Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220481

Data-Driven Modeling of Randomized Controlled Trial Outcomes

Zhehuan Chen[a], Yilu Fang[a], Hao Liu[a] and Chunhua Weng[a,1]
[a] *Department of Biomedical Informatics, Columbia University, New York, NY, USA*

Abstract. Anecdotally, 38.5% of clinical outcome descriptions in randomized controlled trial publications contain complex text. Existing terminologies are insufficient to standardize outcomes and their measures, temporal attributes, quantitative metrics, and other attributes. In this study, we analyzed the semantic patterns in the outcome text in a sample of COVID-19 trials and presented a data-driven method for modeling outcomes. We conclude that a data-driven knowledge representation can benefit natural language processing of outcome text from published clinical studies.

Keywords. outcome, randomized controlled trials, knowledge representation

1. Introduction

As the volume of medical evidence expands quickly, it is imperative to enable scalable machine comprehension of medical evidence and increase its accessibility for patients, clinicians, and researchers. The Participant, Intervention, Comparator, Outcome (PICO) framework is widely adopted for retrieving medical evidence [1]. In this framework, the outcome specifies anticipated measures, improvements, or affects [2], such as "*prolongation of remission,*" "*longer survival,*" and "*blood glucose levels.*" Efforts using natural language processing (NLP) to automate outcome extraction have been growing [3][4][5]. Efforts have also been made to standardize the representation for outcomes. For example, EBM-NLP has defined outcome categories such as Physical Health, which includes Pain, Adverse Effects, and Mortality; Mental/Behavioral Impact, which includes Mental health, Participant Behavior, and Satisfaction with Care; and Non-health Outcome, which includes Quality of Intervention, and Resource Use, and Withdrawals from Study [6]. Zarin *et al.* classified outcome measures as domains, specific measurements, specific metrics used to characterize each participant's result, and methods for data aggregating for each outcome measure [7]. The taxonomy of patient outcome proposed by Wilson and Cleary is more focused on the outcomes related to or affecting health-related quality of life, with emphasis on the causal relationship between different health concepts [8]. Lin *et al.* has proposed an ontology for treatment outcome in cancer and defined high-level classes as Assessment tools, domain, Measure, Relationship, and Value type [9].

[1] Corresponding author, Dr. Chunhua Weng, Department of Biomedical Informatics, Columbia University, 622 W 168 Street PH-20, New York, NY 10027, USA; E-mail: chunhua@columbia.edu.

Despite these existing efforts, outcome text in RCT publications is still not amenable for computing due to their complexities and the fact that the aforementioned ontologies are used primarily for manual knowledge engineering rather than by automated NLP pipelines. There is an unmet need to develop an outcome text knowledge representation that is data standards-based and interoperable with NLP systems. In this paper, we present an original method to develop a data-driven knowledge representation for clinical outcomes using example COVID-19 RCT abstracts obtained from PubMed.

2. Methods

2.1. Data

We retrieved the abstracts of 50 COVID-19 RCT abstracts randomly selected using indexed metadata from the MEDLINE database. Following the standard definition from the PICO framework [1], we manually annotated the PICO statements. After removing the duplicate, vague (e.g., "*significance*" or "*common type*"), and coarse outcomes (e.g., "*reaching the primary outcome*"), and correcting grammatical errors and misclassified outcomes, we obtained 408 distinct outcome text snippets for complexity analysis.

2.2. Complexity Analysis

The complexity analysis includes syntactic and semantic complexity.

2.2.1. Syntactic Complexity

Each outcome text snippet is decomposed into phrases, where ScispaCy with the pre-trained model "en-core-sci-lg" is used for part-of-speech tagging, and the Berkeley Neural Parser (https://parser.kitaev.io/) is used for constituency parsing. To achieve the best interpretability of a biomedical concept, for outcomes that can be decomposed into phrases, we only consider the most granular noun phrases (i.e., NP) and verb phrases (i.e., VP) in them, while for the rest, the entire outcomes will be retained as phrases with their phrasal categories. We categorize and analyze the outcomes based on the number of phrases they contain and their syntactic patterns.

2.2.2. Semantic Complexity

We identify the semantic elements of the outcomes and tally the proportion of their presence and align their categories with those in the previously published outcome analyses or ontologies. For instance, "Measuring object" is adopted from the disease-treatment ontology and refers to an affected object [10]; "Specific measurement" (referring to "Outcome measurement") and "Specific metric" are selected from the PICO ontology (Cochrane PICO Ontology); "Time frame" is adopted from the ClinicalTrials.gov (https://prsinfo.clinicaltrials.gov/results_definitions.html); and "Boolean connector" and "Exclusion connector" are adopted from the complexity analysis of eligibility criteria from Ross *et al.* [11]. We create "Defining connector" and "Statistical method" semantic elements. We further extend the "Exclusion connector" to "Negation cue," referring to all possible negation cues. The explanation of each semantic element is as follows:

a. **Boolean connector**: "*and,*" "*or,*" comma, etc. Besides, "*with*" can be counted as a Boolean connector if it could be changed to AND with no information loss (e.g., "*hospitalization with intensive care*" can be converted to "*hospitalization AND intensive care*"). Also, partially specific lists can be viewed as OR statements (e.g., "*hematologic indicators including c-reactive protein*" could be converted to "*hematologic indicators OR c-reactive protein*").
b. **Negation cue**: the negation cues such as "*not*" and "*other than.*"
c. **Measuring object**: the entity or event to be measured, such as "*c-reactive protein,*" "*anxiety,*" and "*hospitalization.*"
d. **Defining connector**: the adjective, verb or prepositional phrases describing the measuring object. For example, in the outcome "*viral replication in cells infected with sars-cov-2,*" "*viral replication in cells*" is the measuring object and "*infected with sars-cov-2*" is the defining connector. In the outcome "*number of patients turning negative,*" "*number of patents*" is the measuring object, and "*turning negative*" is the defining connector.
e. **Specific measurement**: the measuring methods or instruments which can be used independently, such as "*assessment scale*" and "*6-min walk test.*"
f. **Specific metric**: the specific data (e.g., "*concentration*") for the assessment of the extent to which the outcome has been achieved.
g. **Statistical method**: the statistical method to present the result.
h. **Time frame**: temporal descriptors (e.g., "*at 15 days*" in "*clinical status at 15 days*"), references to temporal events (e.g., "*during hospitalization*") or a combination of both (e.g., "*combined adverse reactions 7 days after injection*").

3. Results

3.1. Syntactic Analysis

According to Figure 1, 38.5% of outcomes contained more than one phrase. Within this subgroup of outcomes, 67.5% of the outcomes follow the syntactic pattern "NP + NP." There can be as many as 7 phrases in an outcome. For outcomes grouped by their number of phrases, the most frequently observed syntactic pattern corresponds to outcomes that contain multiple noun phrases. Some examples are provided in Table 1. According to Table 2, we identified that elements "Measuring object", "Specific metrics" and "Specific measurement" can each be outcomes independently, while the remaining 38.5% of the outcomes all contained more than one semantic element. The most common semantic element combination was the <u>measuring object with specific metric</u>. The specific metric specifies the data to collect for the outcome. It either came right after the measuring object, such as "*antibody level*" (measuring object: "*antibody,*" specific metric: "*level*") and "*lymphocyte count*" (measuring object: "*lymphocyte,*" specific metric: "*count*"), or before the measuring object with a preposition in between, such as "*concentrations of multiple inflammatory molecules*" (measuring object: "*multiple inflammatory molecules,*" specific metric: "*concentrations*"). The second and third common semantic element combinations were the <u>measuring object</u> <u>with a time frame</u> (e.g., "*7-day adverse reactions*") and <u>with a defining connector</u> (e.g., "*coronavirus nucleic acid from throat and nasal swab,*" where the measuring object is "*coronavirus nucleic acid*" and the defining connector is "*throat and nasal swab*"), respectively. Furthermore, 46 (11.3%) of the outcomes were with more than two semantic elements.

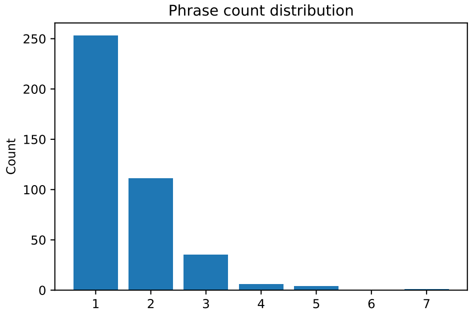

Figure 1. Phrase count distribution.

Table 1. Statistics of the syntactic patterns of the outcomes and their examples.

Phrase Count	Syntactic Patterns	Outcome Count (%)	Examples
1	NP	238 (58.3%)	*depression, symptom severity, t-cell counts*
	VP	10 (2.5%)	*subsequently died, discharged alive from hospital, cough, intubated*
	AdjP	3 (0.7%)	*not hospitalized, rash or itchy, severely ill*
2	NP + NP	106 (26.0%)	*clinical status at 11 days, duration of icu stay, ordinal scale of disease severity, time to fever resolution*
	NP + VP	5 (1.2%)	*time to discharge, case confirmed, patient withdrew*
3	NP + NP + NP	33 (8.1%)	*time to clinical improvement within 28 days, improvement from baseline of two points, change in symptom severity over 14 days*
	NP + NP + VP	2 (0.5%)	*number of patients turning negative, time from randomization to discharge*
4	NP + NP + NP + NP	6 (1.5%)	*time from starting the medication until discharge from hospital, percentages of patients with detectable viral rna at various time points*
5	NP + NP + NP + NP + NP	4 (1.0%)	*median number of days from symptom onset to start of study treatment*
7	NP + NP + NP + NP + NP + NP + NP	1 (0.2%)	*time from randomization to either an improvement of two points on a seven-category ordinal scale or discharge from the hospital*

3.2. Semantic Analysis

Table 2. Semantic patterns of the outcomes with one or two elements. For those outcomes with more than two semantic elements, they will be assigned to all possible semantic pattern groups with two elements.

Semantic Pattern	Outcome Count (%)
Measuring object alone	228 (55.9%)
Measuring object & Specific metric	130 (31.9%)
Measuring object & Time frame	46 (11.3%)
Measuring object & Defining connector	30 (7.4%)
Specific metric & Statistical method	24 (5.9%)
Specific measurement alone	19 (4.7%)
Measuring object & Boolean connector	15 (3.7%)
Measuring object & Negation cue	7 (1.7%)
Specific metric & Specific measurement	7 (1.7%)
Measuring object & Specific measurement	5 (1.2%)
Specific metric alone	4 (1.0%)

4. Discussion and Conclusion

A significant portion of outcomes contains more than one phrase or one semantic element, indicating the syntactic and semantic complexity in outcome text and implying the need to simplify outcome text. Approximately 38.5% of the outcomes contained more than one phrase, and almost the same portion of outcomes contained more than one semantic element, indicating phrase segmentation may potentially help reduce both the syntactic complexity and the semantic complexity of outcomes and improve the accuracy of their knowledge representations. For example, by phrase segmentation, the outcome "*respiratory secretion at day 4*" can be decomposed into two noun phrases "*respiratory secretion*" and "*at day 4*", where the former is semantically a measuring object and can be mapped to the standard observation concept "*Respiratory secretion*", and the latter is a time frame. This study has limitations. We only analyzed COVID-19 RCT abstracts. The generalizability of the semantic patterns to other disease domains remains to be tested. Second, the analyzed RCT abstracts are of a relatively small sample size. More evaluations are needed to test the completeness of knowledge in our representation. The PICO statements and the outcomes' semantic elements were manually annotated in this study; however, we have developed a parser to support the automatic recognition of the PICO statements [3]. A tool for semantic element extraction will be further developed.

References

[1] Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club. 1995;123(3):A12-3.
[2] Davies KS. Formulating the evidence based practice question: a review of the frameworks. Evidence Based Library and Information Practice. 2011;6(2):75-80.
[3] Kang T, Zou S, Weng C. Pretraining to Recognize PICO Elements from Randomized Controlled Trial Literature. Stud Health Technol Inform. 2019;264:188-92.
[4] Kehl KL, Xu W, Lepisto E, Elmarakeby H, Hassett MJ, Van Allen EM, et al. Natural Language Processing to Ascertain Cancer Outcomes From Medical Oncologist Notes. JCO Clin Cancer Inform. 2020;4:680-90.
[5] Humbert‐Droz M, Izadi Z, Schmajuk G, Gianfrancesco M, Baker MC, Yazdany J, et al. Development of a natural language processing system for extracting rheumatoid arthritis outcomes from clinical notes using the national RISE registry. Arthritis Care & Research. 2022.
[6] Nye B, Li JJ, Patel R, Yang Y, Marshall IJ, Nenkova A, et al., editors. A corpus with multi-level annotations of patients, interventions and outcomes to support language processing for medical literature. Proceedings of the conference Association for Computational Linguistics Meeting; 2018: NIH Public Access.
[7] Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials. gov results database—update and key issues. New England Journal of Medicine. 2011;364(9):852-60.
[8] Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life: a conceptual model of patient outcomes. Jama. 1995;273(1):59-65.
[9] Lin FP, Groza T, Kocbek S, Antezana E, Epstein RJ. Cancer Care Treatment Outcome Ontology: A Novel Computable Ontology for Profiling Treatment Outcomes in Patients With Solid Tumors. JCO clinical cancer informatics. 2018;2:1-14.
[10] Khoo CS, Na J-C, Wang VW, Chan S. Developing an ontology for encoding disease treatment information in medical abstracts. DESIDOC Journal of Library & Information Technology. 2011;31(2).
[11] Ross J, Tu S, Carini S, Sim I. Analysis of eligibility criteria complexity in clinical trials. Summit on Translational Bioinformatics. 2010;2010:46.