MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI230948

Toward Real-World Reproducibility: Verifying Value Sets for Clinical Research

Scott L. DUVALL[a], Craig G. PARKER[b], Amanda R. SHIELDS[c], Patrick R. ALBA[d], Julie A. LYNCH[a], Michael E. MATHENY[e], and Aaron W. C. KAMAUU[b,1]
[a]*Department of Veterans Affairs, United States*
[b]*Navidence, United States*
[c]*Quanten LLC, United States*
[d]*University of Utah, United States*
[e]*Vanderbilt University Medical Center, United States*
ORCiD ID: Scott L. DuVall https://orcid.org/0000-0002-4898-3865, Craig G. Parker https://orcid.org/0000-0002-9363-279X, Amanda R. Shields https://orcid.org/0000-0002-3443-7986, Patrick R. Alba https://orcid.org/0000-0002-4898-3865, Julie A. Lynch https://orcid.org/0000-0002-5176-5447, Michael E. Matheny https://orcid.org/0000-0003-3217-4147, Aaron W. C. Kamauu https://orcid.org/0000-0001-5954-115X

Abstract. Standardized operational definitions are an important tool to improve reproducibility of research using secondary real-world healthcare data. This approach was leveraged for studies evaluating the effectiveness of AZD7442 as COVID-19 pre-exposure prophylaxis across multiple healthcare systems. Value sets were defined, grouped, and mapped. Results of this exercise were reviewed and recorded. Value sets were updated to reflect findings.

Keywords. OHDSI, OMOP, RWD, operational definitions, value sets

1. Introduction

Secondary use of real-world healthcare data is becoming increasingly integrated into regulatory decision-making for medicine approvals. The adoption of robust, reproducible methods for generating evidence from these data is critical. Standardized operational definitions of clinical concepts are a core component of a reproducible approach.

In line with this "best practice" approach, standardized operational definitions were developed for a global study describing use and effectiveness of AZD7442. AZD7442 is a combination of tixagevimab/cilgavimab, two neutralising antibodies targeting the SARS-CoV-2 spike protein, that received FDA Emergency Use Authorization (EUA) in December 2021 for COVID-19 pre-exposure prophylaxis (PrEP) in patients with moderate to severe immunocompromising (IC) medical conditions. These operational definitions included value sets, which, where possible, were drawn from common use (e.g., Charlson Comorbidity Index [1-3], Value Set Authority Center [4]), validated constructs (e.g., eMERGE Phenotype KnowledgeBase [5]), and other published resources [6,7]. However, as terminologies and dictionaries are constantly evolving,

[1]Corresponding Author: Aaron Kamauu,. aaron@navidence.com

these value sets need to be reviewed prior to implementation. A technology-enabled approach for handling multiple, complex value sets was piloted within the United States Department of Veterans Affairs (VA) electronic medical record (EMR) data using the VA Informatics and Computing Infrastructure (VINCI) which is mapped to the Observational Health Data Sciences and Informatics' Observational Medical Outcomes Partnership (OMOP) Common Data Model [8].

2. Methods

The operational definitions encompassed 9 broad categories of immunocompromised patients. The Food and Drug Administration issued an Emergency Use Authorization for AZD7442 specifying categories of immunocompromise, including but not limited to:

- Active treatment for solid tumor and hematologic malignancies.
- Hematologic malignancies associated with poor responses to COVID-19 vaccines regardless of current treatment status (e.g., chronic lymphocytic leukemia, non-Hodgkin lymphoma, multiple myeloma, acute leukemia).
- Receipt of solid-organ transplant or an islet transplant and taking immunosuppressive therapy.
- Receipt of chimeric antigen receptor-T-cell or hematopoietic stem cell transplant (within 2 years of transplantation or taking immunosuppression therapy).
- Moderate or severe primary immunodeficiency (e.g., common variable immunodeficiency disease, severe combined immunodeficiency, DiGeorge syndrome, Wiskott-Aldrich syndrome).
- Advanced or untreated human immunodeficiency virus (HIV) infection (people with HIV and CD4 cell counts <200/mm3, history of an acquired immunodeficiency syndrome [AIDS]-defining illness without immune reconstitution, or clinical manifestations of symptomatic HIV).
- Active treatment with high-dose corticosteroids (i.e., ≥20 mg prednisone or equivalent per day when administered for ≥2 weeks), alkylating agents, antimetabolites, transplant-related immunosuppressive drugs, cancer chemotherapeutic agents classified as severely immunosuppressive, and biologic agents that are immunosuppressive or immunomodulatory (e.g., B-cell depleting agents).

To allow for clinical discretion with regard to care of potentially immunocompromised patients, these categories were expanded to also include:

- Patients with solid tumors not on treatment.
- Patients with end-stage renal disease or on dialysis.

These value sets utilized ICD-10-CM, CPT, ICD-10-PCS, HCPCS, RxNorm codes and uncoded drug names. The following methods were used to verify value sets and identify gaps.

- Intensional (implicit algorithm) value sets as well as extensional (explicit) lists of concepts for each category were defined from published sources as above.

- For value sets based on an intensional definition (e.g., asthma = ICD-10-CM codes J45.x), algorithms were processed against the relevant dictionary to generate the extensional list of concepts and codes.
- For value sets based on an extensional definition, all codes were processed to see if there were codes with additional granularity that were not included (e.g., mild intermittent asthma = J45.2 can expand to J45.20, J45.21, and J45.22: mild intermittent asthma, uncomplicated; with acute exacerbation; and with status asthmaticus, respectively).
- For all value sets, we grouped the codes of a given type (e.g., diagnosis, procedure, medications) and ran broad queries against the source EMR data to identify any codes that did not map or those that resulted in zero instances in the patient population. Given the size and breadth of healthcare across the VA, most concepts were expected to have been documented. Those without direct mappings and those with zeros instances were reviewed.
- For all value sets, codes were mapped with corresponding concepts in OMOP, and then mapped back to all codes in the original terminology that mapped to the same OMOP concept (Figure 1). The resulting list of concepts were compared to the original value sets and differences were evaluated to identify potentially relevant additional concepts.
- For all medication value sets, codes or the uncoded drug names were mapped to OMOP concepts. All concepts were then mapped to the ingredient(s) the drugs contained and then to all drugs containing those ingredients. Other drugs in the same VA drug class or same Anatomical Therapeutic Chemical Third, Fourth, or Fifth level classes were also identified. Each of these medications were reviewed for inclusion.

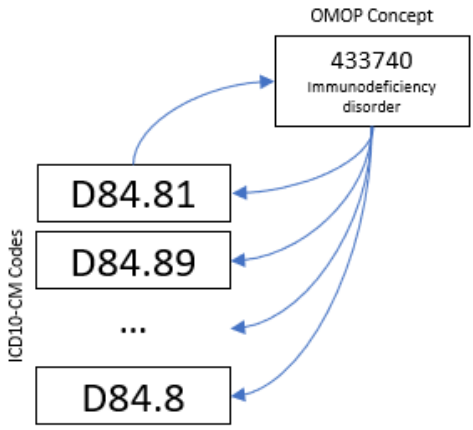

Figure 1. Mapping between original code in the value set to the associated OMOP concept and then to all other codes mapped to that same OMOP concept.

3. Results

There were 22 total value sets were used for operational definitions of immunocompromised conditions for AZD7442 eligibility: 11 diagnosis value sets (1479 concepts), 5 procedure value sets (297 concepts), 5 medication lists (280 medications), and 1 lab value set. There were 9 value sets were based on intensional definitions, whereas the others were extensional lists of distinct concepts.

There were 298 concepts (across all types) that had zero instances in the VA EMR data. Upon review, 1 concept was a typographical error and the remainder were determined to be non-terminal or category-level codes.

There were 339 concepts (across all types) mapped to 1458 OMOP concepts, that further mapped to 2992 concepts that were not included in the original value sets. Of these, only 7 diagnosis codes and 18 medications were determined to be relevant to the intent of the respective value set and therefore were added for our research program.

4. Discussion

When evaluating concepts that had zero instances in the VA EMR data, we found many that have always been non-terminal codes. Use of non-terminal codes in patient data was unexpected. In value sets from some sources, non-terminal concepts were included for completeness and hierarchical reference.

In processing the concept crosswalk via OMOP, we found that all 7 diagnosis codes were identified as terminal codes from 2015-2020 that were replaced with new children codes in October 2020. When processing the intensional definition on the newest version of ICD-10-CM, we only captured the new terminal codes and did not include parent codes. However, as these codes once were terminal codes, we needed to include them in our value sets to capture all instances of their use in retrospective data queries. This activity helped identify this issue and guide us to enhancing our value sets to capture the relevant current and former terminal codes that might have been used in medical documentation.

For the medications, all 18 found in the crosswalk were drug names that did not automatically match to a drug name or RxNorm concept in the VA system. As such, the team manually looked each up using the OMOP Athena tool to find the most relevant RxNorm ingredient concept. We added these to our value set.

The other 2967 concepts identified via OMOP crosswalk were not relevant for the intent of our study. Most of these were due to the initial code mapping to an OMOP concept that was not relevant. For example, ICD-10-CM code O98.711 <HIV disease complicating pregnancy, first trimester> mapped to multiple concepts via OMOP, including “First trimester pregnancy”, “Second trimester pregnancy”, “Third trimester pregnancy”, “Finding related to pregnancy”, etc.; however, those concepts are not relevant to this study. The original ICD-10-CM code of O98.711 is part of a value set for diagnosis of HIV/AIDS. In this case we were able to rule-out the mapping at this concept level.

In other cases, mapped codes were ruled out at the individual code level. For example, ICD-10-CM code I25.750 <Atherosclerosis of native coronary artery of transplanted heart with unstable angina> mapped to other atherosclerosis ICD-10-CM codes via OMOP. However, for this study, the original ICD-10-CM code is part of a

value set for diagnosis of prior solid organ transplant. Therefore, only the transplant-related diagnosis code is relevant and the other atherosclerosis diagnoses are not.

In addition, ICD-10-CM code C7A.012 <Malignant carcinoid tumor of the ileum> mapped to ICD-10-CM code D3A.012 <Benign carcinoid tumor of the ileum> via OMOP mapping to concept "Carcinoid tumor of ileum". However, our cancer diagnosis value sets do not include benign tumors.

5. Conclusions

This is a pilot in development of a formal robust and reproducible methodology for validating, enhancing, and updating value sets used in operational definitions for clinical research. This extensive exercise leverages standard terminologies, dictionaries, and data models to provide a comprehensive set of relevant concepts. Specific application to the AZD7442 program, provided confidence in these operational definitions that are being deployed via effectiveness study protocols across several countries around the world.

References

[1] Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE, Ghali WA. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005 Nov;43(11):1130-9, doi: 10.1097/01.mlr.0000182534.19832.83.

[2] Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, Januel JM, Sundararajan V. Updating and validating the charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol. 2011 Mar;173(6):676-82, doi: 10.1093/aje/kwq433.

[3] Beyrer J, Manjelievskaia J, Bonafede M, Lenhart G, Nolot S, Haldane D, Johnston J. Validation of an International Classification of Disease, 10th revision coding adaptation for the Charlson Comorbidity Index in United States healthcare claims data. Pharmacoepidemiol Drug Saf. 2021 May;30(5):582-93, doi: 10.1002/pds.5204.

[4] de novo code lookup informed by National Library of Medicine Value Set Authority Center [Internet]. Bethesda: National Library of Medicine; c2022 [cited 2022 Apr 10]. Available from: https://vsac.nlm.nih.gov/welcome.

[5] PheKB [Internet]. Nashville: Vanderbilt University; c2017 [cited 2022 Nov 21]. Available from: https://phekb.org/phenotype/1578.

[6] Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, Payne PRO, Pfaff ER, Robinson PN, Saltz JH, Spratt H, Suver C, Wilbanks J, Wilcox AB, Williams AE, Wu C, Blacketer C, Bradford RL, Cimino JJ, Clark M, Colmenares EW, Francis PA, Gabriel D, Graves A, Hemadri R, Hong SS, Hripscak G, Jiao D, Klann JG, Kostka K, Lee AM, Lehmann HP, Lingrey L, Miller RT, Morris M, Murphy SN, Natarajan K, Palchuk MB, Sheikh U, Solbrig H, Visweswaran S, Walden A, Walters KM, Weber GM, Zhang XT, Zhu RL, Amor B, Girvin AT, Manna A, Qureshi N, Kurilla MG, Michael SG, Portilla LM, Rutter JL, Austin CP, Gersing KR. The National COVID Cohort Collaborative (N3C): rationale, design, infrastructure, and deployment. J Am Med Inform Assoc. 2021 Mar;28(3):427-43, doi: 10.1093/jamia/ocaa196.

[7] Sentinel Initiative [Internet]. Boston: Sentinel Operations Center; c2022 [cited 2022 Apr 5]. Available from: https://www.sentinelinitiative.org/methods-data-tools/methods/master-protocol-development-covid-19-natural-history.

[8] Lynch KE, Deppen SA, DuVall SL, Viernes B, Cao A, Park D, Hanchrow E, Hewa K, Greaves P, Matheny ME. Incrementally transforming electronic medical records into the observational medical outcomes partnership common data model: a multidimensional quality assurance approach. Appl Clin Inform. 2019 Oct;10(5):794-803, doi: 10.1055/s-0039-1697598.