

Representing ICD-11 Post-Coordination in OMOP Without Changing the Model

Xiaohan Tanner Zhang¹ ; Paul Nagy¹; Christian Reich²; Christopher G. Chute¹

¹ Johns Hopkins University School of Medicine, Baltimore, MD, USA. ² OHDSI.

Background

In ICD-11, a diagnosis is often built by post-coordination, a stem code plus one or more extension codes combined into a cluster at coding time(1). Common extensions are modifiers such as laterality and severity, but ICD11 also always for qualifier and relational extensions. This right ontology provides the opportunity to avoid massive duplication of coding, but it requires thought when applying to the oHDSI vocabulary structure. OMOP and the OHDSI vocabularies are designed around precoordinated, one standard concept per meaning, with no post-coordinated expressions. The discussion on whether OHDSI Vocabulary should support post-coordination has been open since 2019 and it has been asked what use case would actually require it(2).

The use case is analytic. Because ICD-11 keeps manifestation, anatomy, histopathology, and causal agents on separate axes, a study can select on one axis without the rest, which a precoordinated representation cannot do: every diabetic kidney complication regardless of subtype, or a rare disease whose cluster no single precoordinated concept captures. Collapsing it into one concept loses that detail. We think it can be kept in OMOP today, with mechanisms the community already supports and no fundamental change to the model, and we offer this convention for discussion.

Methods

An ICD-11 cluster joins a stem code to its extension codes: "&" attaches an axis to the stem, "/" separates one statement from the next. The axes are not all alike. Some name a second disease (etiology, manifestation), some refine the diagnosis (severity, stage, laterality, anatomy, course), and some change how it is asserted (history of, suspected, ruled out). Because they differ in kind, they cannot enter OMOP the same way, so the convention first classifies each axis, then applies the handling that fits.

Decide: what the axis is	Class	OMOP handling (CDM 5.4, no model change)
Value is another disease / stem that could stand alone (etiology, manifestation, 'due to')	Relational	Two CONDITION_OCCURRENCE rows. Precoordinate the link where a 'due to' concept exists; FACT_RELATIONSHIP only as an analytics-weak last resort.
Value is Chapter X intrinsic clinical detail that cannot stand alone (severity, stage, laterality, topology, course)	Modifier	MEASUREMENT / OBSERVATION concept in a proposed ICD-11 Modifier vocabulary (Cancer-Modifier pattern), bound to the condition via CDM 5.4 measurement_event_id / observation_event_id; patient-specific detail; definitional facts go to the vocabulary.
Value changes the assertion's operational status (history of, family history, suspected, rule out, negation)	Qualifier	Condition_status / Condition+past date / OBSERVATION; never active; never an asserted condition. We refuse to decompose this category into a diagnosis.
Residue that fits none of the above (substance / chemical, medication context)	Other	Case by case; flag for human review. Carried in the SSSOM companion regardless, so the source cluster still round-trips.

Figure 1. The Axis Type Gate: decide by what the axis is, after Chute & Elkin (1997)

The classifier is Chute and Elkin's distinction between modifiers and qualifiers(3), with a relational class and a residual class added (Figure 1). A **relational** axis carries a disease that could stand on its own; it maps to one or two conditions, below. A **modifier** is an intrinsic detail that cannot stand alone, and it comes in two kinds. Details that are definitional to the disease belong in the vocabulary as a concept relationship, recorded once, not per patient. Detail that varies by patient (severity, stage, laterality) is patient data: a standard modifier concept in MEASUREMENT or OBSERVATION, bound through the CDM 5.4 event link(4) (measurement_event_id; meas_event_field_concept_id = 1147127). A **qualifier** axis changes the status of the assertion, not its content, so it is never asserted as an active condition. Each routes to its existing OMOP home: suspected and ruled out to condition_status, history of to the

condition with a past date, family history to OBSERVATION, negation interpreted by the standard OMOP convention. The gate's job is the routing and the guarantee that none is recorded as a current condition the patient does not have.

A relational axis is precoordinated first. SNOMED is deliberately rich in "due to," "secondary to," and "associated with" concepts, so most etiology and manifestation pairs already exist as one combined standard concept, which is a single `CONDITION_OCCURRENCE` row and needs no link. Only when no combined concept exists do we decompose: store the two diagnoses as two `CONDITION_OCCURRENCE` rows and connect them in `FACT_RELATIONSHIP` with the axis as the relationship (due to, associated with, followed by, has causative agent). `FACT_RELATIONSHIP` is the only condition-to-condition link OMOP provides, since `CONDITION_OCCURRENCE` has no event field, but the community has cautioned against leaning on it because the analytic tools rarely consume it(5); it is a fallback, not the default.

OMOP's oncology model already does this for cancer: instead of proposing a concept for every diagnosis-stage-grade combination (the combinatorial explosion won't fit OMOP), it keeps the diagnosis in `CONDITION_OCCURRENCE` and stores each detail as a separate standard Cancer Modifier concept in `MEASUREMENT`, linked back to the diagnosis row(6).

Whatever class an axis falls into, the whole cluster is retained. A companion SSSOM mapping set(7) records each ICD-11-to-OMOP mapping with its predicate, justification, author, and date, plus a round-trip to the canonical ICD-11 code, so the original cluster can be rebuilt and no detail is thrown away. The one new artifact the convention needs is a small ICD-11 Modifier vocabulary of standard concepts, governed the way the Cancer Modifier vocabulary already is.

Results

One post-coordinated cluster goes into OMOP without changing the schema. A modifier axis, type 2 diabetes mellitus with diabetic chronic kidney disease, stage 3, becomes one condition plus a "stage 3" modifier concept in `MEASUREMENT` bound by the CDM 5.4 event link (Table 1). A relational axis, chronic kidney disease due to type 2 diabetes mellitus, precoordinates to one SNOMED concept where it exists and otherwise decomposes to two conditions joined by a `FACT_RELATIONSHIP` "due to" edge.

Tier	When it applies	What OMOP records
1. Precoordinate (default)	a combined SNOMED concept exists	one <code>CONDITION_OCCURRENCE</code> : <code>condition_concept_id</code> = 43531578, Chronic kidney disease due to type 2 diabetes mellitus
2. Decompose	no combined concept exists	two rows, CKD (#101) and type 2 diabetes (#102), joined in <code>FACT_RELATIONSHIP</code> : ('Due to': 101->102) plus the inverse ('Cause of': 102->101)
3. SSSOM (always)	under either tier	describe the precoordinate or decompose convention

Table 1A relational axis rendered three ways: chronic kidney disease due to type 2 diabetes mellitus

This is not hypothetical. Post-coordination is already routine in national coding: Norway's guide(8) builds the dagger-and-asterisk convention into everyday ICD-10 coding, recording an underlying disease and its manifestation as two linked codes. That is exactly the relational axis the convention handles, and data in this shape already exists, so OMOP needs a faithful place to keep it.

Conclusion

OMOP can already carry ICD-11's multi-axis detail with no change to the model. Each axis is classified by Chute and Elkin's gate and handled with a mechanism the community already supports: precoordination where a concept exists, the oncology event-link pattern for modifiers, situation concepts for qualifiers, and SSSOM for provenance and a round-trip. The one new artifact is a community-governed ICD-11 Modifier vocabulary of standard concepts, like the

Cancer Modifier vocabulary. We ask the OHDSI Vocabulary community to adopt the convention and to settle two open items: a relationship set for the relational axes, and a shared way to evaluate reliability. The takeaway is that ICD-11's granularity can become usable for observational research instead of being lost at mapping time.

References

1. World Health Organization. ICD-11 reference guide [Internet]. 2022. Available from: <https://icd.who.int/>
2. OHDSI Forums. ICD-11 / SNOMED [Internet]. 2019. Available from: <https://forums.ohdsi.org/t/icd11-snomed/8180>
3. Chute CG, Elkin PL. A clinically derived terminology: qualification to reduction. Proc AMIA Annu Fall Symp. 1997;570–4.
4. Observational Health Data Sciences and Informatics. OMOP common data model version 5.4 [Internet]. Available from: <https://ohdsi.github.io/CommonDataModel/cdm54.html>
5. OHDSI Forums. Representing laterality and general questions about post-coordination [Internet]. 2022. Available from: <https://forums.ohdsi.org/t/representing-laterality-general-questions-about-post-coordination/16872>
6. OHDSI Oncology Working Group. Oncology conventions: condition modifiers [Internet]. Available from: <https://ohdsi.github.io/OncologyWG/conventions.html>
7. Matentzoglou N, Balhoff JP, Bello SM. A simple standard for sharing ontological mappings (SSSOM). Database. 2022;2022:baac035.
8. Helsedirektoratet. Kodeveiledning 2026: regler og veiledning for klinisk koding i spesialisthelsetjenesten. Oslo: Helsedirektoratet; 2026.