Hierarchical clustering of microbial resistance profiles and ventilation protocols using the oncology extension

Jared Houghtaling[a], Frederic Jung[a], Ankur Krishnan[b], Marc Padros Goossens[b], Frank Leus[b], Lauren Maxwell[b], Tom Feusels[a] and Freija Descamps[a]

[a] edenceHealth NV
[b] European Clinical Research Alliance on Infectious Diseases (Ecraid)

Background:

Ecraid is, together with edenceHealth, in the process of transforming data from a perpetual observational study (POS) focused on ventilator-associated pneumonia (VAP) to the OMOP Common Data Model (CDM) as part of the European Health Data and Evidence Network (EHDEN). VAP, often caused by multi-drug resistant bacteria, is a frequent complication of mechanically-ventilated ICU patients. Distinguishing between bacterial colonization of the respiratory tract and infection of the lung tissue is challenging; such a challenge leads to unnecessary use of broad-spectrum antibiotics in many patients, which in turn produces antimicrobial resistance (AMR). The anticipated dataset - in OMOP CDM format - will enable powerful standardized analytics and participation in international federated studies on prevention, diagnosis and treatment of VAP and hospital-associated pneumonia (HAP). It will also serve as a participant-level data source that can be used by the E-CDC and national surveillance authorities to improve AMR surveillance and Antimicrobial Stewardship (AMS) in the EU **[1]**.

POS-VAP data is both complex and unique relative to other observational health sources due to its strong inter-event dependencies. For example, a patient's sputum sample might be collected and cultured to identify microorganism(s). Multiple microorganisms might be isolated in the sample and each tested for their specific antimicrobial susceptibility, after which microbiological identification and antimicrobial susceptibility testing could then be performed to determine a microbiological cure. These types of events provide limited information on their own, but when linked together properly, they can lend insights into the unique microbial state of a given patient. Similarly, VAP onset is determined based on patients' invasive mechanical ventilation status and presence of clinical signs of pneumonia, while a clinical cure is determined on day 10 by clinical signs of pneumonia. VAP onset and clinical cure episodes contain imaging procedures and findings, as well as any clinical signs of pneumonia. Additionally, it is important to link each patient's microbiological identification and microbiological cure episodes, as well as their VAP onset and clinical cure episodes; these relationships can provide significant clinical insights into disease diagnoses, presentation, progression, treatment, and prognosis.

Joining events together in this fashion is a task also shared by oncology-focused OMOP transforms in the context of treatment regimens and disease progression, which is why we have chosen to take advantage of the oncology extension available in OMOP CDM 5.4. The work presented here, thus, describes a core element of the ongoing data harmonization process: linking observational concepts together in an hierarchical manner across domains. We will describe our technical approach to this clustering challenge, discussing advantages and limitations, and present suggestions for expanding the oncology extension tables to accommodate diverse types of hierarchically structured data in the future.

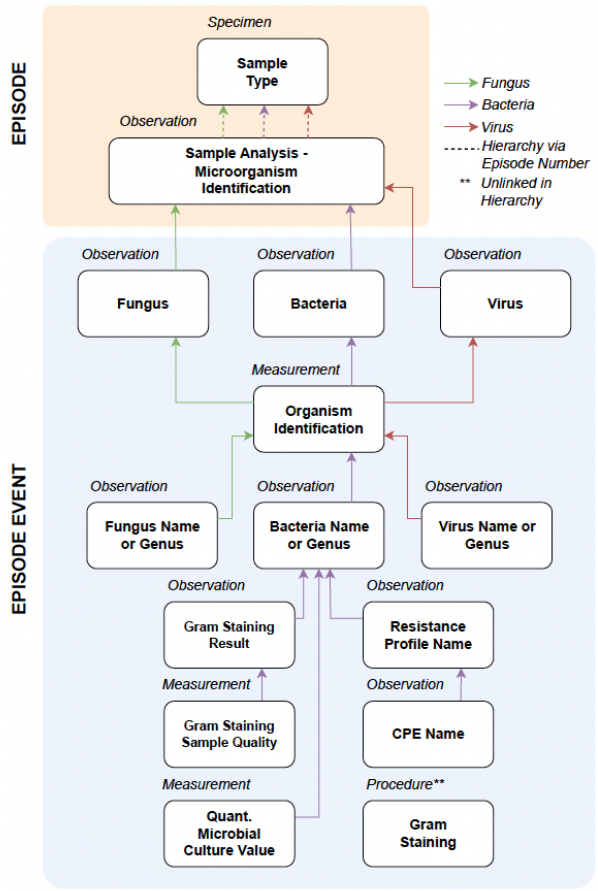

Figure 1. Hierarchical structure and linking of events related to microorganisms and their various attributes and antibiotic resistance profiles. Note that each block may represent more than one event per hierarchical tree (e.g. a single bacterial species might have more than one resistance profile, but each of those profiles will be linked to that specific bacteria's name or genus, and all will exist in the EPISODE EVENT table for that given sample).

Methods:

We have created hierarchical clusters for two different subsets of the POS-VAP data: (1) samples collected and measurements performed related to antimicrobial resistance, and (2)

protocols implemented for invasive mechanical ventilation and onset of VAP as per clinical criteria for pneumonia. We linked events in each cluster unidirectionally from child to great-grandparent using the *_event_id*, and **_event_field_concept_id* fields in the MEASUREMENT and OBSERVATION tables. Moreover, we created nested episodes for each type of cluster, and linked those episodes to all events in the tree using the EPISODE and EPISODE_EVENT tables (see **Figure 1, 2**). Clusters themselves are isolated according to patient and date; a single patient can have multiple clusters related to different ICU admissions.

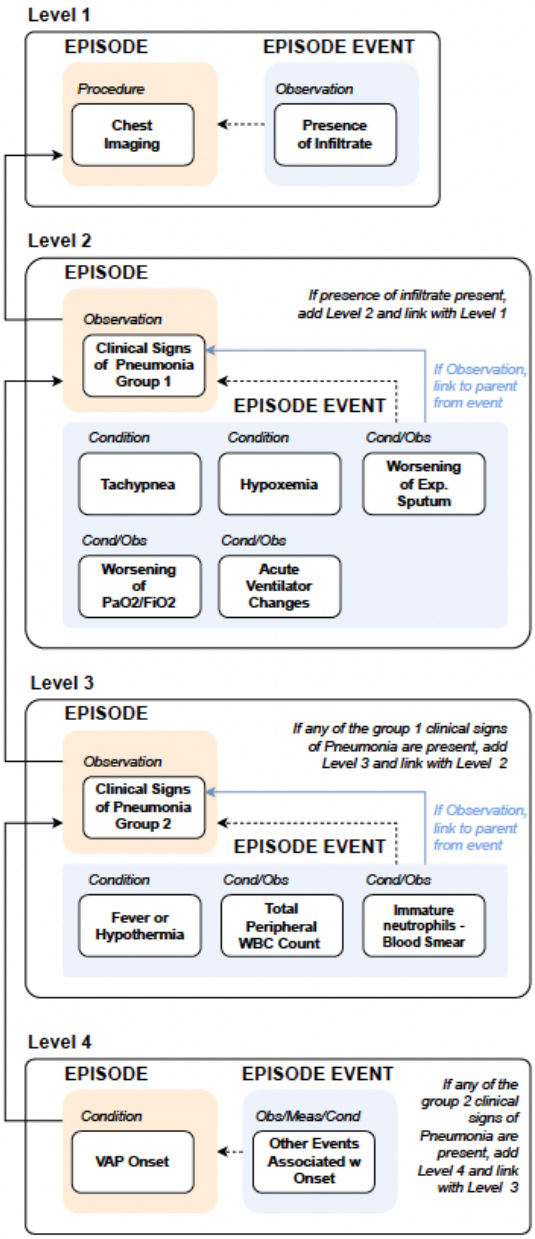

Figure 2: Hierarchical structure and linking of events related to VAP onset and relevant clinical criteria deduced through chest imaging. Note that each block may represent more than one event per hierarchical tree.

The intended analytical functionality from these clusters is two-fold: (1) we expect they will enable better understanding of broader patient presentations, and (2) we will utilize them to capture relation-based covariates (e.g. quantity and type of child AMR for a parent bacterial species), as well as context-based covariates (e.g. the combination of fungi and bacteria present in a given sputum sample as well as their combinatory AMR profile) in downstream modeling efforts. One future research aim we hope to address with the help of this functionality is to improve definitions for clinical and microbiological cures in VAP and similar infectious syndromes so they may serve as validated clinical endpoints in trials evaluating new, effective antimicrobial agents in severe infections.

Results:

We have successfully implemented the logic described above into our Extract, Transform, and Load (ETL) processes that convert VAP registry data to OMOP CDM format. More specifically, we linked the events together using patient and date information once they had been transformed into the STEM table, but prior to their transformation to the OMOP relational tables. One major advantage of implementing this logic is that it enables us to harness the relational structure embedded within the source data for OMOP-specific analyses. The oncology extension also makes it possible to quickly extract and investigate particular microorganism profiles or ventilation protocols via their associated episodes and episode events. One disadvantage, or perhaps limitation, of the approach is that the oncology extension is not designed for capturing these types of information. For example, we chose the EPISODE domain standard concepts ‘Disease Episode’ (32533) and ‘Treatment Cycle’ (32532) for the microorganism and ventilator hierarchies, respectively, but neither concept adequately describes the types of episodes we capture with the source data. Moreover, we designed the hierarchies such that each concept has only one parent (see **Figure 1**); in reality, the concepts have multiple ancestor events. In order to address this later challenge, we are in the process of designing an extension inspired by the OMOP vocabularies structure, with a CONCEPT-ANCESTOR-like table containing hierarchical relationships and degrees of separation.

Conclusion:

Creating references or associations between observational events within the OMOP CDM is a controversial topic **[2]**. However, it is generally accepted that there are instances where observational events stray from being patient-centric to being interdependent, such as biological relationships between patients (e.g. mother and fetus), or in our case, characteristics related to a particular microbe within a patient sample. While the FACT_RELATIONSHIP table has existed for many years within the OMOP CDM, and despite impassioned efforts to promote its use, its integration across the various OHDSI tools remains limited at best. Instead, many groups have created extensions to the OMOP CDM, like the pregnancy extension to capture relational information **[3]**. The oncology extension made it possible to capture the majority of the relational information in the POS-VAP dataset; with enough interest in the broader OHDSI community, we expect to propose a microorganism-specific extension that builds on existing oncology components and addresses the limitations described in the sections above. Such an extension

would enhance this work and could enable patient-specific and accurate monitoring of microbial drug resistance across broad hospital networks.

References:

[1] Kostyanev, Tomislav, et al. "COMBACTE LAB-Net: building a European laboratory network for clinical trials on anti-infectives." Future Microbiology 16.9 (2021): 635-647.

[2] https://forums.ohdsi.org/t/fact-relationships-searching-for-an-extensible-approach/18393

[3] Jones, Sara, et al. "Who is pregnant? defining real-world data-based pregnancy episodes in the National COVID Cohort Collaborative (N3C)." medRxiv (2022).