Content Coverage Evaluation of the OMOP Vocabulary on the Transplant Domain Focusing on Concepts Relevant for Kidney Transplant Outcomes Analysis

Sylvia Cho[1]   Margaret Sin[1]   Demetra Tsapepas[2,3]   Leigh-Anne Dale[4]   Syed A. Husain[5]
Sumit Mohan[5,6]   Karthik Natarajan[1]

[1] Department of Biomedical Informatics, Columbia University, New York, New York, United States
[2] Department of Surgery, Columbia University, New York, New York, United States
[3] Department of Transplantation, New York Presbyterian Hospital, New York, New York, United States
[4] Department of Medicine, Columbia University Medical Center, New York, New York, United States
[5] Division of Nephrology, Department of Medicine, Columbia University Medical Center, New York, New York, United States
[6] Department of Epidemiology, Columbia University Mailman School of Public Health, New York, New York, United States

Address for correspondence Karthik Natarajan, PhD, Department of Biomedical Informatics, Columbia University, 622 West 168th Street, PH-20, New York, NY 10032-3784, United States (e-mail: kn2174@cumc.columbia.edu).

Abstract

Keywords
► transplants
► registries
► standard
► vocabulary
► OMOP
► UNOS

Background   Improving outcomes of transplant recipients within and across transplant centers is important with the increasing number of organ transplantations being performed. The current practice is to analyze the outcomes based on patient level data submitted to the United Network for Organ Sharing (UNOS). Augmenting the UNOS data with other sources such as the electronic health record will enrich the outcomes analysis, for which a common data model (CDM) can be a helpful tool for transforming heterogeneous source data into a uniform format.

Objectives   In this study, we evaluated the feasibility of representing concepts from the UNOS transplant registry forms with the Observational Medical Outcomes Partnership (OMOP) CDM vocabulary to understand the content coverage of OMOP vocabulary on transplant-specific concepts.

Methods   Two annotators manually mapped a total of 3,571 unique concepts extracted from the UNOS registry forms to concepts in the OMOP vocabulary. Concept mappings were evaluated by (1) examining the agreement among the initial two annotators and (2) investigating the number of UNOS concepts not mapped to a concept in the OMOP vocabulary and then classifying them. A subset of mappings was validated by clinicians.

Results   There was a substantial agreement between annotators with a kappa score of 0.71. We found that 55.5% of UNOS concepts could not be represented with OMOP standard concepts. The majority of unmapped UNOS concepts were categorized into transplant, measurement, condition, and procedure concepts.

Conclusion   We identified categories of unmapped concepts and found that some transplant-specific concepts do not exist in the OMOP vocabulary. We suggest that adding these missing concepts to OMOP would facilitate further research in the transplant domain.

Background and Significance

Organ transplant is the preferred treatment option for patients with organ failure.[1–3] The total number of transplants performed in the United States has grown by 19.8 percent since 2012, and 33,606 transplants were reported in 2016.[3] Therefore, continuous efforts to improve the outcomes of transplant recipients within and across transplant centers is significantly important. The Centers for Medicare and Medicaid Services (CMS) currently mandates reporting of key clinical variables from transplant centers to track and monitor their performance on transplant outcomes.[4] To facilitate this process, individual transplant centers submit patient-level data to the United Network for Organ Sharing (UNOS). These data are then analyzed by the Scientific Registry of Transplant Recipients and eventually reported to CMS and the public.

Although the current transplant outcomes registry effectively captures clinical data at the time of wait listing and transplantation, it has significant limitations.[4–6] Predefined clinical variables are obtained when patients are wait-listed, transplanted, and followed up posttransplant at 6 months, 12 months, and every year thereafter. In between these time points, there might be useful and meaningful patient data recorded in the electronic health record (EHR). Furthermore, transplant outcomes research is largely limited to the data variables that are already included in UNOS. We suggest that these challenges can be solved by augmenting the UNOS data with different data sources such as the EHR which could provide additional information to substantiate outcomes analysis.[7]

However, the challenge of augmenting UNOS registry data with other data sources exists on both the individual-center level and on a multicenter level. There is a challenge of disparate data sources having different data structures and coding systems within individual centers, and also barriers to interoperability resulting from disparate data models and concept representations among different transplant centers. This is an underlying problem when integrating observational databases which are stored in different formats and representations.

The problem of disparate observational databases could be solved by transforming the dataset into the format of a common data model (CDM), which is a common conceptual layout of data that enables the integration of multiple datasets through consistent data structure and unambiguous concept representation.[7] Transforming a dataset to a CDM requires standardizing the schema and terminology, which can be achieved by mapping heterogeneous terminologies and schemas.[8–10] However, this must be done vigilantly as there can be potential data loss if a terminology with detailed description is mapped to one that has fewer details.[9,11] Significant data loss due to lack of standardized concept

codes may impact the integrity of research.[8] Thus, a proper feasibility evaluation on mapping concepts from a source database to the target CDM format is necessary to produce reliable research results.

There are many existing CDMs including PCORnet, informatics for integrating biology in the bedside (i2b2), and Observational Medical Outcomes Partnership (OMOP) CDM, which are used in large nationwide initiatives.[12] Previous studies have evaluated multiple CDMs on various quality dimensions and have demonstrated that the OMOP CDM best satisfies the quality dimensions and is an appropriate model for comparative effectiveness and outcomes research.[7,9,13,14] More importantly, there are a rich set of open-source analytic tools that leverage OMOP CDM, which is why OMOP CDM is often chosen over other data models.[7,12,15] Thus, despite an existing effort to transform transplant data into the i2b2 format, the focus of this study will be on assessing the feasibility of mapping concepts from the UNOS database to concepts in the OMOP vocabulary.[16]

Objectives

There have been previous studies on concept mapping,[9,11,17,18] including a study by Dale et al that assessed whether the concepts in the UNOS registry can be successfully mapped to the OMOP vocabulary.[19] While the preliminary analysis reported by Dale et al was informative of vocabulary mappings between UNOS and OMOP, the analysis relied on one annotator to conduct concept mappings. This study aimed to apply a more formal assessment of UNOS content coverage in OMOP by including multiple annotators with complementary domain knowledge—clinical and informatics. This analysis reflects the revised mapping and the effort to have multiple annotators to ensure that there was broad consensus on linkages for content coverage of the OMOP vocabulary with respect to UNOS.

Methods

There are several steps involved in mapping the concepts in the UNOS registry to an OMOP vocabulary standard concept. First, we identified the UNOS registry forms and the concepts within the forms that need to be mapped. Second, two annotators (S.C. and M.S.) used a tool named USAGI, developed by the Observational Health Data Sciences and Informatics (OHDSI) community, to map the UNOS concepts to OMOP standard concepts.[20] USAGI is a tool which provides an interface that facilitates users to map source codes to OMOP standard concepts.[20] It uses a term similarity approach by utilizing synonyms for concepts in the OMOP vocabulary and

using a term similarity score to automatically map source code descriptions to the OMOP vocabulary.[21] The annotators were trained in biomedical informatics, including knowledge/concept representation and standard vocabularies, and also had background in health care. Third, we evaluated the reliability of mapping done by the annotators. Finally, we did a validation on a subset of our mappings with two clinicians in the transplant field.

Description of the Transplant Registry Forms

UNOS maintains a data entry system which is called the Transplant Information Electronic Data Interchange.[3] The system contains transplant-related data that are collected through forms submitted by transplant centers, organ procurement organizations, and histocompatibility laboratories across the country. The major forms that are used to collect most of the data are the transplant candidate registration (TCR) form, the transplant recipient registration (TRR) form, the transplant recipient follow-up (TRF) form, and the donor registration forms.[6] For each of these forms, there are respective forms for each organ including kidney, heart, intestine, lung, liver, and pancreas, as well as a set of forms for immunosuppression.[6]

The UNOS collection forms and its data dictionary can be accessed on the UNOS Web site.[22] The UNOS data dictionary was used as our source data; it includes the fields in the above forms as well as the fields' corresponding answer choices. In this article, we will refer to the set of fields in the various forms as the "question concepts," and the set of allowable answer choices from look-up tables as "look-up concepts" when we need to distinguish between the two categories.

Concept Cleaning Process

First, concepts on patient-identifiable demographic information such as name and address were removed. Redundant fields from the UNOS data dictionary were removed considering the context and semantics of the concepts. For example, (1) the source term "weight in kg" or "serum creatinine" existed in most of the forms including TCR, TRR, and TRF of all organs, donor registration forms, and more. Since the meaning of these terms is equivalent in all forms, we removed all source terms except one; (2) the source term "acute rejection" was included multiple times in different look-up tables. However, we did not remove the duplicate "acute rejection" terms because depending on which look-up table or form it comes from, the meaning of "acute rejection" can be different. For example, the answer "acute rejection" associated to the field "pancreas cause of graft failure" has the meaning "acute rejection of pancreas transplant," whereas the answer "acute rejection" for the field "kidney cause of graft failure" has the meaning "acute rejection of renal transplant."

Mapping UNOS Source Concepts to OMOP Standard Concepts

The OMOP standard vocabulary is a repository of vocabularies used in the research community, currently including over 70 vocabularies (e.g., SNOMED, ICD9, RxNorm, LOINC, etc.).[23,24] The purpose of OMOP vocabulary is to standardize the disparate formats and conventions of various vocabularies into a common structure.[23] Only one concept among all concepts that represent the same clinical event is selected as standard and is used in the OMOP CDM.[23] For example, there are many codes that define atrial fibrillation such as MeSH code D001281, SNOMED code 49436004, and ICD9CM code 427.31, but only the SNOMED code is designated as standard and is used to represent data in the OMOP CDM format.[23] The concepts in OMOP vocabulary can be accessed through ATHENA, a tool that enables researchers to search or download standardized vocabularies.[15,25]

Prior to mapping the UNOS source concepts to OMOP standard concepts, some general rules were established: (1) each concept would be mapped to the corresponding standard terminology that is generally considered the norm. For example, laboratory tests are mapped to LOINC codes and medications are mapped to RxNorm; (2) source concepts are mapped to OMOP standard concepts only if they are semantically equivalent. Additionally, concept descriptions provided by UNOS for each form were referenced to understand the exact meaning and context of concepts. To ensure consistent mappings, the annotators followed these rules and performed 20 mappings together using USAGI before individually mapping concepts.

Evaluation

There were two evaluations performed on concept mapping: (1) examining the agreement among the two annotators; (2) calculating the number of UNOS concepts that were not mapped to a concept in the OMOP vocabulary and categorizing their theme.

The degree of agreement between the annotators was evaluated using Cohen's kappa.[26] Agreement was measured at the level of whether UNOS concepts were mapped or not mapped, which allowed us to evaluate the *content coverage* of the OMOP vocabulary for the transplant domain. Each annotator could *map* or *not map* the UNOS source concept to an OMOP standard concept, therefore we classified mappings as follows: (1) both annotators mapped the UNOS concept to an OMOP concept, (2) both annotator did not map the UNOS concept to an OMOP concept, (3) annotator 1 mapped the UNOS concept to an OMOP concept but annotator 2 did not, and (4) annotator 1 did not map the UNOS concept to an OMOP concept but annotator 2 did. Category (1) not only includes concepts that were mapped to the same concept code but also concepts mapped to different concept codes because the source concepts were deemed mappable by both annotators despite being mapped to different codes.

In the second part of our evaluation, we measured the extent of UNOS source concepts that could not be mapped. The two annotators reviewed concepts in disagreement to come to a consensus on whether the concepts can be mapped or not. After this review, the final batch of unmapped UNOS concepts was iteratively organized by the research group into categories that were empirically derived. Both annotators quickly reviewed the concepts together and devised a list of potential themes. The annotators individually labeled the unmapped concepts with the list of themes devised together, but the annotators also created new categories if needed. The results were discussed to reach a consensus on the labeled

Table 1 Total number of unique concepts in each registry form

Forms	Form description	All concepts (N)	Unique concepts (N)
TCR (heart, heart/lung, kidney, intestine, liver, kidney/pancreas)	Transplant candidate registration	698	105
TRR (heart, immunosuppression, pancreas, liver, kidney/pancreas, kidney, intestine)	Transplant recipient registration	1,181	147
TRF (thoracic, kidney, pancreas, liver, kidney/pancreas, immunosuppression)	Transplant recipient follow-up	599	78
CDR	Deceased donor registration	509	253
DCD	Serial data file	11	4
DHS	Donor histocompatibility	47	3
RHS	Recipient histocompatibility	105	61
LIEX	Liver recipient explant pathology	33	11
LDR	Living donor registration	235	123
LDF	Living donor follow-up	92	23
MALa	Malignancy	75	62
Look-up tables	Response look-up tables	2,701	2,701
Total		6,286	3,571

categories. Questionable concepts were determined by another domain expert (K.N.) in biomedical informatics.

Validation

We validated our results on a subset of transplant concepts that met the following conditions: (1) concepts in the UNOS kidney transplant forms and (2) concepts relevant for kidney transplant outcomes analysis. As a result, the concepts found in the TCR-kidney, TRR-kidney, TRF-kidney, and donor registration forms were validated. Among concepts in these forms, one clinician (S.M.), who has extensive experience in kidney transplant outcomes research, identified the pertinent concepts needed for transplant outcomes analysis. To our knowledge there are no predefined standard or common data elements in the solid organ transplant domain, thus we relied on expert opinion.[27,28]

A second clinician (D.T.) validated the kidney concepts identified by the first clinician. In addition, the second clinician validated the annotator mappings and corrected mappings where there was disagreement. In addition, in cases where concepts were left unmapped by both annotators, the second clinician examined the OMOP vocabulary to find a relevant concept where possible.

Results

A total of 6,286 concepts existed in all the forms prior to concept cleaning of the UNOS forms. After removing duplicate concepts, 3,571 unique concepts remained. The details of the total number of concepts in each form are described in ►**Table 1**.

The results of mapping the UNOS source concepts to the OMOP standard concepts are presented in ►**Table 2**. Among the 3,571 source concepts, 35% were mapped and 50% remained unmapped to OMOP standard concepts by both annotators. This showed that there was substantial agreement between the annotators with a kappa score of 0.71.[26] Examples of mapping UNOS source concepts to an OMOP standard concept are shown in ►**Table 3**. The UNOS source term "acute

Table 2 Results of mapping UNOS source concepts to OMOP standard concepts for both annotators

Different Scenarios for concept mapping	Concepts (N)	Percentage
Both annotators were able to map UNOS source concepts to OMOP standard concepts	1,268	35.51
Annotator 1 and annotator 2 mapped the UNOS source concept to the same OMOP concepts	877	24.56
Annotator 1 and annotator 2 mapped the UNOS source concept to different OMOP concepts	391	10.95
Both annotators were unable to map UNOS source concepts to OMOP concepts	1,794	50.24
Annotator 1 mapped, but annotator 2 was unable to map the UNOS source concepts to OMOP concepts	221	6.19
Annotator 1 was unable to map, but annotator 2 mapped the UNOS source concepts to OMOP concepts	288	8.06
Total number of concepts	3,571	

Abbreviations: OMOP, Observational Medical Outcomes Partnership; UNOS, United Network for Organ Sharing.

Table 3 Example result of UNOS concepts represented with OMOP vocabulary

UNOS source concept	OMOP target concept				
	Concept ID	Concept name	Concept domain	Concept class	Vocabulary
Acute rejection	4200464	Acute rejection of pancreas transplant	Condition	Clinical finding	SNOMED
Cystic fibrosis	441267	Cystic fibrosis	Condition	Clinical finding	SNOMED
Lobectomy	4054047	Lobectomy	Procedure	Procedure	SNOMED
LVAD	4235161	Left ventricular assist device	Device	Physical object	SNOMED
Has the recipient ever had a diagnosis of HCC?	46270540	History of hepatocellular carcinoma	Observation	Context-dependent	SNOMED
DR	3021667	HLA-DR locus [Type]	Measurement	Laboratory test	LOINC
Korean	38003585	Korean	Race	Race	Race
Zortress (everolimus)	40175824	Everolimus oral tablet [Zortress]	Drug	Branded drug form	RxNorm
Simultaneous kidney–pancreas	2721092	Simultaneous pancreas–kidney transplantation	Procedure	HCPCS	HCPCS

Abbreviations: HCPCS, Healthcare Common Procedure Coding System; OMOP, Observational Medical Outcomes Partnership; UNOS, United Network for Organ Sharing.

rejection" was mapped to the OMOP concept "acute rejection of pancreas transplant" because the term "acute rejection" was extracted from the look-up table for "pancreas cause of graft failure." Similarly, the source term "simultaneous kidney–pancreas" refers to a specific type of transplant procedure, so it was mapped to the procedure concept in the Healthcare Common Procedure Coding System (HCPCS) vocabulary.

After a second-round review of the UNOS concepts, we found that approximately half of the UNOS source concepts cannot be represented with the OMOP vocabulary. There were an additional 187 concepts that the annotators disagreed on in the initial review but later came to consensus that the concepts cannot be mapped to an existing OMOP concepts. Details of the results are presented in ►**Fig. 1**.

During the second-round review of unmapped concepts, 13 concept categories were empirically derived as shown in ►**Table 4**. The commonly occurring concept categories were medical condition (23%), transplant (17%), and
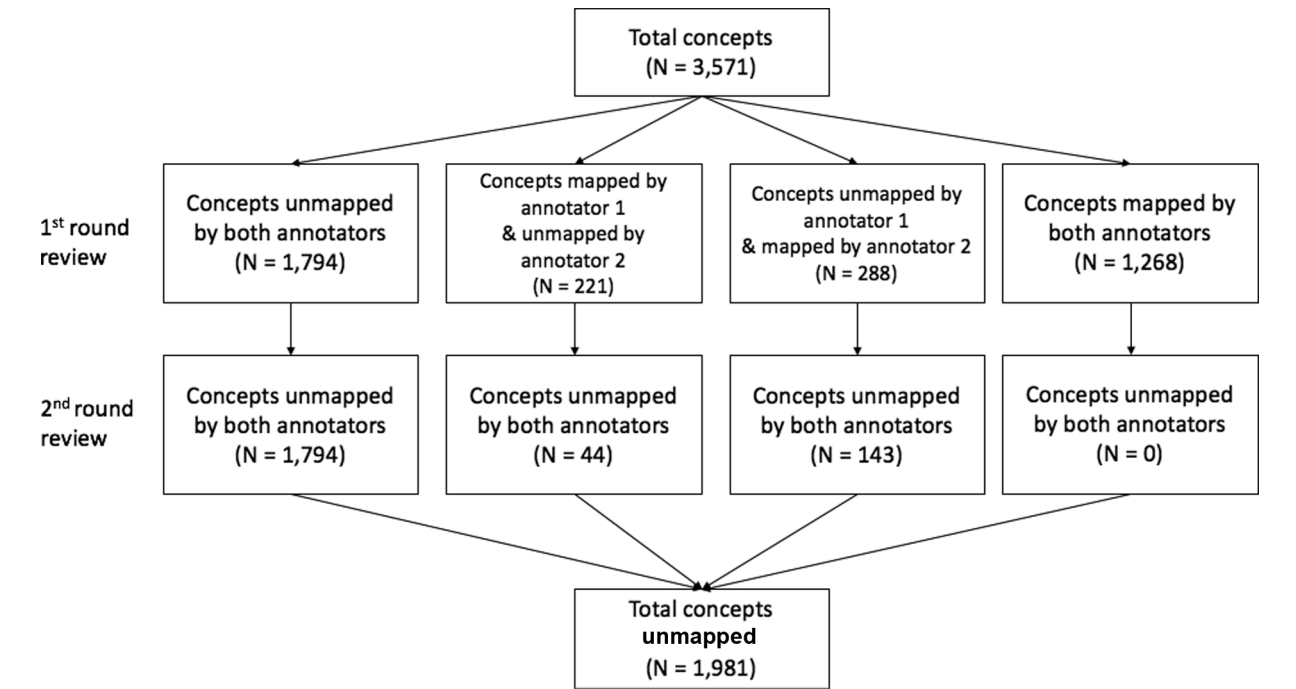

Fig. 1 Number of unmapped concepts after the second review.

Table 4 Categories of unmapped concepts

Category	Count	Percentage	Examples
Measurement	612	30.89	% Macro vesicular fat, anti-CMV serology results, 1:04, 1:05, 10:01, 11:01 (DPA1, DPB1 HLA)
Condition	450	22.72	Diffuse cholangiopathy, "fibrosis expansion of some portal areas, with or without short fibrous septa," incidental carcinoma, cirrhosis type A, drug-treated COPD
Transplant	337	17.01	Pancreas with kidney different donor, total cold ischemia time right kidney, multiorgan noncluster, kidney graft status (received on pump), put on ice
Time	181	9.14	Ventilator support for ≤48 hours, intubated at 72 hours, 9 year after graft failure
Procedure	112	5.65	Orthotopic bicaval, left thoracotomy, celiac axis with pancreas (arterial reconstruction), sequential kidney
Device	61	3.08	Abiomed AB5000, Berlin Heart, Evaheart, Toyobo (all life support)
Administration	50	2.52	Public insurance—Medicare and Choice, Public insurance—Medicare unspecified, loss of health insurance, free care
Medical history	48	2.42	More than 5 previous pregnancies, history of hypertension diuretics, Chagas history
Demographic	47	2.37	Eskimo, grade school (0–8), age in months, inability to find work
Status	28	1.41	Mild decrease in activity level, 100%: fully active, normal, 10%: no play; does not get out of bed
Treatment	25	1.26	Diabetes treatment, induction, growth hormone therapy
Drug	21	1.06	Oral hypoglycemic agent, T10B9 (Medimmune), Mizoribine (Bredinin)
Quantity	8	0.4	"If abnormal, # of vessels with >50% stenosis": "left kidney/number of glomeruli visualized"
Total	1,981	100	

Abbreviations: CMV, cytomegalovirus; COPD, chronic obstructive pulmonary disease.

procedure (6%). The most commonly occurring category for unmapped concepts was the measurement category (31%), but 500 out of 612 concepts assigned to this category were of different human leukocyte antigen (HLA) values. Similarly, approximately 140 out of 181 concepts assigned to the time category were concepts such as 40 years, 50 years, and 40 years after graft failure.

In the validation stage, we found that the total number of question concepts in the kidney transplant forms was 492. Among these question concepts, 49 of them were concepts that the annotators mapped to the same OMOP concept code. We excluded these concepts from the total number of kidney transplant concepts that need to be validated. Among the remaining 443 question concepts, 157 concepts were determined by the clinician to be necessary for outcomes analysis. For example, serum creatinine at the time of transplant, cigarette use of deceased donor, preoperative blood pressure of living donor, and organ received on ice/pump were selected as important variables in outcomes analysis. Concepts such as "skin type of deceased donor" and "% macro/micro vesicular fat of living donor" were considered unnecessary for outcomes analysis.

There were 30 look-up tables that correspond to these 157 question concepts, totaling 266 look-up concepts. Among 266 look-up concepts, 119 of them were concepts that the annotators agreed on (mapped to the same OMOP concept code). Thus, the clinician only validated the remaining 147 look-up

concepts. Therefore, the total number of kidney transplant concepts that we conducted validation on (including both the question and look-up concepts) was 304. The clinician confirmed that there were 219 concepts that were not expressed in the OMOP vocabulary and would need to be incorporated to conduct most transplant-specific outcomes analysis on the OMOP CDM. In addition, the decision on the mappings of 14 concepts out of 304 concepts were changed by the second clinician during validation. This means that approximately 95.4% of the subset of mappings done by annotators were considered valid. The mappings between UNOS and OMOP concepts can be found in our GitHub repository.[29]

Discussion

In this study, we investigated the content coverage of the OMOP vocabulary on concepts in the UNOS transplant registry. We mapped 3,571 UNOS source concepts to standard concepts in the OMOP vocabulary. Kappa score of the two annotators showed substantial agreement on whether a UNOS concept maps to an OMOP concept, and thus confirms the result of our annotation to be trustworthy. We found that approximately half of the UNOS source concepts cannot be represented with an OMOP standard concept. This can potentially lead to a significant amount of data loss when standardizing the UNOS transplant data into an OMOP CDM. We classified the unmapped concepts into different

categories which were derived empirically, and found that UNOS concepts in the measurement, condition, transplant, and procedure categories were the most commonly un-mapped UNOS concepts.

Although about half of the UNOS concepts were not covered by the OMOP standard vocabulary, we found that a small portion of UNOS concepts were concepts that would not exist in a terminology that abides by the Desiderata (e.g., concept orientation).[30] Many of the unmapped concepts in the Measurement category and Time category were numerical values.[30] For instance, there is a set of concepts that represent the number of years after transplantation and the number of years after graft failure (e.g., "1 YEAR," "5 YEAR," "5 YEAR AFTER GRAFT FAILURE," "50 YEAR AFTER GRAFT FAILURE," etc.). These concepts do not necessarily have to exist in a terminology, but rather the meaning of these concepts could be conveyed in conjunction with the schema of the CDM. For example, the measurement values can be represented by including the concept of the corresponding laboratory test (e.g., HLA Ab) into the MEASUREMENT table and linking it with the value in the "value_as_number" field instead of mapping the values to a concept code.[31] Further-more, even if a "transplant date" concept does not exist in the OMOP vocabulary, we could infer the same meaning by putting the concept "transplant" into the PROCEDURE_OC-CURRENCE table and link it with the "procedure_date" field.[32] Taking this into consideration, the actual number of concepts that can be mapped might increase compared with what was determined in this study. Nevertheless, we find that there is a gap in the current OMOP vocabulary in providing enough content coverage on transplant-specific concepts such as "cold ischemia time" in expressing a specific organ or its laterality, which is an essential variable when studying transplant outcomes.

There were several challenges when mapping the UNOS source concepts to the OMOP standard concepts. First, varying levels of concept granularity made concept mapping challeng-ing. For example, in UNOS, concepts for Medicare were very specific (e.g., public insurance—Medicare and choice, public insurance —Medicare unspecified), whereas the OMOP vocab-ulary only had a broader concept "Medicare." On the other hand, sometimes UNOS concepts were broader (e.g., graft failure) when the OMOP concept only contained narrower concepts such as "primary graft failure" (OMOP ID = 4087398), "bone graft failure" (OMOP ID = 4308707), and "skin graft failure" (OMOP ID = 4308404). Second, it is difficult to map complex and composite concepts to a single OMOP concept. For example, complex concepts such as "pancreas with kidney different donor" or "unable to participate in academics due to disease" would rarely exist as a single concept in an OMOP vocabulary. In addition, a composite concept such as the "durable power of attorney/healthcare proxy" could not be mapped because only part of its concept could be mapped to an OMOP concept. While the "durable power of attorney" has a corresponding OMOP concept "active durable power of attor-ney for healthcare," the "healthcare proxy" does not have a mappable OMOP concept. Lastly, the OMOP vocabulary had some inconsistent content coverage.[33] For instance, the UNOS

concept "extracranial tumor" could not be found in OMOP; however, the OMOP vocabulary had "intracranial tumor". Similarly, the concept "malposition of liver" existed in UNOS, but this concept did not exist in the OMOP vocabulary while it had other similar concepts such as "malposition of heart" and "malposition of uterus".

While it was time consuming for each annotator to manually map 3,571 concepts, manual mapping was impor-tant as it provided the opportunity for human reasoning.[34] Relying only on USAGI's automatic mapping would have returned suboptimal results. Our study of mapping UNOS concepts to OMOP concepts can lead to additional benefits by contributing to the standardization of UNOS transplant registry in the OMOP CDM format. Although UNOS is well populated, it is known that there are some missing and incorrect data.[35] Currently, data are manually entered sepa-rately into the transplant outcomes registry, which contrib-utes to problems such as inaccurate data reporting in registries, delays in data inclusion, or missing data. Mapping all concepts to OMOP concepts would be the initial step of enabling a real-time, automatic, and standardized popula-tion of the registry with data extracted from the EHR.[36,37] Furthermore, our work contributes to the journey toward conducting a large-scale multisite study among national and international transplant centers. The transplant community will have the ability to investigate differences in transplant outcomes between races and regions with data not limited to a single transplant center, but with data aggregated from other transplant centers in the United States or abroad. Our work in mapping the UNOS concepts to OMOP concepts is an essential step toward achieving this goal.

However, there were a few limitations in our study. First, there were only two annotators. Having an odd number of annotators for tie-breaker situations might have improved some of the mappings. We tried to mitigate this weakness by asking a clinician on concepts that the annotators disagreed on and validating the results with a clinician who has experience in solid organ transplantation. Although only a subset of concepts was validated, not all UNOS concepts are relevant for outcomes analysis and thus focusing on the essential concepts was a more efficient option considering the limited availability of clinicians' time and the ultimate goal of adopting OMOP for outcomes analysis for the transplant domain. In addition, since the annotators were not clinicians, it is possible that the annotators tried a limited scope of search terms within USAGI when searching for potential OMOP concepts to map to. As USAGI recommends concepts based on term similarity, the annotators tried to utilize this function as much as possible. However, if needed, the annotators could ask clinicians or use search engines and materials provided by UNOS to understand concepts and identify potential search queries. Despite the limitations, the authors would like to put the focus of our work in providing the result of mappings to the research community so that a collab-orative effort could be made in improving the mappings.

In our future studies, we plan to expand the validation to other organs. We will also be requesting to add the concepts that we determined to be necessary for kidney transplant outcomes research into the OMOP vocabulary. In addition,

we plan to transform the UNOS dataset to follow the OMOP CDM format and release the ETL code to the OHDSI community.

Conclusion

We found that approximately half of the UNOS concepts could not be mapped to OMOP standard concepts, which could lead to significant data loss when transforming UNOS registry data into an OMOP CDM. Moreover, we recognized that a major portion of unmapped concepts belongs to the transplant, measurement, condition, and procedure categories. To bridge the gap, we suggest that adding these concepts from UNOS to the OMOP vocabulary would be able to facilitate further research related to organ transplant.

Clinical Relevance Statement

This study analyzes whether the current OMOP vocabulary is sufficient in its content coverage for concepts in the UNOS registry and finds gaps in the vocabulary. We hope that this could motivate the stakeholders to actively pursue adding pertinent missing concepts. In addition, the concept mapping between the UNOS concepts and OMOP concepts can be used as a reference to anyone who is interested in transforming their UNOS data into an OMOP CDM format.

Multiple Choice Questions

1. Why do we suggest using a common data model (CDM) in the transplant field?
    a. It solves the challenge of integrating UNOS data from one institution with UNOS data from another institution by providing a standard format.
    b. It solves the challenge of integrating UNOS data with disparate data sources (e.g., EHR, claims data) by providing a standard format.
    c. It solves the challenge of integrating UNOS data within an institution from different time points by providing a standard format.
    d. It solves the challenge of integrating UNOS data within an institution documented by different care providers.

    Correct Answer: The correct answer is option b. Transforming disparate datasets into a common format enables the integration of different data sources. UNOS data from different institutions would not need a CDM as all of them follow the UNOS data format. UNOS data from different time points would not change the data format as long as UNOS has not changed the data structure or concepts used over time. Also, data documented by different providers in the UNOS system do not change either the structure or the concepts used in the system.

2. Which of the followings did not emerge as a category for unmapped concepts between UNOS and OMOP?
    a. Transplantation
    b. Condition

    c. Procedure
    d. Patient identifiable demographics

    Correct Answer: The correct answer is option d. We removed all patient-identifiable demographic concepts as they would not be important for outcomes research.

Conflict of Interest
S.M. reports grants and other from Angion Pharmaceuticals, personal fees from Kidney International Reports, outside the submitted work.

References

1 Abecassis M, Bartlett ST, Collins AJ, et al. Kidney transplantation as primary therapy for end-stage renal disease: a National Kidney Foundation/Kidney Disease Outcomes Quality Initiative (NKF/KDO-QITM) conference. Clin J Am Soc Nephrol 2008;3(02):471–480
2 Tonelli M, Wiebe N, Knoll G, et al. Systematic review: kidney transplantation compared with dialysis in clinically relevant outcomes. Am J Transplant 2011;11(10):2093–2109
3 United Network for Organ Sharing. US organ transplants, deceased donors set record in 2016. Available at: https://www.unos.org/us-organ-transplants-set-record-in-2016/. Published 2017. Accessed June 21, 2020
4 Howard RJ, Cornell DL, Schold JD. CMS oversight, OPOs and transplant centers and the law of unintended consequences. Clin Transplant 2009;23(06):778–783
5 Srinivas TR, Taber DJ, Su Z, et al. Big data, predictive analytics and quality improvement in kidney transplantation- a proof of concept. Am J Transplant 2017;17(03):671–681
6 Massie AB, Kucirka LM, Segev DL. Big data in organ transplantation: registries and administrative claims. Am J Transplant 2014; 14(08):1723–1730
7 Cho S, Mohan S, Husain SA, Natarajan K. Expanding transplant outcomes research opportunities through the use of a common data model. Am J Transplant 2018;18(06):1321–1327
8 Rijnbeek PR. Converting to a common data model: what is lost in translation?: commentary on "fidelity assessment of a clinical practice research datalink conversion to the OMOP common data model" Drug Saf 2014;37(11):893–896
9 Ogunyemi OI, Meeker D, Kim H-E, Ashish N, Farzaneh S, Boxwala A. Identifying appropriate reference data models for comparative effectiveness research (CER) studies based on data from clinical information systems. Med Care 2013;51(08, Suppl 3):S45–S52
10 Steindel SJ. International classification of diseases, 10th edition, clinical modification and procedure coding system: descriptive overview of the next generation HIPAA code sets. J Am Med Inform Assoc 2010;17(03):274–282
11 Voss EA, Makadia R, Matcho A, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. J Am Med Inform Assoc 2015;22 (03):553–564

12 Klann JG, Joss MAH, Embree K, Murphy SN. Data model harmonization for the All Of Us Research Program: transforming i2b2 data into the OMOP common data model. PLoS One 2019;14(02):e0212463

13 Kahn MG, Batson D, Schilling LM. Data model considerations for clinical effectiveness researchers. Med Care 2012;50(Suppl): S60–S67

14 Garza M, Del Fiol G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. J Biomed Inform 2016;64:333–341

15 Observational Health Data Sciences and Informatics. Software tools. Available at: https://www.ohdsi.org/software-tools/. Accessed June 20, 2020

16 Maier C, Christoph J, Schmidt D, et al. Experiences of transforming a complex nephrologic care and research database into i2b2 using the IDRT tools. J Healthc Eng 2019;2019:5640685

17 Zhou X, Murugesan S, Bhullar H, et al. An evaluation of the THIN database in the OMOP Common Data Model for active drug safety surveillance. Drug Saf 2013;36(02):119–134

18 Matcho A, Ryan P, Fife D, Reich C. Fidelity assessment of a clinical practice research datalink conversion to the OMOP common data model. Drug Saf 2014;37(11):945–959

19 Dale L-A, Mohan S, Natarajan K. Advancing outcomes analysis within and across transplant centers using the OMOP CDM. OHDSI Symposium. Available at: https://www.ohdsi.org/web/ wiki/lib/exe/fetch.php?media=resources:unos_mapping.pdf. Published 2016. Accessed June 20, 2020

20 Ryan P, Duke J, Hripcsak G. Open-source big data analytics in healthcare. Available at: https://www.ohdsi.org/wp-content/ uploads/2014/07/OHDSI-Tutorial-PreFinal-mod.pdf. Accessed June 21, 2020

21 Observational Health Data Sciences and Informatics. USAGI. Available at: http://www.ohdsi.org/web/doku.php?id=documentation: software:usagi. Published 2017. Accessed June 21, 2020

22 United Network for Organ Sharing (UNOS). Data collection forms. Available at: https://transplantpro.org/technology/data-collection-forms/. Accessed June 21, 2020

23 Reich C, Ostropolets A. Chapter 5: Standardized vocabularies. In: The Book of OHDSI. 2019. Available at: https://ohdsi.github.io/ TheBookOfOhdsi/StandardizedVocabularies.html#fn25. Accessed August 12, 2020

24 Maier C, Lang L, Storf H, et al. Towards Implementation of OMOP in a German University Hospital Consortium. Appl Clin Inform 2018; 9(01):54–61

25 Observational Health Data Sciences and Informatics. ATHENA. http://athena.ohdsi.org. Accessed June 20, 2020.

26 McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(03):276–282

27 US National Library of Medicine. NIH common data elements (CDE) repository. Available at: https://cde.nlm.nih.gov/home. Accessed June 21, 2020

28 Common Data Element Repository (CDER) Library. Available at: https://repository.usaspending.gov/cder_library/. Accessed June 21, 2020

29 Cho S, Sin M, Tsapepas D, et al. UNOS-to-OMOP GitHub repository. Available at: https://github.com/sylviacho/UNOS-to-OMOP/wiki. Published 2020. Accessed July 12, 2020

30 Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998;37(4–5):394–403

31 Observational Health Data Sciences and Informatics. Common data model: Measurement table. OHDSI. Available at: http:// www.ohdsi.org/web/wiki/doku.php?id=documentation:cdm: measurement. Published 2017. Accessed June 21, 2020

32 Observational Health Data Sciences and Informatics. Common data model: Procedure_occurrence table. OHDSI. Available at: http:// www.ohdsi.org/web/wiki/doku.php?id=documentation:cdm:procedure_occurrence. Published 2017. Accessed June 21, 2020

33 Lee D, Cornet R, Lau F, de Keizer N. A survey of SNOMED CT implementations. J Biomed Inform 2013;46(01):87–96

34 Friedman CPA. A "fundamental theorem" of biomedical informatics. J Am Med Inform Assoc 2009;16(02):169–170

35 Tsapepas D, King KL, Husain SA, Mohan S. Evaluation of kidney allocation critical data validity in the OPTN registry using dialysis dates. Am J Transplant 2020;20(01):318–319

36 Marsolo K. EHR-linked registries for studies of populations. In: Hutton JJ, ed. Pediatric Biomedical Informatics: Computer Applications in Pediatric Research. Dordrecht: Springer; 2012:133–148

37 Seneviratne MG, Seto T, Blayney DW, Brooks JD, Hernandez-Boussard T. Architecture and implementation of a clinical research data warehouse for prostate cancer. EGEMS (Wash DC) 2018;6(01):13