AI-Assisted Precision and Granularity Assessment of In-Source Japanese Diagnosis Mappings

Background: Standardizing data from certain geographical regions, such as Japan, remains challenging despite their widespread representation. Current standardization efforts focus on high-concept drug mapping [1] and unified terminology representation, including the work of the Medical Information System (MEDIS) [2]. MEDIS distributions include vendor-provided mappings to ICD-10. However, we hypothesized a mismatch: MEDIS terms likely capture more granular clinical entities than ICD-10's broad classification categories can represent. This work lays the ground for further efforts in Japanese healthcare data standardization, with direct implications for improving Extract, Transform, Load (ETL) processes and enhancing the reliability of multi-national observational research.

Methods: We analyzed MEDIS source terms and their native ICD-10 mappings. We used English translation present in the source for all the analyses. To assess ICD-10 mapping precision, we applied a vectorizer-embedder algorithm with Euclidean distance measurement, previously developed and validated for regimen mappings to OHDSI vocabularies [3]. The algorithm generated mapping suggestions between source terms and ICD-10 codes, which Subject Matter Experts (SMEs) evaluated to identify suboptimal native mappings. SMEs were medical doctors with experience working with OHDSI vocabularies and their mappings.

To evaluate mapping granularity and its impact on clinical specificity, we compared two distinct mapping pathways. Pathway 1 (Direct mapping): SMEs manually mapped MEDIS source terms directly to SNOMED CT and other Standard OHDSI vocabularies, preserving the clinical granularity inherent in the source terminology. Pathway 2 (ICD-10-mediated mapping): We computationally mapped MEDIS terms through their native ICD-10 codes to Standard vocabularies using the OHDSI concept_relationship table, simulating the current standard workflow. Our primary outcome metric quantified information loss: the proportion of mappings that resulted in broader (less specific) or more granular (more specific) Standard concepts in Pathway 2 relative to Pathway 1. This comparison directly measures whether the ICD-10 intermediate step introduces systematic loss of clinical specificity. Figure 1 presents a schematic overview illustrating all possible granularity outcomes assessed by this methodology. All analyses utilized OHDSI vocabularies version 5.0 (27-AUG-2025).

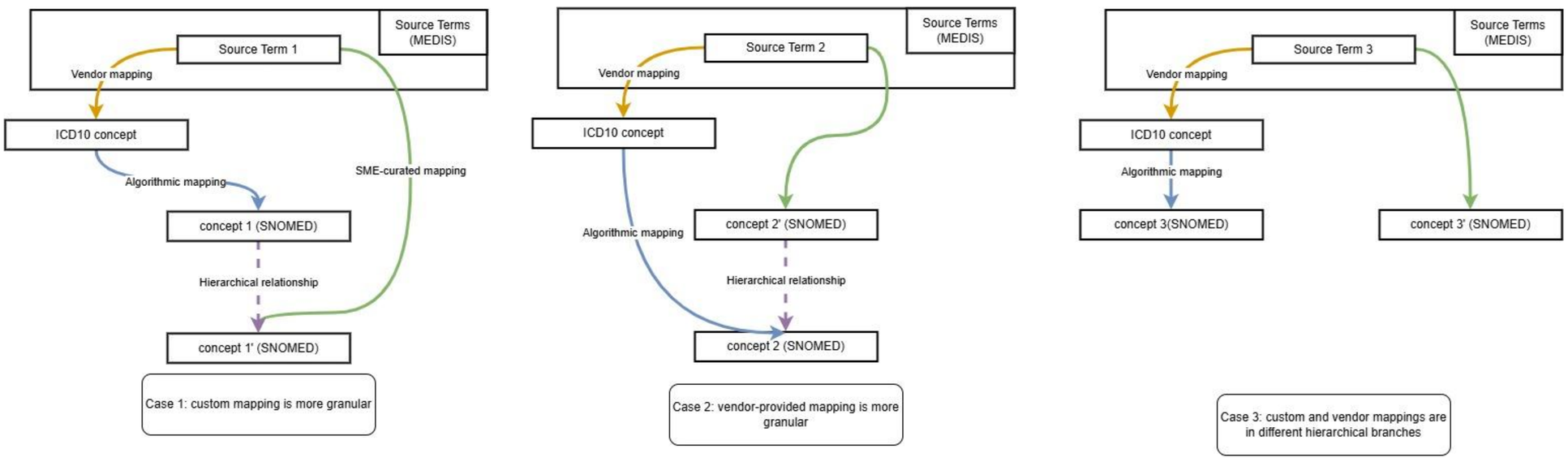

Figure 1 Possible outcomes of mapping granularity comparison

Results: Of 27,563 ICD-10 mappings evaluated, 8,420 (30.5%) aligned with Large Language Model (LLM) generated suggestions. For the 19,143 mappings that did not align, SMEs manually compared the native and LLM-generated mappings to assess mapping quality. We randomly selected 3,000 of these discordant mappings (15.7%) for human review.

Results (cont.): SME review found that 2,671 (89%) of the source mappings were optimal, while 329 (11%) were suboptimal. Applying this proportion to all discordant mappings suggests an overall suboptimal mapping rate of approximately 2,04% in the source vocabulary (Figure 2).

In the second phase, we compared the level of detail between the existing vendor mapping and an SME-developed mapping. We assessed whether each mapping used broader or more specific diagnostic codes by assessing their position within the hierarchy of OHDSI vocabularies. We found that 6,964 (25.3%) codes were mapped to a broader (more general) concept in the vendor mapping compared to the custom mapping, while 2,974 (10.8%) codes were mapped to a more specific concept in the vendor mapping (Figure 3). Overall, the vendor mapping more frequently favored broader, more general codes over specific ones.

Discussion and conclusion: Our hybrid computational-expert approach identified a low rate of technically suboptimal mappings (2.04%). However, examining semantic granularity revealed that over one-quarter of ICD-10-mediated mappings resulted in clinically broader concepts than the original MEDIS source terms, representing substantial information loss.

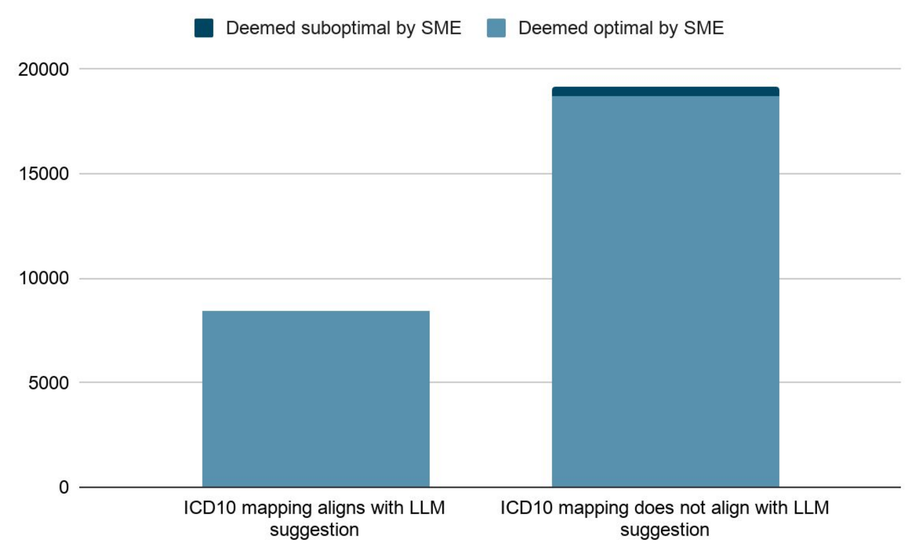

Figure 2 Proportions of vendor-provided mappings to ICD10 aligning with the LLM-suggested mappings

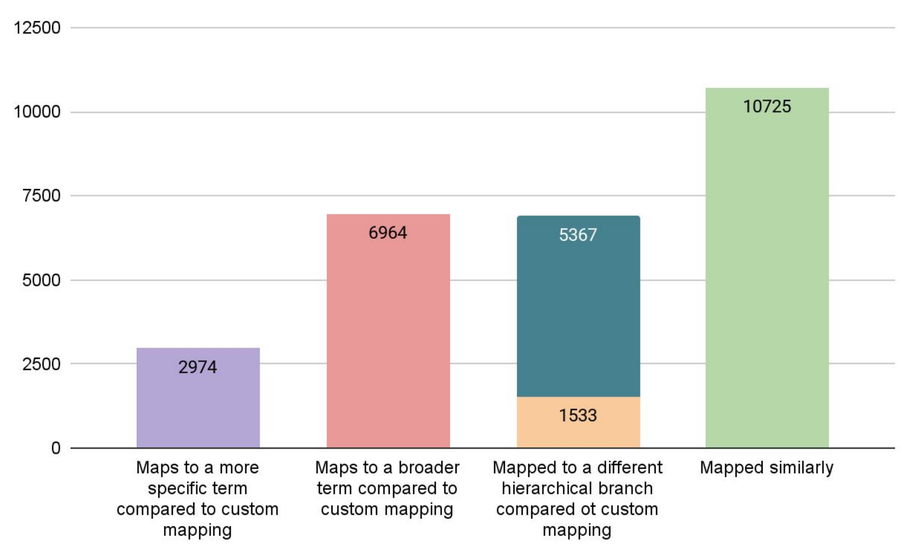

Figure 3 Hierarchical relationships of mappings to SNOMED (blue - text similarity to source term is higher in custom mapping, yellow - text similarity to source term is higher in vendor mappings)

Discussion and conclusion (cont.): When SME-developed and vendor-derived SNOMED mappings diverged hierarchically, we used computed text similarity scores as an objective arbitration metric. In 5,367 of 6,900 cases (77.8%), SME-developed mappings demonstrated superior semantic alignment with source terms.

These results underscore the need for standardized methodologies to systematically evaluate mapping quality and granularity across vocabularies. Future development of comprehensive quality frameworks incorporating multiple evaluation metrics and AI-assisted validation tools could enable more reproducible, scalable mapping processes, ultimately strengthening the reliability of evidence generated from global observational health data networks.

Tetiana Orlova, Mikita Salavei, Maria Khitrun, Tatsiana Skuhareuskaya

References: 1 2 3