eHealth2015 – Health Informatics Meets eHealth
D. Hayn et al. (Eds.)

doi:10.3233/978-1-61499-524-1-15

Standardized Quality Assurance Forms for Organ Transplantations with Multilingual Support, Open Access and UMLS Coding

Julian VARGHESE, Sarah SCHULZE SÜNNINGHAUSEN and Martin DUGAS
Institute of Medical Informatics, University of Muenster, Germany

Abstract. Quality assurance (QA) is a key factor to evaluate success of organ transplantations. In Germany QA documentation is progressively developed and enforced by law. Our objective is to share QA models from Germany in a standardized format within a form repository for world-wide reuse and exchange. Original QA forms were converted into standardized study forms according to the Operational Data Model (ODM) and shared for open access in an international forms repository. Form elements were translated into English and semantically enriched with Concept Unique Identifiers from the Unified Medical Language System (UMLS) based on medical expert decision. All forms are available on the web as multilingual ODM documents. UMLS concept coverage analysis indicates 92% coverage with few but critically important definition gaps. New content and infrastructure for harmonized documentation forms is provided in the domain of organ transplantations enabling world-wide reuse and exchange.

Keywords. Transplantation, Documentation, Unified Medical Language System, Quality Assurance, Health Care

1. Introduction

Organ transplantation is a live-saving treatment when patients suffer from end-stage diseases of major organs [1, 2]. According to the German Organ Transplantation Foundation (DFO), around 7,000 organs have been transplanted[1] in Germany from 2012 to 2013; transplanted organs were kidney, liver, lung, heart, pancreas and small intestine in the order of transplantation frequency [3]. For all of those organ transplantations obligatory documentation for German quality assurance (QA) is applied except for the very rare small intestine transplants. Therefore, we will refer to the five major organs whenever the term *organ transplantation* is used.

Existing ways of quality assurance (QA) are necessary to measure, maintain and further develop the medical quality of medical procedures such as organ transplantations among different health service providers. Many hospitals provide organ transplantations in Europe, which are coordinated by Eurotransplant (ET) [4] in terms of organ allocation. However, QA is carried out by different institutes in different countries. A standardized multilingual form-based approach can help to enable harmonious and transparent data acquisition for all hospitals providing QA for transplantation services.

[1] Living donor transplants and domino transplants not counted.

Since 2000 all German health service providers are obliged by law to apply QA for certain services [5]. A distinct form-based way of QA documentation has evolved within 30 medical specialties containing more than four million patient cases, indicating outstanding efforts in German QA at international level [6]. Currently, organ transplantations are also under obligatory quality assurance in Germany. That is, administrative and clinical data of patients (organ donors and recipients) is documented mandatorily through a form-based approach, which defines all necessary data items to be collected from patients that undergo organ transplantation procedures.

Those forms were established by the German Institute for Applied Quality Improvement and Research in Health Care (AQUA) [7] (commissioned by the German Federal Joint Committee) as a long-term consensus-driven result by multidisciplinary medical specialists [5-7].

The original AQUA forms are defined in a structured way using an internal data model lacking standardization and open access to enable scientific exchange. Furthermore, language dependence hampers reuse or edition on an international level and missing semantic codes for data items make it difficult to interpret the meaning of data items from a machine-readability point of view [8]. Innovative approaches of interoperable medical forms need to be established to support implementation of those forms into different health-related information systems for harmonious structured clinical data storage and linkage to routine documentation. Regarding this aspect, it is important to use existing standard data models and semantic codes from international terminologies for data items within those forms. Thus, each data item is defined in a more structured way and linked to a medical context unambiguously such that form items get machine-readable and language-independent. The first feature is beneficial for clinical data management, data analysis and data integration, where electronic forms are used for medical documentation and form-based data retrieval [8]. The last feature enables the world-wide reuse of medical forms where QA is less developed. To achieve this, international vocabularies with a large coverage for clinically relevant medical terms need to be used such that medical terms (e.g. organ rejection reaction, Diabetes Mellitus type 2) can be mapped semantically correctly to data elements of the forms.

The objectives of this work are threefold:

1. Request and retrieve the latest German QA forms regarding organ transplantations and process them into a standardized model of study forms: Operational Data Model (ODM) [9] is an XML-based meta data model specified by the Clinical Data Interchange Standards Consortium (CDISC) [10] and supported [11] by regulatory agencies such as Food and Drug Administration (FDA) [12] and European Medicines Agency (EMA) [13]. For more information on alternative models and our preference for ODM, refer to [8].

   Translation of German into English data items and semantic coding using the Unified Medical Language System (UMLS) [14]. All forms will be accessible on a medical form repository, which has been developed, based on our previous work [14]. This way, we will provide the first medical E-forms for QA in organ transplantation with structured data items and international semantic annotations in a technical standard format and ability for international discussion, editing and reuse. Based on English translation and language-independent medical context of items provided by UMLS codes, further languages can be added.

   To give a more mature example of our methods of form sharing and multilingual support, a general follow up form has been published previously with semantic coding and English translation, upon which currently 17 languages (German, English, French, Portuguese, Spanish, Mandarin, Hindi, Japanese, etc.) have been added, see [15].

2. Common data elements and clusters of common data elements in the domain of organ transplantation will be provided. Furthermore a coverage analysis of UMLS will show the possible extent of critical concept definition gaps for medical QA in the domain of organ transplantation.

Supplement files including links to all processed forms, full data tables and large figures are available on: [16].

2. Methods

2.1. Forms retrieval

All available QA forms for organ transplantations were retrieved from the German Institute for Applied Quality Improvement and Research in Health Care (AQUA) [7]. Permission for scientific analysis and publication was granted. Supplementary AQUA files that include further filling out explanations were not taken into account.

The original forms were provided as PDF files and a data model for all forms using a relational Microsoft Access data-base [17]. Thus, structured information of every data item (as item name, datatype, measuring unit, min/max value, mapping to code lists, whether or not an item is mandatory) is available.

Altogether, 16 forms were retrieved for all organ transplantation under QA. Every German health care facility that provides any of the organ transplantations is obliged to fill out the corresponding forms. The list of form names, their corresponding organ transplantation procedure and number of data items is available in the supplement file: Forms-Transplant.xlsx on [16].

2.2. Form standardization and semantic coding

A converter was implemented to convert the original forms into ODM study forms according to the current ODM specification 1.3.2.

A manual review of all original forms was carried out by three medical experts - including one physician with certified English proficiency (IELTS Band 7.5) - to

1) translate item names from German into English and 2) apply semantic coding using the UMLS Metathesaurus (version 2014AB) based on medical expert consensus decision. Both steps were realized within the ODM format.

Due to variations in concept granularity and some existing concept duplicates within UMLS and thus posing ambiguities when mapping medical terms into UMLS codes, we make use of previously published coding principles for pre-coordinated and post-coordinated concept codes [18]. Thus, form data items were assigned to UMLS codes with medical proficiency and previously used codes will be reused wherever possible. Data items whose values can be directly inferred from previous data items within the same form will not be assigned to medical concepts. E.g. the items “Year of Implantation” and “Year of Implantation is unknown” will be coded only once. Thus, concept frequency analysis as a next step will not be distorted.

It has to be noted, that only the concept domain was regarded when mapping data items to medical concepts but not the value domain. After having assigned UMLS codes, the medical context of each data item is defined language-independently.

2.3. UMLS Coverage Analysis

During the process of semantic coding all medical concepts that could not be mapped with sufficient clinical specificity (according to coding principles, [18]) will be collected. These concepts will be marked with internal codes for identification.

2.4. Frequency Analysis of Medical Concepts

To provide a list of common data items, overall frequency of medical concepts that represent all data items was calculated by counting same UMLS codes. Cumulative frequency analysis illustrated how well all concept occurrences in all forms could be covered by only focusing on the most frequent concepts.

Clinical categories of most common concepts were manually identified to represent the most common concepts within a hierarchical heat map tree based on manual expert review. Thus, clinical categories are represented by tree nodes, which are based on hierarchical relations based on previous work [18] to cluster most common concepts in clinical trials.

3. Results

3.1. Forms standardization and UMLS Coverage analysis

All 16 original QA forms with 433 data items could be converted to 16 ODM forms. The 433 items refer to 374 medical concepts, of which 132 are unique, based on manual expert decision.

Among the 374 medical concepts occurrences, 344 could be coded via UMLS codes (concept occurrences with missing codes: 30, Coverage: 92%). The 30 concept occurrences with missing UMLS-codes were semantically annotated with internal identifiers to count them for the frequency analysis of all concepts. Table 1 represents an extract of 5 concepts with missing UMLS-codes, which we deemed critically important. The full list of concepts is available in the supplement file: Missing-Codes.xlsx on [16].

Table 1. Extract of important concepts in QA forms, which are not covered by UMLS. Frequency refers to the number of occurrences of that concept in all forms. MELD: Model for End-stage Liver Disease. Total number of concept occurrences n=374.

Concept Name	Frequency
Organ recipient/donor ID as ET-Number	15
Exceptional MELD score	2
Medical Urgency Code ET-Status	1
Domino liver transplantation	1
Weight of resected liver	1

Table 2. Top five extract of the most frequent administrative or demographic concepts, CUI: Concept unique identifier.

Concept Name	CUI	Frequency
Organ recipient/donor ID (ET Number)	-not available-	15
Date of birth	C0421451	15
Gender	C0079399	14
Facility's Section Identifier of service provider	C1547540	12
Medical specialty of service provider	C0037778	12

Table 3. Top five extract of the most frequent clinical concepts.

Concept Name	CUI	Frequency
Cause of death	C0007465	11
Diagnosis/Diagnoses	C0011900	8
Steroids (Pharmacotherapy)	C0038317	6
Azathioprine (Pharmacotherapy)	C0004482	6
Blood group (AB-classification)	C0427624	6

3.2. Frequency Analysis of coded medical concepts

Table 2 and 3 provide a top five extract of the most frequent concepts for demographic/administrative and clinical concepts, respectively. Both tables present for each concept its UMLS code (CUI) and concept frequency in all forms. For the full list of concepts and their UMLS codes, see supplement file: FrequencyAnalysis.xlsx in [16].

Figure 1 illustrates how well all concept occurrences (n=374) in all forms can be covered by only focusing on the most frequent concepts. That is, the cumulative coverage of all 132 unique concepts is shown starting with the most frequent concept "Organ recipient/donor ID (ET Number)" (rank 1, frequency = 15)[1] on the left hand side of the graph, followed by the remaining frequency-ordered concepts.

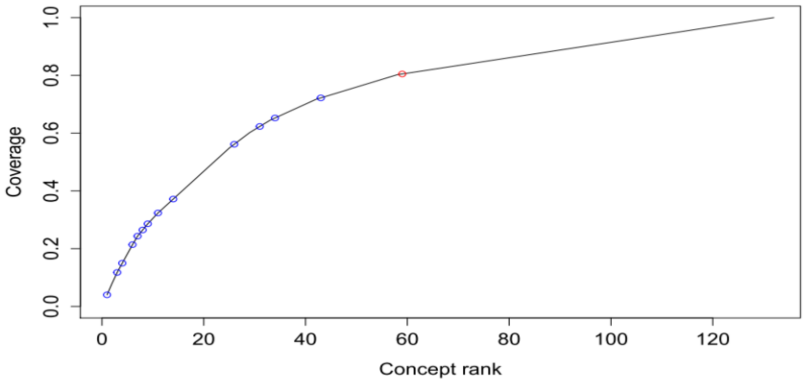

Figure 1. Cumulative frequencies, starting left with most frequent concept 'Organ recipient/donor ID (ET Number)' (rank =1, frequency= 15 => Coverage = 15/374), blue circles indicate slope reductions, red circle marks concept rank 58, after which concept frequency is only one.

[1] The concept „Date of birth" shares the same frequency of 15 and could have also assigned to rank 1

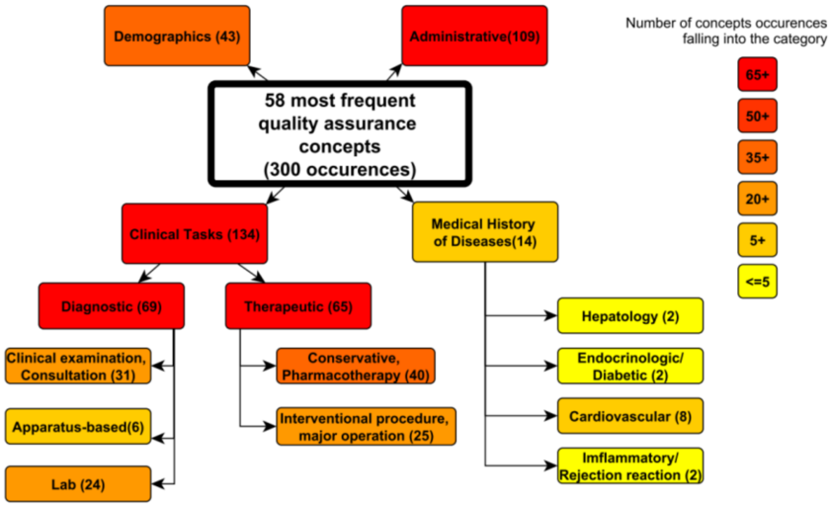

Figure 2. Categorical Heat Map Tree categorizing all 300 concept occurrences for concepts that have at least a frequency = 2. Refer to the text section for detailed description.

Blue circles indicate slope changes of the graph due to frequency reduction of the next most frequent concept. The red circle represents the 58th most frequent concept, which has an absolute frequency = 2 and reached a cumulative coverage of 80%. All other concepts on the remaining right side of the graph share an absolute frequency = 1. Hence, by only using 58 out of all 132 unique concepts (44%), 80% of all concepts occurrences in the QA-forms can be covered.

Figure 2 shows the category tree depicting clinical categories and subcategories fitting these 58 most frequent medical concepts, which occur 300 times in all forms. The color of a node indicates the number of concept occurrences within the QA forms that fall into the specific category represented by that node, similar to a heat map. Every category is annotated with a count number in parenthesis, which states the exact number of times medical concepts from the QA-forms fall into that category. Note that the count number of a parent's node doesn't need to be the sum of its child nodes' count numbers, since the parent node can contain concepts, which cannot be assigned to any of its child nodes.

Medical categories as Diagnostic/Therapeutic and Administrative encompass the most frequent medical concept occurrences for documentation of QA forms. Compared to our previous work [18] to cluster most common concepts for clinical trials the main category "Administrative" was used instead of "Trial/Research-specific concepts" to represent administrative patient-related QA data items like "Health care facility identifier" or "Zip code of discharge". For the full list showing which concept was mapped to which category, refer to the supplement file: FrequencyAnalysis.xlsx on [16].

3.3. Form Sharing for Open Access

All ODM forms are provided on a web-form portal on [14], where forms can be visualized and downloaded as native ODM files and different other structured formats, e.g. SPSS, SQL, CSV and CDA. Links to each of the 16 forms are available in the supplement file: Forms-Transplant.xlsx on [16].

4. Discussion

4.1. UMLS Coverage Analysis

As mentioned before, the original form creation was based on a long-term and expert-driven consensus process commissioned by the German Federal Joint Committee. Hence, data items of those forms have a medical significance based on expert consensus in routine documentation to assess, maintain and improve medical quality for medical treatment. Therefore, our results regarding medical concepts that represent data items of the QA forms and are not defined by UMLS indicate significant definition gaps. Even though most of the concepts occurrences could be covered, 8% remain undefined and are critically important to be incorporated into UMLS or other large medical vocabularies to make those concepts amenable for unambiguous identification, which is an important prerequisite for data harmonization and interoperability of medical data items.

4.2. Medical Concepts and Data Items of Forms

Calculated frequencies and established clustering of concepts are based on the mapping of all data items from the QA forms into medical concepts. It has to be noted that medical concepts represent the concept domain of a data item and not necessarily further information as value-domain-specific information or temporal information.

If the user is interested in which items are linked to concept codes our published form files can be searched for a given UMLS code, which is defined within a data item definition element. Furthermore a data item definition element contains information on measurement units, min and max values wherever this information was given in the original forms.

4.3. Further limitations

Expert-based manual coding of data items is a time-consuming process, application of our methods on a larger set of form data items leads to an increase of coding efforts linear to the size of the item set. Carrying out multiple manual coding by different coders is possible, however it has to be noted, that UMLS is not a classification, it is a meta-vocabulary with some concept duplicates and varying concept granularities leading to potentially different code mappings among different coders. Since we apply our methods on previously published coding principles [18] and a very limited set of different medical experts as coders with consensus-based code mappings, identification of different medical concepts is based on medical expert knowledge and the mentioned UMLS coding issues could be counteracted.

Translation was carried out by a German physician with certified English proficiency (IELTS Band 7.5), but not validated by back-translation into German.

Nevertheless, a form portal is provided to enable world-wide editions of semantic annotations and/or item translations to improve content of the forms.

5. Conclusion

The work provides the first E-forms for quality assurance for organs transplantations containing language-independent data items with semantic codes. A platform is provided for open access to enable world-wide reuse and versioning.

Frequency analysis and clusters of medical concepts that represent data items provide a set of conceptual common data elements. It could be shown, that by only using around 40% of all unique medical concepts, 80% of all concept occurrences could be covered when only focusing on the most frequent concepts.

UMLS-coverage analysis indicates concept coverage of 92% with few but critically important concept definition gaps. A UMLS update of the presented concepts is therefore strongly recommended.

References

[1] J.L. Platt, New directions for organ transplantion, Nature, 1998, 392(6679 Suppl), 11-17.
[2] P. Burra, M. De Bona, Quality of life following organ transplantation, Transplant International, 2007, 20(5), 397-409.
[3] German Organ Transplantation Foundation (DFO), http://www.dso.de/, last access: 30.01.2015.
[4] Eurotransplant (ET), https://www.eurotransplant.org/, 30.01.2015.
[5] A. Bramesfeld, G. Willms, Cross-Sectoral Quality Assurance, § 137a Social Code Book V, Public Health Forum, 2014: 22(2), 14.e1-14.e3.
[6] G. Heller, J. Szecsenyi, G. Willms, B. Broge, Quality measurement using administrative data in mandatory quality assurance, Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen, 2014, 108(8-9), 465-469.
[7] Institute for Applied Quality Improvement and Research in Health Care (Aqua), https://www.aqua-institut.de/en/home/index.html , last access: 30.01.2015.
[8] M. Dugas, F. Fritz, R. Krumm, B. Breil, Automated UMLS-based comparison of medical forms, PloS One, 2013 July 4, 8(7).
[9] Operational Data Model, http://www.cdisc.org/odm, last access: 30.01.2015.
[10] Clinical Data Interchange Standards Consortium, http://www.cdisc.org/odm, last access: 30.01.2015.
[11] U.S. Food and Drug Administration, http://www.fda.gov , last access: 30.01.2015.
[12] European Medicines Agency, http://www.ema.europa.eu/ema , last access: 30.01.2015.
[13] Unified Medical Language System®, http://www.nlm.nih.gov/research/umls, last access: 30.01.2015.
[14] Medical Data Models, https://medical-data-models.org/, last access: 30.01.2015.
[15] Follow Up Form, https://medical-data-models.org/forms/5232, last access: 30.01.2015.
[16] J. Varghese, E-Health Supplement 2015, https://drive.google.com/folderview?id=0BxfnhHTIk8tqelUwMlI2UXlnU1E&usp=sharing, last access: 30.01.2015.
[17] Microsoft Office: Access. http://office.microsoft.com/en-001/access, Accessed 2014 Dec 12, last access: 30.01.2015
[18] J. Varghese, M. Dugas, Frequency Analysis of Medical Concepts in Clinical Trials and their Coverage in MeSH and SNOMED-CT, Methods of Information in Medicine, 2015, 54(1), 83-92.