Comparing Human and AI-Derived DRG Severity of Illness Using OMOP CDM: A Belgian APR-DRG Case Study

Dr. Peter Heirman[1], Prof. Dr. Philippe Kolh[1], Prof. Dr. Ittoo Ashwin[2]

[1]CHU Liège

[2]Université de Liège (ULiège)

1. Background

In Belgium, hospital financing and casemix evaluation are partially determined by care complexity measured through All Patient Refined Diagnosis Related Groups (APR-DRGs), derived from ICD-10-CM/PCS coding registered in the Minimal Hospital Dataset (MZG/RHM). Accurate assignment of APR-DRGs and their associated Severity of Illness (SOI) subclasses depends on comprehensive and consistent clinical coding. However, manual coding is labor-intensive, subject to variability, and increasingly challenged by workforce constraints.[1,2]

Recent advances in natural language processing (NLP) and artificial intelligence (AI) have enabled the development of automated and AI-assisted clinical coding systems. These systems have demonstrated the ability to support human coders by detecting diagnoses from unstructured clinical text and improving coding completeness.[3–6] Since DRG and SOI assignment is sensitive to secondary diagnoses and comorbidities, AI-assisted coding may impact casemix measurement and downstream reimbursement.[7]

At the same time, observational research infrastructures such as the OMOP Common Data Model (CDM) lack a dedicated DRG/SOI table. This complicates systematic comparison of multiple DRG and SOI derivations (e.g., human vs AI, different grouper versions) within a single analytical framework.[8]

2. Methods

We conducted a retrospective analysis of 18,515 inpatient hospitalizations from the first semester of 2025 at CHU de Liège. ICD-10-CM/PCS codes produced by professional human coders were compared with codes automatically generated by the Solventum 360 AI coding model. APR-DRG base groups and SOI levels were independently derived from both coding sources using the same grouper logic.

Hospitalizations in which the AI-derived SOI exceeded the human-derived SOI were identified. A subset of 2,563 such cases was reviewed by a panel of 25 professional coders. Reviewers evaluated whether AI-suggested additional diagnoses were clinically valid or represented hallucinations, based on the full medical record and original human coding.

To enable structured comparison, APR-DRG and SOI results were represented in OMOP CDM using the `OBSERVATION` table, linked to the hospitalization via `visit_occurrence_id`. Separate observations were created for human-derived and AI-derived groupings, with explicit provenance indicators for derivation method and grouping context. This approach allowed multiple DRG and SOI results to coexist per hospital stay without overwriting source data.

3. Results

Among the 18,515 analyzed hospitalizations, 10,312 stays showed identical APR-DRG assignments between human and AI coding. In the selected subset of 2,563 hospitalizations with higher AI-derived SOI, expert review demonstrated that a substantial proportion of AI-suggested diagnoses were clinically valid and had been omitted during initial manual coding.

After expert validation, the corrected impact corresponded to a net gain equivalent to 7.67 hospital beds, representing a 15% increase relative to the mean length of stay of 5.79 days within this cohort. The observed effect was primarily attributable to additional secondary diagnoses detected by the AI model that influenced SOI assignment.

The OMOP-based representation enabled direct comparison of DRG and SOI outcomes across derivation methods, preserving provenance and supporting downstream analyses of coding variability, severity shifts, and grouping stability.

4. Conclusions

Fully autonomous AI-based coding for routine hospitalizations is not yet sufficiently reliable to replace human coders. However, a hybrid human–AI approach offers clear advantages. Human expertise mitigates AI-related hallucinations and errors, while AI systems effectively identify missed diagnoses that materially affect DRG and SOI assignment.

Representing multiple DRG and SOI derivations as visit-linked observations within OMOP CDM provides a scalable, transparent, and reproducible framework for evaluating coding strategies, AI models, and grouper versions, and supports future transitions including ICD-11-based grouping.

5. References

References

[1] Otero Varela L., Doktorchik C., Wiebe N., *et al.* (2022). *International Classification of Diseases clinical coding training: an international survey.* Health Inf. Manag. 2024-05, Vol.53 (2), p.68-75.

[2] Geissler A., Quentin W., Busse R. (2015). *Heterogeneity of European DRG systems and potentials for a common EuroDRG system.* Int. J. Health Policy Manag. 2015-05, Vol.4 (5), p.319-320

[3] Liu Q., Song P., Huang H., *et al.* (2021). *DRG grouping by machine learning: from expert-oriented to data-based method.* BMC Med. Inform. Decis. Mak. 2021-11, Vol.21 (1), p.1-11, Article 312

[4] Liu J., Capurro D., Nguyen A., *et al.* (2021). *Early prediction of diagnostic-related groups and hospital cost by processing clinical notes.* NPJ Digit. Med. 2021-07, Vol.4 (1), p.103-8, Article 103

[5] Gartner D., Kolisch R., Neill D.B., Padman R. (2015). *Machine learning approaches for early DRG classification and resource allocation.* J. Health Inform. 2015-09, Vol.27 (4), p.718-734

[6] Chang Y.-C., Chen Y.-H., Lin C.-Y., *et al.* (2024). *Evaluating an NLP-driven AI-assisted ICD-10-CM coding system for DRG assignment.* J. Med. Internet Res. 2024-09, Vol.26 (12), p.e58278

[7] Lu A.-T., Wang Y.-C., Tsai M.-S., *et al.* (2025). *Application of AI-assisted clinical coding for DRG-based hospital reimbursement.* JMIR Hum. Factors 2025-02, Vol.12, p.e59961-e59961

[8] Observational Health Data Sciences and Informatics (OHDSI). (2021). *The Book of OHDSI.* Chapter: Common Data Model and Administrative Data Representation.