MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220042

Comparative Assessment of Completeness of CDISC Controlled Terminology

Anas Elghafari[a], Joseph Finkelstein[a]

[a] *Center for Biomedical and Population Health Informatics, Icahn School of Medicine at Mount Sinai, New York, NY, USA*

Abstract

The CDISC Controlled Terminology (CT) defines the terms that may be used to represent clinical trial data in the CDISC standards. Despite its unique importance, there has been limited systematic examination of the coverage of this terminology. In this work, we performed an assessment of the completeness of CDISC CT's coverage by comparing clinical outcomes for multiple sclerosis (MS) available in CDISC CT with two independent high-fidelity benchmarks: (1) 71 expert-selected outcomes catalogued by the National Institute of Neurological Disorders and Stroke (NINDS), and, (2) 66 common outcomes used in MS trials registered on ClinicalTrials.gov (CTG). We employed a semi-automated search and term-mapping process to identify possible CDISC equivalents to the benchmarks' measures. We found that 55% of the NINDS outcomes and 52% of the CTG outcomes are absent from the CDISC Terminology, indicating a need for expanding the terminology to take into account other established standards and real-world practice.

Keywords:

Outcome Measures, Multiple Sclerosis, CDISC

Introduction

The Clinical Data Interchange Standards Consortium (CDISC) publishes a suit of standards designed to facilitate the representation and exchange of clinical studies data. The CDISC standards include a set of foundational standards, most important of which is Study Data Tabulation Model (SDTM), intended specifically for representation of clinical trial data. At the core of CDISC standards lie the CDISC Controlled Terminology (CT). The terminology defines the terms, synonyms, and variable names that may appear in a CDISC dataset. The purpose of the terminology is to bring standardization and uniformity so that the same data elements and measures can be represented in consistent, comparable ways across different studies.

The CDISC standards have come to occupy a place of unique importance in the world of clinical trials due to being recommended by both the United States Food and Drug Administration (FDA) as well as Japan's Pharmaceuticals and Medical Devices Agency (PMDA). Yet, despite their unique importance, there has only been limited systematic examination in the literature of the limitations of these standards and the CT to provide standardized representation for clinical trial data. In this work, we aim at developing computational approaches to evaluate the coverage CDISC Terminology of the outcomes and endpoints of Multiple Sclerosis (MS).

Methods

For this evaluation we use two benchmarks of the relevant outcome measures: (1) we use an automated pipeline [1] to collect outcomes used in ClinicalTrials.gov MS studies and take the most frequent individual-level measure, giving us a benchmark of 66 outcomes (2) We utilize the Common Data Elements Project [2], from The National Institute for Neurological Disorders and Stroke (NINDS), where a set of 71 outcome measures for MS compiled by domain experts. Using these two benchmarks, we apply term mapping pipeline that indexes the CDISC Terminology and allows us to efficiently identify the possible CDISC counterpart for a given outcome, if it exists. We further verify the correctness of the mapping by manual review.

Benchmarks

Benchmark 1: Most Frequent Outcomes from ClinicalTrials.gov

An automated pipeline was used to download 2,226 MS clinical trials from ClinicalTrials.gov, parse their content, and aggregate the outcome measures used in those trials [1]. Using the output of this pipeline, manually reviewed the 100 most frequent outcomes to group related outcomes and to include only individual level outcome measures (that means excluding population aggregate measures like "survival rate". Such aggregate measures are not part of the CDISC terminology.) The result was a list of 66 measures.

Benchmark 2: NINDS's Multiple Sclerosis Outcomes and Endpoints

The National Institute of Neurological Disorders and Strokes has developed a comprehensive catalogue of Common Data Elements, including outcomes, for multiple sclerosis (among other diseases). Those data elements are readily organized by disease area and type on the NINDS website. For Multiple Sclerosis there are 71 outcomes and endpoints that we used as our second benchmark.

Search and Mapping Pipeline

We developed an automated CDISC mapping pipeline that searches through the CDISC terminology to identify the potential counterparts of an input term or data element. The pipeline uses CDISC Library API, described as "the single, trusted, authoritative source of CDISC standards metadata". The API provides the terminology packages in JSON format, providing a clear structure for the various parts (codelists, terms, synonyms, definitions) and allowing easy parsing and integration with the text processing and indexing. Overview of the steps in the pipeline is in Figure 1.

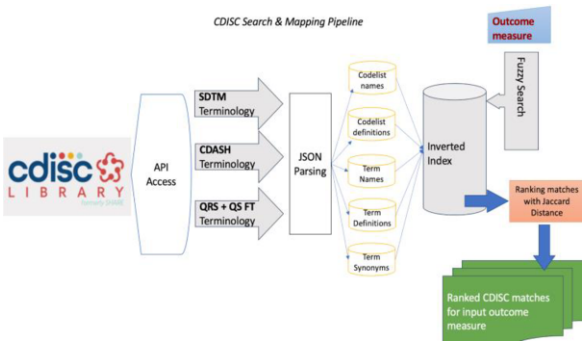

Figure 1 – CDISC Terminology Search & Mapping Pipeline

After the all terms and definitions are parsed from the API's JSON output, all the content is tokenized into words and used to create an inverted index where each token (word) is mapped to the list of Concept IDs that are relevant to this token. On top of the index, we implemented a matching and ranking mechanism that takes the input outcome measure to be mapped, tokenizes it, uses the index to get all relevant records for each token, and then ranks the results based on the string overlap between the input and possible matches. In the case of no match, that result is indicated as well.

Manual Verification of Mapping

Given that the mapping pipeline described uses fuzzy search and candidate ranking to identify possible counterparts to a given input, it is necessary to manually verify the mapping results to exclude the possibility of a mapping error. For the cases where the pipeline identified CDISC terms that correspond to the input, the verification process simply involves checking that the predicted CDISC counterpart is indeed the counterpart to the input (in the few cases where that was not the case, the correct CDISC counterpart was simply the second or third search match in the ranking). For the outcomes where the pipeline found no corresponding CDISC concept, the verification process involved manually checking the CDISC terminology files for outcome name or possible variations and abbreviations, to ensure that it is indeed missing as the pipeline results indicate.

Results

Table 1 – Quantifying CDISC Coverage of MS Outcomes

Outcomes Benchmark	Outcome	Covered in CDISC	Missing from CDISC
NINDS	71	32 (45%)	39 (55%)
ClinicalTrials.gov	66	32 (48%)	34 (52%)
Outcomes in both benchmarks	33	21 (64%)	12 (36%)
Outcomes in either benchmark	104	43 (41%)	61 (59%)

We compiled the two benchmarks as described above and applied the CDISC search and mapping pipeline. After manually verifying the mapping results, we tabulated the outcome measures from both benchmarks. Table 1 provides an overview of the CDISC terminology's coverage of outcomes from both benchmarks. Table 2 lists the 61 benchmark outcomes not found in the terminology

Table 2 – MS Outcomes Missing from CDISC Terminology

MS Outcome Measure missing from CDISC	Source
Multiple Sclerosis Spasticity Scale (MSSS-88)	NINDS
ICIQ-Lower Urinary Tract Symptoms Quality. of Life	NINDS
MOS Modified Social Support Survey (MSSS)	NINDS
Multiple Sclerosis International Quality of Life Questionnaire (MusiQoL)	NINDS & CTG
Multiple Sclerosis Quality of Life Inventory (MSQLI)	NINDS & CTG
Patient Reported Impact of Multiple Sclerosis (PRIMuS)	NINDS
Sexual Satisfaction Scale (SSS)	NINDS
Automated Neuropsychological Assessment Metrics (ANAM)	NINDS
Beery-Buktenica Developmental Test of Visual-Motor Integration (Beery VMI)	NINDS
Brief Visuospatial Memory Test Revised (BVMT-R)	NINDS
California Verbal Learning Test - Children (CVLT-C)	NINDS
California Verbal Learning Test - Second Edition (CVLT-II)	NINDS & CTG
Center for Epidemiologic Studies-Depression Scale (CES-D)	NINDS
Clinical Evaluation of Language Fundamentals - Fifth Edition (CELF-5)	NINDS
Conner's Continuous Perf. Test III	NINDS
Contingency Naming Test (CNT)	NINDS
Delis-Kaplan Executive Function System (D-KEFS)	NINDS
Expressive One-Word Picture Vocabulary Test	NINDS
Nonverbal Selective Reminding Test (NVSRT)	NINDS & CTG
Stroop Test	NINDS & CTG

Test of Memory and Learning Revised (TOMAL-2)	NINDS
Wechsler Abbreviated Scale of Intelligence (WASI)	NINDS
Wechsler Intelligence Scale for Children-V (WISC-V)	NINDS
Wide Range Assessment of Memory and Learning, Second Edition (WRAML-2)	NINDS
Wisconsin Card Sorting Test (WCST)	NINDS
Woodcock-Johnson III Test	NINDS
Grooved Pegboard Test	NINDS
Modified Ashworth Scale for Grading Spasticity	NINDS & CTG
Scale for the Assessment and Rating of Ataxia	NINDS
Activities Specific Balance Confidence Scale (ABC-Scale)	NINDS
Barthel Index	NINDS
Berg Balance Scale (BBS)	NINDS & CTG
Fatigue Scale for Motor and Cognitive Functions	NINDS & CTG
Functional Independence Measure	NINDS & CTG
MOS Pain Effects Scale (PES) Component of Multiple Sclerosis Quality of Life Inventory (MSQLI)	NINDS
Multiple Sclerosis Functional Composite	NINDS & CTG
PROMIS Fatigue Short Form	NINDS & CTG
PROMIS Pain Interference Short Form	NINDS & CTG
Short Form 12- Item Health Survey	NINDS
Time to First Relapse	CTG
Brain Atrophy by MRI	CTG
Dynamic Gait Index	CTG
Box and Block Test	CTG
Action Research Arm Test (ARAT)	CTG
2-Minute Walk Test	CTG
Patient Global Impression of Change (PGIC)	CTG
Godin Leisure-Time Exercise Questionnaire (GLTEQ)	CTG
Aerobic capacity	CTG
Manual Ability Measure-36	CTG
Four Square Step Test	CTG
Beck Depression Inventory	CTG
Cardiorespiratory fitness	CTG
Knee proprioception	CTG
Functional Reach Test	CTG
Modified Sensory Organization Test	CTG
Upper Extremity Function	CTG
Trunk Impairment Scale	CTG
Motricity Index	CTG
Brief International Cognitive Assessment for Multiple Sclerosis	CTG
Central Activation Ratio	CTG
10/36 Spatial Recall Test	CTG

Discussion

Our findings show that the further development of the CDISC terminology requires better integration with the existing standards organization, as well as the incorporation of evidence-based, data-driven approaches that can surface the outcome measures used in practice by analyzing publicly available records.

Main Result

We see that, while the CDISC Terminology covers many of the important MS outcomes, there are still crucial outcome measures that are missing. Notably, the *Berg Balance Scale* is not included in any CDISC Terminology Package despite it being one of the NINDS outcomes and among the 10 most frequent outcomes in MS studies, in addition to being relevant to many disease areas beyond MS. Other established measures that are missing include *California Verbal Learning Test*, and *MS International Quality of Life Questionnaire.* There are gaps pertaining to classes of outcomes, such as spasticity measures. Neither the *Multiple Sclerosis Spasticity Scale* nor *The Modified Ashworth Scale for Grading Spasticity* (both NINDS-catalogued) is included. During our verification process, we found that the word "spasticity" does not appear at all in any CDISC terminology code list or descriptions, indicating no measures included for it at all. A similar gap exists with regards to Ataxia measures. The effect of those gaps is that a study with any of these outcome measures would need to use ad-hoc variable names for the representation of that data, thus preventing the possibility of systemic, inter-study analysis and comparison.

Overall, the coverage of CDISC Terminology can be described as uneven. It appears from this analysis that the further development of the CDISC terminology would benefit from taking into account the CDE curation work done by scientific organizations like the NINDS, and would further benefit from incorporating an evidence-based data-driven approach that discovers the outcomes used in practice by analyzing available trial data.

Comparison with Previous Work

In our survey of the literature we have found only three previous works that involved a critical examination of CDISC standards or terminology. Of the three works, Ranallo et al. [3] is somewhat similar to this work in that it focuses on evaluating the Terminology coverage of a class of outcomes. They found most of the elements they considered to be missing from the CDISC Terminology's QS package that they examined. In addition to differences in the areas of interest, and the terminologies covered (their work also looked at the coverage in LOINC and SNOMED), there are important methodological differences: Ranallo et al. focused on a specific class of Outcome measures: Psychological Assessment Instruments, while this work considers all outcome measures for a disease-area. Secondly, the benchmark used in Ranallo et al. consisted of expert-selected instruments, while we measure against two benchmarks, and we use computational methods to identify common outcomes in practice to use as a benchmark. Lastly, their approach involves manual search and mapping, while we describe a specialized search and mapping pipeline, with manual verification for quality assurance. This automated approach facilitates casting a wider net in the search and mapping to include multiple CDISC CT packages (SDTM, CDASH, QS).

Garza et al. [5] also considered the question of data elements coverage in various standards, including CDISC's SDTM. However, their focus was not on terminology coverage of outcome measures, but rather on how well various data models could represent longitudinal community registry Electronic Health Records. Most of those data elements were in the domains of demographics (contact info, medical history, social history, allergies, etc.) and medical encounters (diagnosis, hospitalization, etc.).

Huser et al. [4] also involved a critical look at CDISC limitations. However, their focus was on the data exchange standard of the CDISC, Operational Data Model (ODM), not on any of the foundational standards or terminology. Additionally, they used a single clinical study and examined ODM's capacity to represent the protocol elements, metadata elements, and the study's Case Report Forms throughout that study's life cycle. That analysis did not include an examination of the Controlled Terminology or outcome measures.

The work we present here also utilizes ClinicalTrials.gov as a source for the evidence-based benchmark we use in our evaluation. We believe this is the first work that utilizes CTG data as a benchmark to evaluate a terminology's completeness. However, data from ClinicalTrials.gov (CTG) has found a variety of innovative uses in the medical informatics research. For example, Anderson et al. used CTG data to study level of compliance with result reporting requirements [6]. Bourgeois et al. used CTG data to compare industry-funded to non-industry-funded trials in terms of the likelihood of reporting positive outcomes [7], while Hartung et al. investigated the discrepancies between results submitted to CTG's results database and those published in peer reviewed journals [8]. With regards to outcomes and other common data elements (CDEs), Huser et al. examined the use of CDEs in real datasets and showed how the CDEs identified change by changing the threshold of commonness [9, 10]. Vodicka et al. analyzed the proportion and characteristics of CTG trials that included patient reported outcomes [11]. Luo et al. proposed a semi-automatic approach for identifying inclusion criteria CDEs [12]. This work differs in its focus on the outcome and endpoints part of the clinical trial data, and the use of those outcome measures to evaluate coverage of a controlled terminology.

Conclusions

We used a semi-automated term mapping pipeline to critically evaluate the coverage of CDISC Controlled Terminology with respect to Multiple Sclerosis outcomes. As benchmarks, we used outcome measures catalogued by NINDS as well as the outcome measures frequent on MS trials registered on ClinicalTrials.gov. Our evaluation found that over half of the benchmark outcomes are absent from the CDISC terminology. This means that the CDISC standards cannot represent those measures in a standardized uniform way that is comparable across studies, and that studies using the CDISC standards will have to resort to ad-hoc terms and variable names to represent those data elements. Thus, undermining one of the main goals of a controlled terminology. As such, we believe that further development of the CDISC terminology would benefit from taking into account other standardization efforts, and from incorporating data-driven evidence-based methods that could identify outcomes common in practice.

Acknowledgements

This work was funded in part by NIH contract UL1TR001433.

References

[1] A. Elghafari, J. Finkelstein Introducing an Ontology-Driven Pipeline for the Identification of Common Data Elements. Stud Health Technol Inform. (2020) Jun 26;272:379-382.

[2] S.T. Grinnon, K. Miller, J.R. Marler, Y. Lu, A. Stout, J. Odenkirchen, S. Kunitz. National Institute of Neurological Disorders and Stroke Common Data Element Project - approach and methods. Clin Trials. (2012) Jun;9(3):322-9.

[3] P.A. Ranallo, T.J. Adam, K.J. Nelson, R.F. Krueger, M. LaVenture, C.G. Chute. Psychological assessment instruments: a coverage analysis using SNOMED CT, LOINC and QS terminology. AMIA Annu Symp Proc. (2013) Nov 16;2013:1333-40.

[4] V. Huser, C. Sastry, M. Breymaier, A. Idriss, J.J. Cimino. Standardizing data exchange for clinical research protocols and case report forms: An assessment of the suitability of the Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM). J Biomed Inform. (2015) Oct;57:88-99.

[5] M. Garza, G. Del Fiol, J. Tenenbaum, A. Walden, M.N. Zozus. Evaluating common data models for use with a longitudinal community registry. J Biomed Inform. (2016) Dec;64:333-341

[6] M.L. Anderson, K. Chiswell, E.D. Peterson, A. Tasneem, J. Topping, R.M. Califf. Compliance with results reporting at ClinicalTrials.gov. N Engl J Med. (2015) Mar 12;372(11):1031-9

[7] F.T. Bourgeois, S. Murthy, K.D. Mandl. Outcome reporting among drug trials registered in ClinicalTrials.gov. Ann Intern Med. (2010) Aug 3;153(3):158-66

[8] D.M. Hartung, D.A. Zarin, J.M. Guise, M. McDonagh, R. Paynter, M. Helfand Reporting discrepancies between the ClinicalTrials.gov results database and peer-reviewed publications. Ann Intern Med. (2014) Apr 1;160(7):477-83.

[9] V. Huser, J.J. Cimino. Linking ClinicalTrials.gov and PubMed to track results of interventional human clinical trials. PLoS One. (2013) Jul 9;8(7):e68409.

[10] V. Huser, J.J. Cimino. Precision and negative predictive value of links between ClinicalTrials.gov and PubMed. AMIA Annu Symp Proc. 2012;2012:400-8. Epub (2012) Nov 3.

[11] E. Vodicka, K. Kim, E.B. Devine, A. Gnanasakthy, J.F. Scoggins, D.L. Patrick. Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007-2013). Contemp Clin Trials. (2015) Jul;43:1-9

[12] Z. Luo, R. Miotto, C. Weng. A human-computer collaborative approach to identifying common data elements in clinical trial eligibility criteria. J Biomed Inform. (2013) Feb;46(1):33-9.

Address for correspondence

Joseph Finkelstein, MD, PhD

Joseph.Finkelstein@mssm.edu