2008	CDM Premier 2008
Sex			
Male	50.9%	49.9%	50.2%
Source of admission			
Routine including births and other sources	72.0%	65.0%	65.6%
Other hospital or healthcare facility	13.7%	19.1%	18.6%
Emergency department	14.8%	15.9%	15.7%
Discharge status			
Routine including births and other sources	94.0%	97.4%	97.4%
Died	0.4%	0.3%	0.3%
Other	5.6%	2.2%	2.2%
Payer			
Medicare/Medicaid/other government payer	45.9%	46.1%	49.2%
Private insurance	46.2%	51.9%	55.9%
Other	7.9%	2.1%	2.4%
Mean length of stay			
Days	3.7	4.0	3.8
Region			
Midwest	18.7%	19.7%	19.7%
Northeast	14.3%	18.4%	18.4%
South	48.7%	43.3%	43.3%
West	18.3%	18.6%	18.7%
Teaching status			
Teaching hospital	41.5%	44.6%	42.9%
Urban vs. rural			
Urban	89.2%	90.5%	90.7%
Bed size			
Small	10.3%	13.9%	16.7%
Medium	17.9%	16.5%	15.7%
Large	71.8%	69.6%	67.6%

Table 3. Inpatient drugs in pediatric population in 2008

Drug Name	Study results per 100 patients	Premier 2008 per 100 patients	CDM Premier 2008 per 100 patients	Difference from CDM Premier
Acetaminophen	14.7	15.0	19.6	▲
Albuterol	5.1	6.3	5.1	▼
Ampicillin	8.0	9.0	8.3	▼
Ceftriaxone	5.6	5.7	5.7	--
Fentanyl	6.6	8.2	7.6	▼
Gentamicin	6.6	6.8	6.8	--
Ibuprofen	6.3	7.5	7.1	▼
Lidocaine	11.0	15.0	14.7	▼
Morphine	6.2	7.3	7.1	▼
Ondansetron	6.2	7.1	6.9	▼

Figure 2:Comparing Premier to Claims database:
Similarities between inpatient condition concepts are not inherent unless stratified by age and gender.

Figure 2. Proportion of CDM Premier vs. CDM Optum

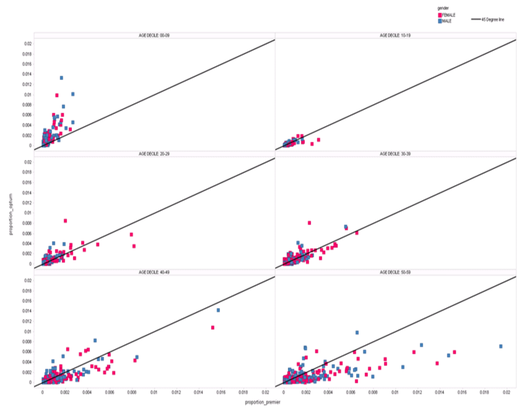


Table 4: Proportions for hypertension stratified:
The proportions for MedDRA concept "Essential hypertension" for both databases are highlighted to show individual proportions are similar between age groups 10-19, 20-29, 40-49 in males and 20-29 in females.

Table 4. Proportions for essential hypertension stratified

Age Decile	Gender	Premier Proportion	Optum Proportion
00-09	Female	0.00008	0.00048
10-19	Female	0.00021	0.00009
20-29	Female	0.00169	0.00157
30-39	Female	0.00556	0.00702
40-49	Female	0.01521	0.01081
50-59	Female	0.02923	0.01430
00-09	Male	0.00011	0.00072
10-19	Male	0.00029	0.00027
20-29	Male	0.00179	0.00193
30-39	Male	0.00550	0.00738
40-49	Male	0.01572	0.01421
50-59	Male	0.03159	0.02072

CONCLUSIONS

Transformation of Premier to OMOP CDM

- Without standard charge code mapping none of the Premier data could have been associated to a concept ID, and out of all mappable codes over 90%were mapped.
- Adding visit logic helps build complete observation period in the CDM, and adds additional information that was not present prior.

Replication study

- The demographic information from the replication study indicate that there is no loss of data when transforming data.
- The drug information from the replication study suggests that by using the OMOP vocabularies can yield more complete results rather than string searches alone.

Comparing to claims database

- Comparing overall inpatient conditions from Premier and Optum with stratification for ages <60 indicates that the databases have similar distributions despite Premier being a hospital database.

REFERENCES

1. Lasky., et al., *Estimating pediatric inpatient medication use in the United States.* Pharmacoepidemiol Drug Saf. 2011 Jan;20(1):p 76-82

Contact: rmakadia@its.jnj.com

This study was conducted by Janssen Research & Development, LLC.