Protocol for finding supplemental oxygen data in electronic health record (EHR) flowsheets for inclusion in the OMOP ETL

Tanner Zhang, Steven Miller, Michael Cook, Alan Coltri, Zachary Wang, Paul Nagy, Justin M Rucci, Galina Lozinski, Matthew Robinson

Background

Observational research for hospitalized patients with respiratory diseases, including clinical prediction, description of clinical trajectory, and causal inference for estimating drug efficacy, requires an accurate longitudinal description of the modality and magnitude of supplemental oxygen support provided.[1, 2, 3] Description of supplemental oxygen delivery, measured by oxygen flow rate or fraction of inspired oxygen (FiO_2) and modality of oxygen delivery are the most important components[2]. Clinicians use supplemental oxygen data recorded within electronic health record (EHR) flowsheets for their clinical decision making. Incorporating flowsheet data from EHRs into the OMOP ETL has been a challenge due to the high volume of entries and high variability of facility-specific customization. We propose a standardized and reproducible protocol to identify the flowsheet elements of interest, map them to their appropriate OMOP concept IDs, and load them into the OMOP Common Data Model (CDM).

Methods

Flowsheet identification

Identifying flowsheet entries containing the data of interest proceeded in a series of standard steps: 1) Perform wildcard search of flowsheet names, display names, and template names to identify initial candidates for oxygen flow, FiO_2, and oxygen device. 2) Perform wildcard search of flowsheet values among flowsheet entries not identified as initial candidates to identify flowsheet entries containing oxygen device data with nonintuitive flowsheet names. 3) Manually inspect distinct template names to identify and remove entries associated with anesthesia records. 4) Create a list of included template and flow measurement identification numbers assembled from the prior steps. 5) Extract flowsheet entries that have both an included template identification number and included measurement identification number.

Semantic mapping

The master list of included respiratory flowsheet entries was compiled from our EHR. We searched for each string in the online Athena database (athena.ohdsi.org), limiting our search by domain (device), vocabulary (SNOMED), and concept (Standard). We matched the closest code to the source data and validated the selection by ensuring the code fits within an appropriate hierarchy (i.e., respiratory equipment) and belongs to the right OMOP domain. The SNOMED codes and corresponding OMOP concept ids were then noted and used to define the mapping of clinical concepts extracted from the EHR to OMOP concept ids. These codes were reviewed and compared to established work from the National Covid Cohort Collaborative(N3C)[4]. As in N3C, concepts that did not have a clear corresponding match in either SNOMED or OMOP were assigned for custom mappings.

Utilization of OMOP CDM tables

Current OMOP CDM tables use the device_exposure table to capture the information about a patient's

exposures to medical instruments and foreign objects for diagnostic or therapeutic purposes; because the measurement table is used for structured values documented for a patient's clinical course, it would be essential to utilize the measurement table to persist more granular data including vent modes, volume and other settings for patients' respiratory support. This would allow for the device_exposure table to be exclusively used to document quantities and time, preserving the specificity of the two tables already established within the OMOP CDM.

To balance granularity and conciseness of entries, the respiratory support data collected would only be documented as multiple entries when the parameters for the ventilator settings change during the patient's clinical course or if the patient is checked multiple times to be constantly in room air. If these factors did not change throughout the patient's clinical course, the device and sequential records for 'room air' in the device_exposure table would be coalesced and merged, and the start time and end time modified correspondingly to represent the complete duration of the device. The time and patient identifier will be used to keep the linkage between records in the device_exposure table and the measurement table.

Results

Oxygen delivery device

Initial query of the flowsheet table yielded 18 distinct measurement names for oxygen delivery with matches to the search query. Manual inspection of distinct template names revealed templates used for anesthesia charting by searching for abbreviations including "AN" (anesthesia), "OR" (operating room), and "PACU" (post-anesthesia care unit). Some measurement entries were present in both included and excluded templates. The resultant measurement values consisted of oxygen delivery devices, such as "Heated High Flow." Distinct measurement values were coalesced when differing source oxygen device descriptions were mapped to the same SNOMED concept. There were 5 entries for which there was no available SNOMED standard terminology. Room air as an oxygen device was assigned the custom OMOP concept ID 2004208005 according to draft guidance from N3C. Other oxygen devices for which no appropriate SNOMED terminology was available were assigned the custom OMOP concept ID 2004208004, but the individual constituents of other oxygen devices were common and represented divergent respiratory support modalities.

Oxygen flow and FiO_2

Wildcard searching using terms "oxygen flow rate," "l/min", "high flow rate," "o2 flowrate," and "o2 therapy flow rate" identified entries with relevant data for oxygen flow. Semantic mapping identified SNOMED term 427081008 "delivered oxygen flow rate" and LOINC term 3151-8 "inhaled oxygen flow rate" as potential terminology for oxygen flow. The SNOMED term was chosen as the LOINC term 3151-8 defines discrete values of 1-6 liters/minute which would inadequately categorize patients on high-flow nasal cannula therapy. Review of wildcard matches for "fio2" yielded appropriate results in almost all cases. Semantic mapping identified SNOMED term 250774007, "inspired oxygen concentration" as an appropriate concept with no other close contenders.

Conclusion

A standardized protocol for identification of flowsheet entries simplifies the daunting process of including flowsheet data in the OMOP ETL. The OMOP CDM is missing standard codes for several commonly utilized methods of invasive and non-invasive respiratory support, which leaves an opportunity for incorporation

of standardized codes into the CDM to permit outcomes tracking for clinical interventions.

References/Citations

1. Seto T, Sung L, Posada JD, Desai P, Weber S, Datta S. Integrating Flowsheet Data in OMOP Common Data Model for Clinical Research. arXiv; 2021. p. arXiv:2109.08235.
2. Hripcsak, G., Schuemie, M. J., Madigan, D., Ryan, P. B., & Suchard, M. A. (2021). Drawing reproducible conclusions from observational clinical data with OHDSI. *Yearbook of medical informatics*, *30*(01), 283-289.
3. Mosier, J., & Subbian, V. (2021). Phenotyping COVID-19 Patients by Ventilation Therapy: Data Quality Challenges and Cohort Characterization. *Stud Health Technol Inform*, 198-202
4. https://github.com/National-COVID-Cohort-Collaborative/Phenotype_Data_Acquisition/wiki/New-Data-Element:-O2-Device-(Tier-1)