

Mapping ICD-10-CA, ICD-9, and ICD-9-CM Terminologies to SNOMED CT to Support OMOP Implementation in Canada

Lisa Lix,¹ Nasiba Ahmed,^{1,2} Georgina Archbold,² Ali Beykzadeh,³ Sasha Blant,^{1,2} Lisa Flaten,⁴ Yoko Ishigami-Doyle,⁵ Alexandra Ouédraogo,⁶ Aditi Patrikar,⁷ Sandra Peterson,^{8,10} Alexandra Roine,² David Yang,⁹ Yinshan Zhao,¹⁰ Alison Park⁷

1. George & Fay Yee Centre for Healthcare Innovation, University of Manitoba, Winnipeg, MB; 2. Health Data Research Network Canada, Canada; 3. DataNB, University of New Brunswick, Fredericton, NB; 4. Manitoba Centre for Health Policy, University of Manitoba, Winnipeg, MB; 5. Health Data Nova Scotia, Dalhousie University, Halifax, NS; 6. Secure Island Data Repository, University of Prince Edward Island, Charlottetown, PE; 7. ICES, Toronto, ON, Canada; 8. Centre for Health Services and Policy Research, University of British Columbia, Vancouver, BC; 9. University of British Columbia, Vancouver, BC, Canada; 10. Population Data BC, University of British Columbia, Vancouver, BC.

Background

Health Data Research Network (HDRN) Canada is a national network of provincial, territorial and pan-Canadian data centres that supports multi-regional health data research across Canada. As part of this work, HDRN Canada is implementing OMOP across participating data centres,¹ including mapping non-standard terminologies commonly used in Canada (e.g., ICD-9, ICD-9-CM, ICD-10-CA) to OMOP standard vocabularies. The primary objective of this project was to map diagnostic codes from the [Charlson](#) and [Elixhauser](#) Comorbidity Indices to the Standard Nomenclature of Medicine - Clinical Terms (SNOMED-CT).

Methods

We identified ICD-10-CA, ICD-9-CM, and ICD-9 diagnostic codes used in the Charlson and Elixhauser Comorbidity Indices from established Canadian and international algorithms.^{2,3} A total of 7,332 diagnostic codes were identified for mapping. To support consistent and granular mappings, code lists were expanded to include all descendant codes within the ICD hierarchy. Codes were mapped and reviewed within their parent groups to ensure consistency across related concepts.

Mapping Approach. A tiered, multi-stage mapping strategy was applied based on whether the source vocabulary was available in the OHDSI ATHENA vocabulary repository. Mapping and review were conducted in pairs by analysts from multiple Canadian data centres. Discrepancies were resolved through group consensus and in consultation with OHDSI vocabulary experts at EPAM Systems, Inc.

ICD-9-CM (Available in ATHENA). For ICD-9-CM codes available in ATHENA, existing mappings to standard OMOP concepts were systematically reviewed by independent analysts for semantic accuracy, consistency, and adherence to OMOP conventions. An additional 17 legacy ICD-9-CM codes present in our internal code list were absent from ATHENA and were therefore manually mapped. These comprised 11 HIV-related codes, four tobacco abuse codes, and two codes for uterine rupture. Issues and recommended improvements were documented and submitted to EPAM vocabulary experts for review.

ICD-9 (Not available in ATHENA). ICD-9 codes were first matched to semantically equivalent ICD-9-CM codes. Where equivalence was established, ICD-9-CM mappings were reused. ICD-9 codes without a suitable equivalent were manually mapped to standard OMOP concepts.

ICD-10-CA (Not available in ATHENA). ICD-10-CA codes were first matched to ICD-10, ICD-10-CM, ICD-10-CN, and ICD-9-CM codes in ATHENA. To improve match rates, labels were harmonized by standardizing

spelling (e.g., American to Canadian/British variants) and correcting typographical and minor lexical differences.

Where matches were identified, existing OMOP mappings to standard concepts in ATHENA were leveraged. Unmatched codes were manually mapped using OHDSI's Usagi tool in conjunction with ATHENA, SNOMED CT browsers, and ICD coding manuals. All mappings were independently reviewed and discrepancies were resolved through consensus and consultation with EPAM vocabulary experts.

Validation of Charlson Comorbidity Index (CCI) Code Lists. We evaluated the SNOMED CT CCI code lists derived from our ICD-10-CA and ICD-9-CM mappings by comparing them with an established SNOMED CT-based CCI algorithm code list.⁴ This code list (hereafter the Fortin list) was first updated to align with a more current OMOP vocabulary (v20250827), which included replacement of deprecated concepts, correction of precision errors, and incorporation of updated "maps to" relationships.

Agreement was assessed at the concept level, allowing for both exact matches and partial matches in cases where one ICD code mapped to multiple standard concepts. In these cases, a match was considered present if at least one of the mapped concepts overlapped with the Fortin list. We quantified agreement using positive percent agreement from both perspectives (HDRN-to-Fortin and Fortin-to-HDRN) and calculated the Jaccard coefficient⁵ to measure similarity between the lists.

Results

Mapping Results

All 7,332 diagnostic codes (3,040 ICD-10-CA, 2,692 ICD-9-CM, and 1,600 ICD-9) from the Charlson and Elixhauser Comorbidity Indices were successfully mapped to OMOP standard concepts. For ICD-9-CM, 2,675 codes (99.4%) leveraged existing ATHENA mappings, while 17 (0.6%) required custom mapping because these legacy codes were absent from ATHENA (Table 1). Following review, 2,371 (88.1%) codes mapped to a single standard concept and 321 (11.9%) mapped to multiple standard concepts. Mapping improvements were recommended for 371 codes (13.9%).

For ICD-9, 1,257 codes (78.6%) were mapped through semantically equivalent ICD-9-CM concepts and 343 (21.4%) required custom mapping (Table 1). Most ICD-9 codes mapped to a single standard concept (1,397; 87.3%), while 203 (12.7%) required one-to-many mappings. One-to-many mappings were more common from custom-mapping codes (46.1%) than from leveraging existing mappings (3.6%).

Of the 3,040 ICD-10-CA codes, 1,925 (63.3%) leveraged existing ATHENA mappings, while 1,115 (36.7%) Canadian-specific codes required custom mapping (Table 2). Overall, 1,980 codes (65.1%) mapped to a single standard concept and 1,057 (34.8%) required one-to-many mappings, a higher proportion than observed for ICD-9 and ICD-9-CM and reflecting the complexity of Canadian-specific terminology. Custom-mapped codes were far more likely to require one-to-many mappings than auto-mapped codes (82.5% vs. 7.2%). Changes were also recommended for 268 (13.9%) auto-mapped codes.

Table 1. ICD-9 / ICD-9-CM to SNOMED CT mapping results summary.

	Total N (%)	Auto-mapped N (%)	Custom-mapped N (%)
Charlson & Elixhauser comorbidity indices codes			
ICD-9-CM			
Unique codes mapped	2692	2675	17
Unique codes unmapped	0	0	0

Before review:			
Mapped to 1 standard concept	-	2458 (91.9)	-
Mapped to 2+ standard concepts	-	217 (8.1)	-
After review:			
Mapped to 1 standard concept	2371 (88.1)	2358 (88.1)	13 (76.5)
Mapped to 2+ standard concepts	321 (11.9)	317 (11.9)	4 (23.5)
No change from ATHENA mapping	-	2304 (86.1)	-
Any change from ATHENA mapping	-	371 (13.9)	-
<i>Any suggested replacement*</i>	-	252 (9.4)	-
<i>Any suggested addition*</i>	-	155 (5.8)	-
<i>Any suggested deletion*</i>	-	36 (1.3)	-
ICD-9			
Unique codes mapped	1600	1257	343
Unique codes unmapped	0	0	0
Mapped to 1 standard concept	1397 (87.3)	1212 (96.4)	185 (53.9)
Mapped to 2+ standard concepts	203 (12.7)	45 (3.6)	158 (46.1)

*Some codes had more than one mapping change, therefore the sum of 'Any suggested replacement', 'Any suggested addition', and 'Any suggested deletion' does not equal 'Any change from previous mapping'.

Table 2. ICD-10-CA to SNOMED CT mapping results summary.

	Total N (%)	Auto-mapped N (%)	Custom-mapped N (%)
Charlson & Elixhauser comorbidity indices codes	3040	1925	1115
Unique codes mapped	3040	1925	1115
Unique codes unmapped	0	0	0
Before review:			
Mapped to 1 standard concept	-	1774 (92.2)	-
Mapped to 2+ standard concepts	-	151 (7.8)	-
After review/mapping:			
Mapped to 1 standard concept	1980 (65.1)	1786 (92.8)	195 (17.5)
Mapped to 2+ standard concepts	1057 (34.8)	139 (7.2)	920 (82.5)
No change from previous mapping	-	1657 (86.1)	-
Any change from previous mapping	-	268 (13.9)	-
<i>Any suggested replacement*</i>	-	206 (10.7)	-
<i>Any suggested addition*</i>	-	39 (2.0)	-
<i>Any suggested deletion*</i>	-	56 (2.9)	-

*Some codes had more than one mapping change, therefore the sum of 'Any suggested replacement', 'Any suggested addition', and 'Any suggested deletion' does not equal 'Any change from previous mapping'.

Charlson Comorbidity Index Code List Validation

Moderate agreement was observed between HDRN Canada ICD-10-CA and ICD-9-CM derived SNOMED CT CCI code lists and the Fortin list (Jaccard coefficient $J = 0.48$ for ICD-10-CA; 0.56 for ICD-9-CM), with

higher concordance for ICD-9-CM mappings (Table 3). Agreement increased substantially when SNOMED CT hierarchical relationships were considered, suggesting that most discrepancies were due to differences in granularity rather than true semantic disagreement. A large proportion of non-overlapping concepts were reconciled through inclusion of descendant and ancestor relationships, demonstrating the influence of hierarchical structure on observed concordance.

These findings suggest that, while direct concept overlap varies by source vocabulary, apparent differences in comorbidity definitions are substantially reduced when SNOMED CT hierarchical relationships are considered, indicating high concordance in underlying concept representation.

Table 3. Comparison of updated Fortin SNOMED CT-based algorithm for the Charlson Comorbidity Index vs. HDRN Canada direct mapping of Quan ICD-9-CM and ICD-10-CA Charlson Comorbidity Index diagnosis code sets to standard concepts, by condition.

	Updated Fortin List	HDRN Canada Quan ICD-10-CA Mapping	HDRN Canada Quan ICD-9-CM Mapping
Unique ICD codes	N/A	1825	1277
Total Concept IDs	1515	883	975
Unique Concept IDs	1510	875	960
Unique Concept IDs by condition			
Myocardial infraction	24	11	10
Congestive heart failure	41	16	26
Peripheral vascular disease	58	22	37
Cerebrovascular disease	128	67	65
Dementia	40	20	19
Chronic pulmonary disease	91	50	36
Rheumatic disease	57	24	10
Peptic ulcer disease	112	36	106
Mild liver disease	43	28	23
Diabetes, without chronic complications	26	23	20
Hemiplegia or paraplegia	46	33	36
Renal disease	33	28	32
Malignancy, except skin neoplasms	682	432	499
Moderate or severe liver disease	16	12	7
Metastatic solid tumor	45	29	34
AIDS/HIV	11	1	6

Conclusion

This work substantially expands OMOP vocabulary coverage for Canadian clinical terminologies, particularly ICD-10-CA, which has not previously been comprehensively mapped. The findings also show that ICD-10-CA–derived SNOMED CT concept sets are broadly consistent with established SNOMED-based comorbidity definitions, with agreement improving substantially when SNOMED CT hierarchical relationships are considered.

A limitation is that validation was limited to code list concordance and did not assess phenotype performance in OMOP data. Future work includes evaluating the impact of these mappings on cohort identification and assessment of comorbidity burden in real-world datasets. We also plan to further expand and validate Canadian terminology mappings and contribute them to ATHENA to promote

adoption of validated Canadian mappings in the OHDSI community.

References

1. Lix L, et al. Implementation of the OMOP Common Data Model in a Multi-Site Canadian Network. *2025 Global OHDSI Symposium Collaborator Showcase*. Available from: <https://www.ohdsi.org/2025showcase-138/>
2. Lix L, Smith M, Pitz M, et al. Cancer Data Linkage in Manitoba: Expanding the Infrastructure for Research. Winnipeg (MB): Manitoba Centre for Health Policy; 2016. Available from: http://mchp-appserv.cpe.umanitoba.ca/reference/Candata_web_final.pdf
3. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43(11):1130-9. doi:10.1097/01.mlr.0000182534.19832.83. PMID: 16224307
4. Fortin SP, Reps J, Ryan P. Adaptation and validation of a coding algorithm for the Charlson Comorbidity Index in administrative claims data using the SNOMED CT standardized vocabulary. *BMC Med Inform Decis Mak*. 2022;22:261. doi: 10.1186/s12911-023-02205-4. PMID: 36207711; PMCID: PMC9541054.
5. Podani J. The wonder of the Jaccard coefficient: From alpine floras to bipartite networks. *Flora Mediterranea*. 2021 Jul;31:105-23.