

MAPPING MEDICAID AND MEDICARE CLAIMS TO THE OMOP COMMON DATA MODEL: LESSONS LEARNED FROM THREE DATA SOURCES

Traci E. Yamashita MS¹; Patricia St. Clair ScB²; Lisa M. Schilling MD, MSPH¹

¹University of Colorado, Aurora CO; ² University of Southern California Leonard D. Schaeffer Center for Health Policy & Economics, Los Angeles CA

BACKGROUND & OBJECTIVES

Mapping claims data to the OMOP Common Data Model (CDM) to conform to the clinical CDM is complex due to

- Different structure of claims data, e.g., unit of a claims file is the claim number-line number
- Inconsistent claims file format/content

Objective is to share lessons learned from mapping

- Medicaid claims from two states
- Medicare claims from CMS

RESULTS

MAPPING TO OMOP CDM VERSION 4

OMOP CDM TABLE	MEDICAID CHALLENGES [C] & WORK-AROUND [W]	MEDICARE CHALLENGES [C] & WORK-AROUND [W]
Organization	C: Must be created from claims and provider files; not a natural construct in claims structure; W: Followed TRUVEN method	C: : Must be created from claims and provider files; not a natural construct in claims structure; W: Not populating table since CDM for claims only
Care Site	Same challenges and work-around as Organization table.	Same challenges and work-around as Organization table.
Provider	C: Must be created from claims and provider files; Complicated algorithm to handle Organization and Care Site; W: Followed TRUVEN method	C: Must be created from claims. Parts A/B have attending/referring and other providers. Both UPIN and NPI identifiers used in claims. W: In progress. Keeping all provider IDs found, with most frequent specialty.
Person	C: Limited Data Sets so Person table primarily used for clinical-claims linkage – fields that are mapped to CDM come from clinical records; W: Populate with linkage fields only	C: Time varying beneficiary residence used to analyze geographic/neighborhood effects; only one location allowed in Person table; W: Add person-location table with person_ID, location_ID, start & end date
Payer Plan Period	C: (i) Some data sources are unavailable to validate algorithms that collapse gaps in eligibility less than 31 days into single eligibility period; (ii) end user may want to characterize fluctuation in Medicaid eligibility; W: Preserve detail eligibility information for end user	C: Unclear how to organize complexity of plan information which is critical to sample selection and separating missing info from zeros; W: Separate entries that identify unique statuses for a) Part A/B and HMO enrollment; b) Part D enrollment and benefit phase; c) dual eligibility and LIS status
Visit Occurrence	C: Inconsistent field population by state (values don't conform to OMOP vocab & large number of missing values); W: Choose simplest algorithm to define visits & document for end user C: Level of detail required for end user is dependent on research question; W: Define visit to accommodate clinical comparisons, but add place of service in Procedure Occurrence table to preserve detail.	C: Unclear how to define visit from physician visit, home health, and hospice claims; visit ties together conditions, treatments, and their costs; W: Set rules for combining claims into visits when appropriate
Condition Occurrence	C: Varying number of conditions received from states, e.g., primary diagnosis vs more complete list; W: Map all conditions received with clear documentation for end user	C: Unclear which provider, referring/attending or other. W: Use referring/attending. Add fields to link to performing or other providers associated with the claim.
Procedure, Occurrence/Cost	C: Claims unit is claim number/line number; W: Preserve claim unit detail for end users interested in specific procedure codes, e.g., record is generated for each claim number/line number where procedure code is not missing	C: OMOP vocabs don't include HIPPS codes (used to determine payments) for IP rehab, SNFs, home health; W: In progress. Request addition of HIPPS field to Procedure Occurrence table and OMOP vocab
Drug, Exposure/Cost	Same challenges and work-around as Procedure table. C: Population of cost fields vary widely by state; W: Map all costs received with clear documentation for end user	C: Part D claims costs (used to analyze the cost burden to various payers, e.g., patient, low income subsidies, other state programs) have structure different from OMOP Drug Cost table; W: Modify Drug Cost table to add fields for Part D cost information

METHODS

	MEDICAID #1	MEDICAID #2	MEDICARE
DATA SOURCE	State Medicaid agency	Data contractor	CMS
PRE-PROCESSING	Minimal – includes reversals/voided claims	Cleaned	Cleaned
FILE FORMAT/ CONTENTS	State-determined	APCD formatting	CMS formatting
ACCESS TO DATA	BAA for use of data from one of two partner sites	BAA for use of data from 2 partner sites	DUA with CMS
END USE	Clinical & Claims for multiple clinical sites	Clinical & Claims for multiple clinical sites	Claims Only
SAMPLE	EHR linked claims 2010-2012	EHR-linked claims 2009-2012	(i) HRS-linked claims. 1991-2008 (ii) 20% national sample, 2002-2011
CLAIM TYPES / FILES RECEIVED	Medical Claims, Pharmacy Claims, TPL, TPL Medicare Only, Eligibility, Provider, Surgical	Medical Claims (hdr/line/DX/ procedures), Pharmacy Claims (hdr/line), Member, Person, Provider	Inpatient, Outpatient, SNF, Home Health, Hospice, Physician, DME
OMOP TEMPLATE	TRUVEN mapping specifications to CDMV4 were used as a guide		Starting with HUMANA

CONCLUSIONS

Accurate representation of claims data in the OMPO CDMV4 schema requires:

- 1) Simple mapping algorithms to:
 - Accommodate inconsistent field population by various claims sources
 - Support comparisons across various sources
- 2) Preservation of the ability to retrieve original claims detail to:
 - Assess validity and reliability of the mapping process, ensure fidelity to source data
 - Support versatility of end-user hypothesis-driven data needs
- 3) Close collaboration with managers of source data to provide information beyond data dictionaries

CDM provides structure to combine clinical and claims data, however expensive human effort is required to convert claims structure to CDM. Capitalizing on movement of state Medicaid offices toward standardized all-payer claims files may reduce human effort and improve interoperability and harmonization.