TOCSOC: A temporal ontology for comparing the survival outcomes of clinical trials in oncology

Deendayal Dinakarpandian, Michaela Liedtke,
Mark A. Musen
Department of Medicine, Stanford University
Stanford, CA
dinakar@stanford.edu

Bhavish Dinakar
College of Chemistry, University of California, Berkeley
Berkeley, CA

***Abstract*—The outcome of clinical trials for cancer is typically summarized in terms of survival. However, different trials for the same disease may use different measures of survival, or use differing vocabulary to refer to the same outcome measure. This makes it harder to automate an objective comparison of treatments. We propose a temporal ontology of survival outcome measures that a) helps to standardize the vocabulary for reporting survival outcomes and b) makes it possible to automatically rank the relative efficacy of different treatments. The approach has been illustrated by examples from the oncology literature. The temporal ontology and the accompanying reasoner are freely available on Github (https://github.com/pdddinakar/TOCSOC).**

Keywords—temporal ontology; survival outcome; oncology; clinical trials; reasoning

I. Introduction

The outcome of clinical trials for cancer is often summarized in terms of survival. This may be a rate, for example a 5-yr survival of 50% or a duration, for example a median survival time of 4 years. Ideally, if all potential treatments for a specific cancer were compared in terms of a common metric, it would be straightforward to rank them in terms of their effectiveness. In reality, clinical trials often use a wide variety of survival outcome measures. The scientific, ethical and pragmatic reasons for this heterogeneity are listed below:

A. Variation in study design. Long term studies may use survival measures over longer periods of times than short term studies.

B. Differences in life expectancy. Life expectancy after diagnosis varies greatly among cancers. For instant, the 5-yr survival rate for malignant melanoma exceeds 90% but is less than 20% for lung cancer (1). Thus, studies to improve the treatment might seek to look at longer time periods for melanoma compared to lung cancer.

C. Tracking disease control. For cancers that are incurable, the pragmatic goal is sometimes to retard its progress. In such cases, progression-free survival rather than measures of mortality may be used as a metric to capture phases of stable disease.

D. Consolidating gains in therapy. In contrast to incurable cancers, the availability of highly effective treatments for some cancers makes it possible to induce longer periods of remission (potentially a cure) where there is no evidence of disease. Rather than measures of mortality, measures like disease free survival are useful in such cases.

E. Limited recruitment and retention in studies. Patients are prone to drop out of studies, particularly in cancer. Progressive attrition of participants sometime forces investigators to use short term measures to report outcomes rather than wait for the originally planned longer term measures. For example, 2 or 3 yr. survival statistics might be reported instead of 5 yr. statistics.

F. Early termination on ethical grounds. If a therapy is highly successful compared to standard therapy, a decision to terminate the study and publish early might be made. Conversely, if the treatment itself causes unacceptable harm to trial participants, the trial may be terminated prematurely. In both cases, measures of shorter term survival may be included in the corresponding publication.

Even when the same survival measure is used, different studies use different terms to refer to the same concept, and different papers use the same term to refer to differing outcome measures. Oncologists typically use their expert knowledge to resolve these ambiguities and evaluate the relative merits of different therapies. This could be in the context of drafting best practice guidelines or for individualized patient care.

This paper proposes the use of a temporal ontology of terms for summarizing the results of clinical trials in oncology. The use of an ontology can reduce the ambiguity in specifying results. Additionally, the inclusion of temporal relationships within the ontology can help partially automate the comparison between treatments whose effectiveness has been summarized with different but related measures. We first describe the source of the vocabulary and the process to create the temporal ontology. This is followed by a description of the reasoning used to rank treatments for a specific cancer. We give examples from real world data and conclude with a discussion of limitations and future plans.

II. CREATION OF THE TEMPORAL ONTOLOGY

Overall survival (OS) is a commonly used measure of the effectiveness of cancer therapy. It is defined as the length of time from either the date of diagnosis or the start of treatment that patients are still alive (2). In other words, such a commonly used term has two different interpretations that is obvious only to a human reader. We searched the Bioportal (3) collection of ontologies for a perfect match to the term "Overall survival." The following four independent resources include OS as a term: "National Cancer Institute Thesaurus (NCIT) (4)," "Experimental Factor Ontology (EFO) (5)," "Cancer Care: Treatment Outcome Ontology (CCTOO) (6)" and "Interlinking Ontology for Biological Concepts (IOBC) (7)." As CCTOO (6) is specific to cancer treatment, we selected this ontology for further exploration.

Out of a total of 1133 terms in the ontology, we found 35 terms (First column in Table 1) containing the token "survival," which were scattered throughout the ontology. CCTOO is based on IS_A and IS_ASSESSED_BY relationship between terms. In contrast, our goal was to create a temporal ontology with the relationship NOT_GREATER_THAN (NGT) between the terms. The rationale for this is the fact that many events in cancer outcomes that precede another could also be simultaneous. For example, though several symptoms (events) of cancer may not be fatal, the timing of some symptoms may coincide with death.

An exhaustive approach to determine if an NGT relationship exists between every pair of terms would require 595 comparisons. In order to this more efficiently, we first sorted the terms based on their suffixes to group related concepts together - the terms were reversed, sorted based on the reversed strings and reversed again to obtain the original terms. This procedure resulted in a sorted list of terms (Second column in Table 1), such that neighboring terms sharing suffixes were more likely to have a temporal relationship with each other. For example, the first five terms in the second column in Table 1 are all survival rates, and all types of "Progression-free survival" are grouped together.

These were manually checked and arranged into a hierarchical list, where each indent corresponds to the NGT relationship. Since definitions were missing for most of the CCTOO terms, we referred to the following resources, in order, to establish and add the meanings of the terms: NCI dictionary (2), the NCI Outcome Measures Glossary (8, 9), the DATECAN initiative (10) and finally Pubmed (11) searches for papers containing the terms. We edited the hierarchy based on the following criteria:

A. Highly specific terms were removed, e.g., Breast cancer specific survival. Since the intended use of the proposed temporal ontology is in the context of a specified disease, it is redundant to explicitly include disease names in the names of survival measures.

B. Synonyms were merged together, e.g., "Disease-free survival" was chosen as the canonical term for "Relapse-free survival."

C. Ambiguous terms not useful for comparing durations or rates were removed, e.g., Long term survival. Since time duration is expected to be explicitly stated in summarizing an outcome, "Long term survival" is not a useful concept to standardize.

TABLE I.

TERMS FROM CCTOO	SUFFIX SORTED TERMS
Distant recurrence-free survival	Disease free survival rate
Biochemical relapse-free survival	Relapse-free survival rate
Long term survival	Progression-free survival rate
Local relapse-free survival	Event-free survival rate
Event-free survival rate	Overall survival rate
Invasive disease-free survival	Breast cancer specific survival
Failure-free survival	Disease-specific survival
Metastasis-free survival	Prostate cancer-specific survival
Overall survival rate	Regional recurrence free survival
Treatment-free survival	PSA progression free survival
Distant failure-free survival	Symptomatic skeletal event free survival
Locoregional failure-free survival	Recurrence-free survival
PSA progression free survival	Local recurrence-free survival
Overall survival	Distant recurrence-free survival
Disease-specific survival	Failure-free survival
Progression-free survival	Locoregional failure-free survival
Symptomatic skeletal event free survival	Distant failure-free survival
Local progression-free survival	Disease-free survival
Distant disease-free survival	Invasive disease-free survival
Immune-related progression-free survival	Biochemical disease-free survival
Radiographic progression-free survival	Distant disease-free survival
Relapse-free survival	Relapse-free survival
Progression-free survival rate	Biochemical relapse-free survival
Event-free survival	Local relapse-free survival
Disease-free survival	Progression-free survival
Clinical progression-free survival	Radiographic progression-free survival
Local recurrence-free survival	Immune-related progression-free survival
Regional recurrence free survival	Biochemical progression-free survival
Biochemical progression-free survival	Clinical progression-free survival
Prostate cancer-specific survival	Local progression-free survival
Relapse-free survival rate	Metastasis-free survival
Disease free survival rate	Treatment-free survival
Recurrence-free survival	Event-free survival
Biochemical disease-free survival	Overall survival
Breast cancer specific survival	Long term survival

D. A clear distinction between period and rate was made. It is common practice in publications to use the term "survival" to refer to both a duration of time, e.g., median survival time and a rate, e.g., proportion alive after a period of time has elapsed. The reader has to infer this from the context. However, this distinction needs to be explicit in an ontology. Therefore, we added the suffix "time" to all terms to indicate the first interpretation and the suffix "rate" to all terms to indicate the first interpretation.

E. Missing terms were added, e.g., only 5 'rate' terms were present in CCTOO. A corresponding 'rate' term was created for each 'time' term.

The resulting temporally related hierarchy contains 44 terms related by NOT_GREATER_THAN (NGT) relationships. These consist of 22 concepts expressed as both

durations (Fig. 1) and rates (only the first few rows as shown at bottom of figure for brevity). The full version is available as an OWL file created with the help of Protégé (12). While the distinction between rate and time may be clear to a human reader from the context, it is necessary to separate these concepts for machine interpretation. Also, since the motivating goal is to compare treatments, definitions of the concepts Note that the terms "Overall survival time (OS)" and "Disease-specific survival time (DSS)" are in bold on the far right as the deepest concepts. These refer to the longest periods. All terms are NGT DSS, and OS is NGT DSS. This is because OS is agnostic of health or treatment status, while DSS is longer because it excludes deaths from causes unrelated to the disease or its treatment. At the other extreme, "Treatment-free survival time" has the shortest duration and has an NGT relationship with all terms; cancer is likely to return earliest when all treatments, including maintenance, are discontinued. The final hierarchy was checked for accuracy by author M.L., who is an oncologist.

Fig. 1. The TOCSOC temporal hierarchy.

```
Treatment-free survival time
    Failure-free survival time
        Distant failure-free survival time
        Regional failure-free survival time
        Local failure-free survival time
    Disease-free survival time
        Event-free survival time
            Invasive disease-free survival time
                Symptomatic skeletal event-free
                survival time
            Biochemical disease-free survival time
            Recurrence-free survival time
                Distant recurrence-free survival time
                Regional recurrence-free survival
                time
                Local recurrence-free survival time
                Locoregional recurrence-free survival
                time
            Progression-free survival time
                Radiographic progression-free
                survival time
                Biochemical progression-free survival
                time
                Clinical progression-free survival
                time
                Local progression-free survival time
                    Overall survival time
                        Disease-specific survival
                        time

Treatment-free survival rate
    Failure-free survival rate
……
……
```

III. TEMPORAL REASONING FOR TREATMENT COMPARISON

The temporal ontology shown in Fig. 1 may be interpreted as longer time durations from left to right. This temporal ordering of types of survival outcomes can be exploited based on the following key TOCSOC reasoning principle:

Consider treatments T1 and T2 with respective outcome measures O1 and O2, such that O1 has an NGT relationship with O2. If the observed value of O1 is at least as large as the value of O2, then T1 is likely better than T2.

We present several representative cases below to illustrate specific scenarios of reasoning derived from the general TOCSOC rule.

A. Identical measure with different values. If treatments *x* and *y* have overall survival times (often reported as medians) of 5 and 6 years respectively, then it is trivial to conclude that *y* is better than *x*. Now consider a treatment *p* for the same type of cancer where the group was followed for only 5 years, at which point more than half the subjects were still alive. This is usually referred to as median not reached, implying that the overall survival for this group is greater than 5. This implies that *p* is likely better than *x*, but not guaranteed to be better than *y*.

B. Measures of same type but differing in duration. If treatment x results in a 5-yr OS rate of 80% while treatment *y* results in a 4-yr OS rate of 70%, then *x* is better than *y*.

C. Temporally related measures. This is the specific scenario that TOCSOC was envisioned to handle. If treatment *x* results in a (median) progression-free survival (PFS) of 5 years and treatment y results in a median OS of 4 years, then *x* has an OS of at least 5 years (inferred from TOCSOC) and is therefore better than *y*.

D. Comparing rates with periods. When available, survival times should be compared with survival times and rates with rates. However, it may sometimes be necessary to compare rates with times. This is possible to a limited extent. Measures that end with "survival time" are typically the median survival time within a group. For example, if 4 subjects with treatment *x* have survival times {1,2,4,5}, then (median) survival time with treatment *x* is 3 years. This may be interpreted as a survival rate of 50% at 3 years. To be strictly correct, this corresponds to a survival rate of at most 50% since the median for survival times {1,3,3,3} is also 3, even though this is also the maximum survival time; there are no survivors past 3 years.

E. Replicate measures. Different studies may report different outcomes for the same treatment. One option to deal with this situation is to use an average value for each treatment that is weighted by the size of the replicate studies. Another option is to compare treatments based on a bounded range of reported performances, though this is likely to underestimate the difference between treatments.

F. Indeterminable comparisons. Sibling terms (successive terms at the same level of indentation in Fig. 1) are uncomparable by definition. For example, "Biochemical progression-free survival time" may be greater than "Clinical progression-free survival time" in some individuals, but the other way around in others. Even when comparable, it is hard to reach a conclusion if one treatment has an OS of 90 % at 2 years and an alternative treatment has a PFS of 50% at 4 years. A plethora of data can also paradoxically lead to an inconclusive result. If multiple metrics are available for each treatment, then rankings might be different or even reversed based on choice of metric. The pragmatic strategy for this is to

report all rankings along with the rationale, thus serving more as an objective summary of evidence than a ranker.

Based on the above considerations, we implemented a reasoner that takes a temporal ontology and a set of treatments with corresponding survival outcomes as input, and outputs a ranking of treatments. The survival outcome input is specified as either a rate (time period of observation and proportion) or a duration (survival time). The temporal ontology is represented internally as directed acyclic graph in an adjacency matrix. A second directed graph is created corresponding to the ranking of treatments. In silent mode, only unambiguous rankings are returned. In verbose mode, undeterminable rankings (cycles in the graph) are also included in the output. Since the ontology is read dynamically, the reasoner can be used with alternate versions of ontologies based on NGT relationships.

IV. ILLUSTRATIONS FROM LITERATURE

Consider the results of two treatments (the exact details are not relevant) for high risk multiple myeloma shown in the table below:

Trial Reference	Treatment	Disease	Metric	Value
(13)	A	HRMM	OS 5-yr	55%
(14)	AA+B	HRMM	OS 4-yr	54%

Since treatment AA+B has a 4-yr OS that is lower than the 5-yr OS for treatment A, it cannot be better than treatment A.

Now consider the following comparison of treatment AA+B with AAsib that exploits the structure of TOCSOC. The observed outcome for AAsib corresponds to 50% OS at 4.25 years. Since the 4-yr PFS for AA+B is 52%, we can conclude that the 4-yr OS for AA+B is significantly higher than 52% (OS is typically considerably higher than PFS in most cases) and therefore better than AAsib.

Trial Reference	Treatment	Disease	Metric	Value
(14)	AA+B	HRMM	PFS 4-yr	52%
(15)	AAsib	HRMM	Median OS	4.25 yrs.

V. LIMITATIONS & FUTURE PLANS

We have shown the value of recasting an existing ontology into one based on temporal relationships for comparing the effectiveness of different treatments for cancer. This can help rank different treatments for each cancer, especially as multiple new treatments are increasingly becoming available for several cancers. However, it is important to acknowledge that this approach only ranks treatments; it is far from a treatment 'recommender.' Several other considerations often drive choice of therapy. A treatment with a shorter survival time may be selected for reasons of toxicity, cost or patient age. A treatment that is better at preventing distant recurrences than local recurrences may be preferred. The result of comparing a set of treatments may not be valid because of heterogeneity of the underlying disease. Despite diligent efforts to conduct randomized clinical trials, study populations often turn out to contain a mixture of cancers at the molecular level. For improving the rationale of decision making, advances in disease subtyping also need to be taken into account. Each study is likely to have selection biases, both known and unknown in its choice of subjects. While treatment outcomes are often summarized as an average estimate of effectiveness, it is important to take into account the confidence intervals of estimates when comparing them. Further, expanded individual profiles are likely to be taken into account in the era of personalized and molecular medicine.

The present study could be improved in terms of both the ontology employed and the power of the reasoner. This paper restricted itself to using terms from a pre-existing ontology in the useful but narrow perspective of 'survival.' As medical care improves to the point where many more cancers are curable, temporal metrics for the quality of life are likely to become more important. Further, different types of cancer may use specialized metrics to evaluate outcomes. As terms are used more consistently in the literature, more precise temporal relationships could be used. While using a detailed temporal ontology like the W3C OWL Time Ontology (16) would be overkill, it would be helpful to add a few more relationships, e.g., STRICTLY_LESS_THAN could be added where applicable. As such, the first version of TOCSOC is best viewed as an upper ontology. More terms can be incorporated by mining trials registered at sites like "clinicaltrials.gov" for primary and secondary endpoints that have temporal dependencies, some of which may be specific only to certain cancers.

The reasoner is currently conservative in being largely deterministic; it could be enhanced by a Bayesian mode that takes into account prior distributions of the outcomes as well as the temporal relationship between them. Instead of point estimates, full distributions could be taken into account to combine multiple weak signals into more robust evidence for rankings.

ACKNOWLEDGMENTS

This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

REFERENCES

1. Surveillance, Epidemiology, and End Results Program [Internet]. 1. Surveillance, Epidemiology, and End Results Program [Internet]. Available from: https://seer.cancer.gov/.
2. NCI Dictionary of Cancer Terms [Internet]. Available from: https://www.cancer.gov/publications/dictionaries/cancer-terms.
3. Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, et al. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software

applications. Nucleic Acids Res. 2011;39(Web Server issue):W541-5.
4. NCI Thesaurus [Internet]. Available from: https://ncit.nci.nih.gov/ncitbrowser/.
5. Malone J, Holloway E, Adamusiak T, Kapushesky M, Zheng J, Kolesnikov N, et al. Modeling sample variables with an Experimental Factor Ontology. Bioinformatics. 2010;26(8):1112-8.
6. Lin FP-Y, Groza T, Kocbek S, Antezana E, Epstein RJ. The Cancer Care Treatment Outcomes Ontology (CCTO): A computable ontology for profiling treatment outcomes of patients with solid tumors. Journal of Clinical Oncology. 2017;35(15_suppl):e18137-e.
7. Kushida T, Kozaki K, Tateisi Y, Watanabe K, Masuda T, Matsumura K, et al., editors. Efficient construction of a new ontology for life sciences by subclassifying related terms in the Japan Science and Technology Agency thesaurus. International Conference on Biomedical Ontology; 2017.
8. National Cancer Institute. Outcome Measures Glossary [Available from: https://wiki.nci.nih.gov/display/CRF/Outcome+Measures+Glossary.
9. Punt CJ, Buyse M, Kohne CH, Hohenberger P, Labianca R, Schmoll HJ, et al. Endpoints in adjuvant treatment trials: a systematic review of the literature in colon cancer and proposed definitions for future trials. J Natl Cancer Inst. 2007;99(13):998-1003.
10. Gourgou-Bourgade S, Cameron D, Poortmans P, Asselain B, Azria D, Cardoso F, et al. Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials)dagger. Ann Oncol. 2015;26(5):873-9.
11. PubMed [Internet]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/.
12. Musen MA, Protege T. The Protege Project: A Look Back and a Look Forward. AI Matters. 2015;1(4):4-12.
13. Bjorkstrand B, Iacobelli S, Hegenbart U, Gruber A, Greinix H, Volin L, et al. Tandem autologous/reduced-intensity conditioning allogeneic stem-cell transplantation versus autologous transplantation in myeloma: long-term follow-up. J Clin Oncol. 2011;29(22):3016-22.
14. Green DJ, Maloney DG, Storer BE, Sandmaier BM, Holmberg LA, Becker PS, et al. Tandem autologous/allogeneic hematopoietic cell transplantation with bortezomib maintenance therapy for high-risk myeloma. Blood Adv. 2017;1(24):2247-56.
15. Giaccone L, Storer B, Patriarca F, Rotta M, Sorasio R, Allione B, et al. Long-term follow-up of a comparison of nonmyeloablative allografting with autografting for newly diagnosed myeloma. Blood. 2011;117(24):6721-7.
16. W3C. Time Ontology in OWL 2017 [Available from: https://www.w3.org/TR/owl-time/.