

Common Data Model Conversion in AsPEN for SCAN Project

Y. Zhang¹; E. C. C. Lai²; N.C. Hardy¹; L. Lin¹; S. Setoguchi-Iwata¹

¹Duke Clinical Research Institute, Durham, NC; ²National Cheng Kung University, Tainan City, Taiwan

Background

- Asian Pharmacoepidemiology Network (AsPEN) is a multinational research network established to:
 - Provide a mechanism to support the conduct of pharmacoepidemiological research and;
 - Assess safety and effectiveness of medications and other therapeutic modalities in participating countries.
- The Surveillance of Health Care in Asian Network (SCAN) project aims to develop infrastructure for AsPEN and characterize the health and health care utilization of the populations covered in each participating AsPEN database.

Objective

To refine and adapt an existing CDM for healthcare data sources among AsPEN sites participating in the SCAN project, and bridge the gap among the coding systems used in each country or region with OMOP standard vocabularies.

Methods

Data Sources and Participating AsPEN Sites:

•US Medicare claims data, Japan Medical Data Center (JMDC) data, Taiwan National Health Insurance Research Database (NHIRD) data, and Hong Kong Clinical Data Analysis and Reporting System (CDARS) data

Data conversion standards:

•Created a logic map describing the relationship between the source data and the common data model data at both the table level and variable level.

Extraction, transformation and loading (ETL) process:

•Used SAS to developed and implement the ETL process
 •Extracted selected columns from the source data to the transaction table to avoid wasting resources on unneeded columns.

Methods (cont.)

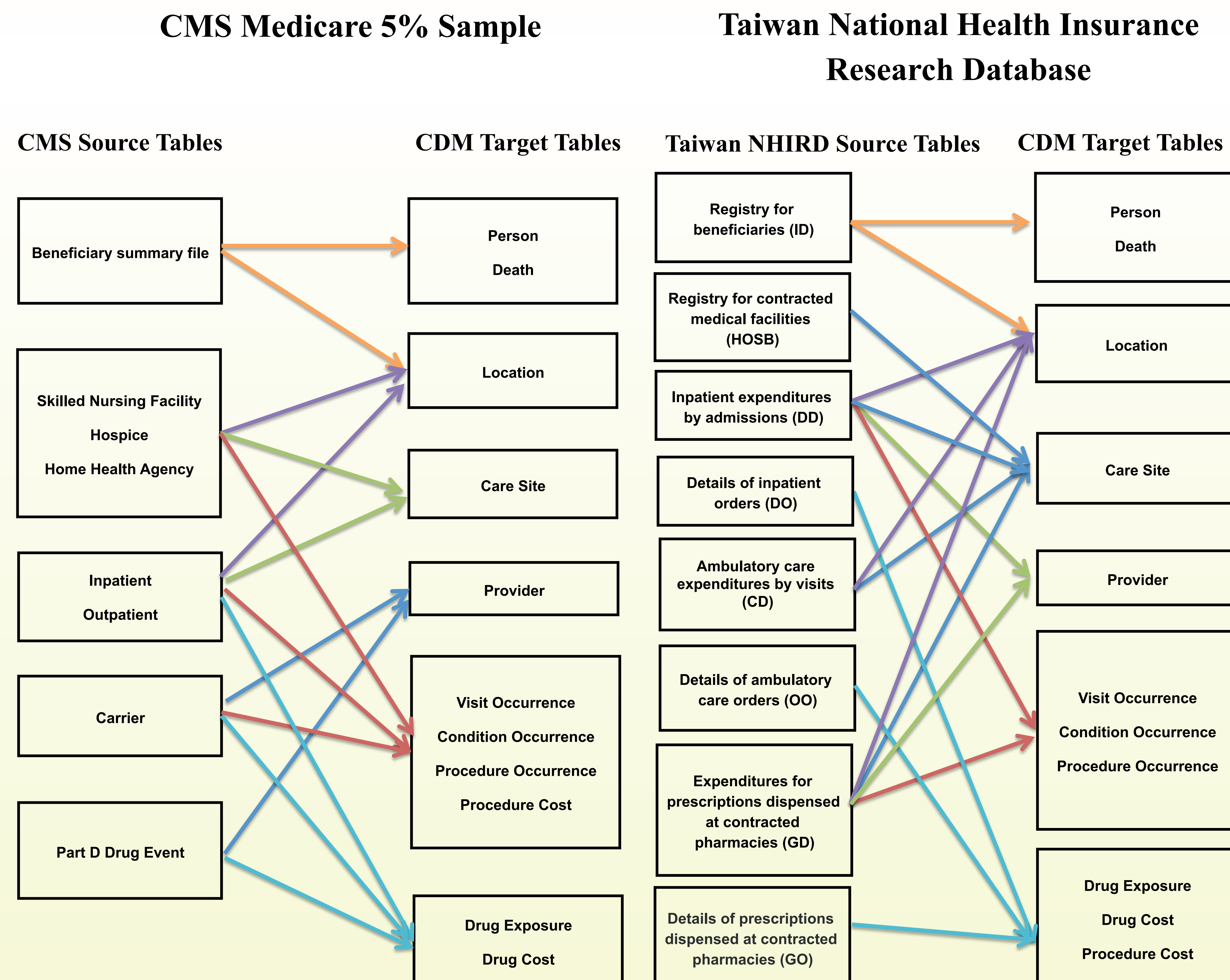
- Customized crosswalks between each local standard and OMOP vocabulary prior to implementation.
- Created administration tables to organize and track the ETL process.

First step of approach:

- Began conversion of two years of Medicare data to the CDM.
- Non-US sites underwent a mapping process from the national drug code (NDC) to RxNorm through drug generic name (in English), dosage strength, and route of administration.

Results

Mapping of Data Tables to OMOP Common Data Model Tables



Funding/Support: Janssen Research and Development

Results (cont.)

- 94% of Taiwanese NDC drug codes were mapped to RxNorm. 5% remained unmapped for 2005-2010 prescription records.

Percentage of Mapped Concepts and Concept Usage Group by CDM Tables, Variables, and Vocabularies

Table	Variable	Vocabulary	Percent Mapped Concepts	Percent Usage
condition_occurrence	condition_concept_id	SNOMED-CT	65.3	83.2
		OMOP Vocabulary v4.3 2013-09-15	34.7	16.8
	condition_type_concept_id	OMOP Condition Occurrence Type	100.0	100.0
drug_era	drug_concept_id	RxNorm	100.0	100.0
		OMOP Drug Exposure Type	100.0	100.0
drug_exposure	drug_concept_id	RxNorm	71.8	97.2
		OMOP Vocabulary v4.3 2013-09-15	28.2	2.8
		OMOP Drug Exposure Type	100.0	100.0
	relevant_condition_concept_id	SNOMED-CT	100.0	3.2
person	ethnicity_concept_id	Ethnicity	100.0	100.0
		HL7 Administrative Sex	100.0	100.0
		CDC Race	100.0	100.0
procedure_occurrence	procedure_concept_id	HCPCS	19.1	11.0
		ICD-9-Procedure	19.1	0.7
		CPT-4	53.9	73.4
		SNOMED-CT	0.0	14.6
		OMOP Vocabulary v4.3 2013-09-15	10.9	0.4
		OMOP Procedure Occurrence Type	100.0	100.0
	relevant_condition_concept_id	SNOMED-CT	100.0	49.6
provider	specialty_concept_id	CMS Specialty	13.7	62.9
		NUCC	86.3	34.5
		OMOP Vocabulary v4.3 2013-09-15	0.0	2.6

Conclusions

- We demonstrated the feasibility of adopting and modifying OMOP CDM for US Medicare data.
- The diversity and unique drug and procedure coding system from each database are a major challenge in CDM conversion.
- Further discussions on coding standardization are needed.