

Standardizing Bedside Nursing Narratives for Interoperable Research: An Episode-Based Nursing Observation Extension to the OMOP CDM

Sujin Gan¹, Chungsoo Kim¹², Youngjin Lee³, Rae Woong Park¹²

¹Department of Biomedical Sciences, Ajou University School of Medicine, Suwon, Republic of Korea,

²Department of Biomedical Informatics, Ajou University, Suwon, Korea; ³College of Nursing, Ajou University, Suwon, Republic of Korea

Background

Nurses continuously observe patients at the bedside and document subtle behavioral, functional, emotional, and physiological changes. These narrative records reflect clinical reasoning, treatment responses, and early signs of deterioration. The CONCERN early warning system showed that patterns in nursing surveillance documentation can identify patient deterioration earlier than conventional approaches.¹ However, most nursing documentation remains unstructured free text with limited standardization, and documentation practices vary across institutions. Although the OMOP Common Data Model (CDM) supports standardized observational research across the OHDSI network, these bedside data remain underrepresented because they are highly contextual and longitudinal.² We therefore designed a nursing observation extension for the OMOP CDM and developed a scalable hybrid expert–LLM pipeline to transform statement-level nursing narratives into standardized, time-resolved semantic concepts.

Methods

We used the electronic health record database of Ajou University Medical Center (AUMC), a tertiary teaching hospital in South Korea, covering January 2023 to December 2024. The study population included adult inpatients (≥ 18 years) admitted to 36 study wards with hospital stays of at least 24 hours. Obstetric, neonatal, and hospice wards, as well as patients with do-not-resuscitate (DNR) orders, were excluded. As a proof-of-concept use case, we processed flowsheet-based bedside nursing records documented within 24 hours before rapid response team (RRT) activation. Records included four statement subtypes: quantitative observation, structured assessment, nursing action, and test/evaluation.

We constructed a feature dictionary consisting of 39 nursing observation features: 35 adapted from the CONCERN framework and four additional features commonly documented in nursing flowsheets but not represented in standard OMOP domains (physician notification, oral intake/diet, sleep pattern, and restraint use).¹ A gold standard dataset was established through two rounds of expert annotation involving 5 clinical nurses. Annotation included independent review, group-level consensus, and final reconciliation. Inter-rater agreement was assessed using Cohen’s kappa.

Statements were labeled using a hybrid pipeline consisting of (i) rule-based normalized exact matching and regex detection and (ii) two-stage LLM classification for unmatched statements. The LLM workflow first identified clinically meaningful statements and then assigned multi-label nursing features. We used Qwen 3.5 (31B), an open-source model with temperature set to 0.0. LLM-based labels were further refined through expert review and deterministic post-correction rules. Performance was evaluated against the gold standard using precision, recall, and F1 score.

We designed two OMOP extension tables: NURSING_EPISODE and NURSING_OBSERVATION (Figure 1). NURSING_EPISODE stores one row per hospitalization episode linked to PERSON, with RRT activation recorded as an episode-level observation. NURSING_OBSERVATION stores one row per statement–

feature pair in long format with minute-level timestamps (Table 1). Existing SNOMED CT and LOINC concepts were reused whenever possible, and custom concepts were created only for unmapped nursing constructs. Each feature was assigned to a standard OMOP domain, enabling temporal linkage with CONDITION, DRUG, MEASUREMENT, and PROCEDURE tables.

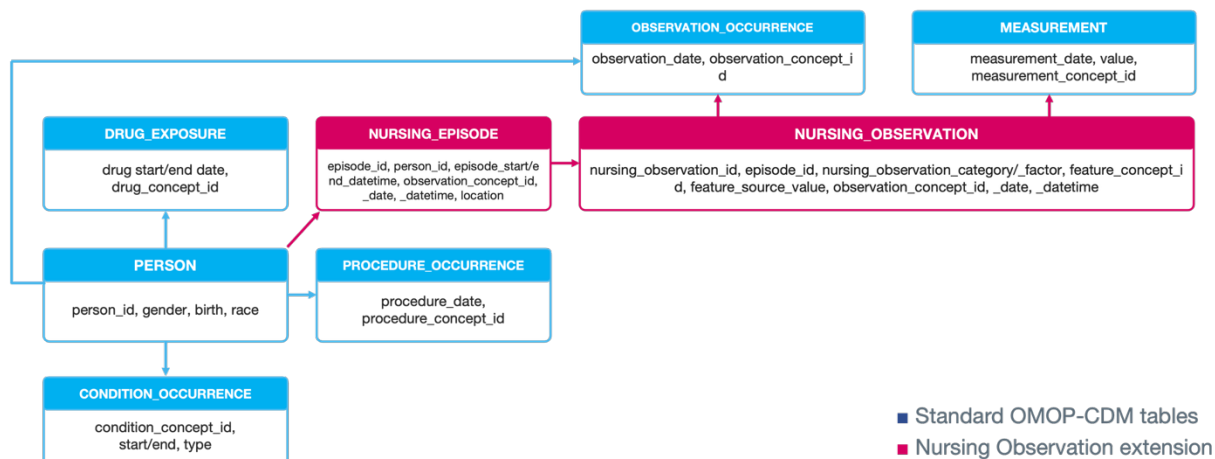


Figure 1. Overview of the proposed nursing observation extension within the OMOP-CDM.

Table 1. Key fields of the proposed nursing observation extension tables.

Table	Key fields
NURSING_EPISODE	episode_id (PK); person_id → PERSON; episode_start/end_datetime; admission/discharge ward; observation_concept_id (ex. RRT event); observation_datetime
NURSING_OBSERVATION	nursing_observation_id (PK); episode_id → NURSING_EPISODE; person_id; observation_datetime; nursing_observation_category/factor; feature_concept_id; feature_source_value (original text); flowsheet subtype; flowsheet_record_id (audit trail)

Results

RRT-associated nursing records from 3,108 patients were transformed into 3,573 episode rows and 364,767 raw nursing observation rows. After preprocessing and removal of non-clinical content, 82,339 statements were mapped to semantic nursing features, producing 139,799 longitudinal nursing observation records after multi-label expansion.

Inter-annotator agreement for the gold standard was high (Cohen's $\kappa = 0.82$), with 96.4% agreement between annotation rounds. Against the expert gold standard, the deployed hybrid pipeline achieved a micro-F1 score of 0.86 and a macro-F1 score of 0.88 across features, with precision approaching 1.00 for most features. Objective physiological features showed consistently high performance, whereas subjective or judgment-dependent nursing assessments showed lower recall and F1 scores (Table 2).

Table 2. Feature-level performance of the hybrid nursing statement classification pipeline.

Feature	Precision	Recall	F1	Feature	Precision	Recall	F1
Improper renal function	1.00	1.00	1.00	General concern	1.00	0.75	0.86
Abnormal temperature	1.00	1.00	1.00	Chest pain	1.00	0.75	0.86
Violence gesture	1.00	0.91	0.95	Abdominal pain	1.00	0.75	0.86
Abnormal heart rhythm	1.00	0.89	0.94	Abnormal mental state	1.00	0.75	0.86
Abnormal rate, rhythm, depth	1.00	0.88	0.93	Monitoring	1.00	0.75	0.86
Diagnosis related to infection	1.00	0.88	0.93	Headache	1.00	0.67	0.80
Fluid volume alteration	1.00	0.85	0.92	Fall risk	1.00	0.62	0.77
Deficit of circulation	1.00	0.83	0.91	Mood disorder	1.00	0.61	0.76
Communication problem	1.00	0.79	0.88	Pain level	0.96	0.57	0.72

Most features were mapped to standard OMOP domains using primarily LOINC and SNOMED CT vocabularies, while custom concepts were created for nursing constructs without standard equivalents. Vital-sign frequency concepts accounted for the largest proportion of observations, including heart rate (24,261 rows; 17.4%), respiratory rate (23,894; 17.1%), blood pressure (21,454; 15.3%), and temperature (21,058; 15.1%). The most frequent patient-status concepts were fall risk (4,192; 3.0%), pain level (2,342; 1.7%), and abnormal mental state (1,706; 1.2%).

Conclusion

The proposed nursing observation extension enables nursing semantics to be represented as OHDSI-compatible concepts that can be directly linked with existing OMOP domains. As a result, bedside nursing narratives can be analyzed together with diagnoses, laboratory results, medications, vital signs, and structured clinical assessments within a unified patient representation. The minute-level longitudinal structure also supports temporal and time-series analyses of patient status and clinical deterioration trajectories. This approach may facilitate future research on clinical deterioration, readmission, quality of care, and nursing workforce outcomes. In addition, the extension was designed to accommodate both CONCERN-derived concepts and future nursing feature sets without major schema modification. Several limitations should be noted. This study was conducted at a single tertiary center, and external validation across multiple institutions is needed to assess generalizability. Because the source corpus consisted primarily of Korean-language clinical text, the cross-linguistic robustness of the LLM pipeline remains to be validated in multilingual settings. In addition, the RRT cohort represented a deterioration-enriched population; therefore, further validation using broader general nursing documentation is needed.

References

1. Rossetti SC, et al. Real-time surveillance system for patient deterioration: a pragmatic cluster-randomized controlled trial. *Nat Med*. 2025. doi:10.1038/s41591-025-03609-7
2. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;216:574-8. PMID: 26262116