C-TrO: An Ontology for Summarization and Aggregation of the Level of Evidence in Clinical Trials

Olivia SANCHEZ-GRAILLET [a,1], Philipp CIMIANO [a], Christian WITTE [a] and Basil ELL [a]

[a] *Cluster of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University, Germany*

Abstract. Evidence-based medicine requires that medical decisions are taken based on available, verified and high quality evidence, in particular in the form of randomized clinical trials (RCTs). Such evidence is mainly obtained from multiple relevant published studies and synthesized in systematic reviews and meta-analyses. Therefore, aggregating and summarizing the level of evidence across different clinical studies are relevant tasks in the medical context. Towards the development of a system that aggregates, compares and rationalizes upon the evidence from multiple clinical trials, we developed an ontology that models such studies. The ontology considers clinical trial elements as well as other pieces of evidence and their relationships, which support the aggregation and rationalization of the level of evidence across studies. On the basis of this conceptualization of clinical trials, we also obtained a knowledge base for the extraction of evidence via SPARQL queries, and an annotation scheme for annotating clinical trial publications. We validated our ontology through a case study on glaucoma, in which the ontology proved to be able to answer the competency questions required for aggregating and rationalizing the results across different studies.

Keywords. ontology, clinical trials, knowledge base, evidence aggregation, evidence-based medicine

1. Introduction

In paradigms that do not consider evidence (e.g., eminence-based medicine) there may exist a risk of bias caused by the lack of a robust statistical analysis, conflict of interest, or decisions made based on personal experience only. In order to avoid this risk, the current paradigm for medical decision making requires evidence-based reasoning [1,2]. Thus, decisions should be made based on available, verified and high quality evidence. Such evidence is typically obtained from randomized clinical trials (RCTs), which are considered as the gold-standard for clinical research.

[1]Corresponding Author: Olivia Sanchez-Graillet, Cluster of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University, Inspiration 1, 33619 Bielefeld, Germany; E-mail: olivia.sanchez@uni-bielefeld.de.

In order to decide which treatment could be comparatively more effective in terms of efficacy and safety relying on high quality evidence, it is necessary to collect evidence from multiple relevant studies, whilst considering the possible presence of bias and the shortcomings of those studies. Therefore, clinicians need to synthesize and compare the aggregated information, which is normally done in the form of systematic reviews and meta-analyses. These meta-studies and reviews try to consider all the evidence contained in multiple controlled clinical studies and related publications available in open-access databases such as PubMed[2]. However, reviewing the vast number of publications requires an extensive manual effort. Besides the fact that this effort is costly in terms of time and money, the corresponding reviews are typically outdated at the time of publication. For these reasons, there is a great interest in automatizing the generation of systematic reviews [3,4,5,6]. This automatization requires to capture the outcomes and values of multiple studies formally in a machine-interpretable way. Furthermore, when formalizing the results from clinical studies to support cross-study aggregation, it is necessary to consider several aspects, such as the use of different terminologies, different units, incomplete information, different ways of reporting results, etc. Therefore, ontologies that are able to represent the characteristics and results of clinical studies on a semantic level, abstracting from the original publication and such that comparison and aggregation of evidence is supported, are needed.

With the aim to support the automatization of systematic reviews and meta-analyses, which implies the aggregation, comparison and rationalization of the level of evidence from multiple clinical trials, we present the clinical trial ontology (C-TrO) which models not only the PICO[3] elements, but also other relevant pieces of evidence that are necessary for carrying out these tasks.

The remainder of this paper is structured as follows: Section 2 describes the design and development of C-TrO, whilst Section 3 describes its structure. The validation of C-TrO through a case study is provided in Section 4. An example of the use of C-TrO in the annotation of text is given in Section 5. Section 6 offers final remarks and conclusions.

2. Ontology Design and Conceptualization

In order to establish the requirements and the suitable development methodology for our ontology, we first describe the objectives of C-TrO. Afterwards, we describe the corresponding methodological aspects.

2.1. Uses and Requirements of the Ontology

The main goals of our ontology are: 1) provide the structure for a knowledge base that stores the information obtained from published clinical trials and other related information sources; 2) provide the logical structure from which it is possible to summarize and aggregate the level of evidence from multiple trials through SPARQL queries; and 3) support an annotation scheme of concepts and their relationships in clinical trial publications. The manually annotated corpus will be used for training/testing an IE system

[2]https://www.ncbi.nlm.nih.gov/pubmed

[3]PICO stands for P: population or patients and problem, I: intervention, C: comparison and O: outcome.

that will extract the concepts and relations from clinical trial abstracts and automatically populate the knowledge base with this information.

Consequently, C-TrO is meant to describe any type of clinical trial for any health condition, and to take into account important information such as risk of bias, results according to a given aggregation method, relative or absolute risks, size of effect, and any other information that could support rationalized decision-making as to which medical treatments are comparatively more efficacious or/and safer than others.

Besides considering the PICO framework, which describes the basic information structure of clinical trials, the structure of the ontology must facilitate the comparison of different interventions across multiple studies by considering *clinical arms* as intervention groups. We shortly review the PICO elements as follows.

2.1.1. PICO Elements

Clinical arms are groups of patients (or any other unit of study) that receive one or more clinical interventions according to a given protocol. *Population* refers to the participants in the clinical trials who have a certain health condition and receive the medical interventions. An *intervention* is a medical treatment applied to a group of participants with the aim to combat diseases or other health disorders. *Comparison* refers to the comparison between two interventions, in which usually one of them serves as control (e.g. placebo). *Outcomes* are the results of the analysis relative to the measures obtained after the intervention(s) applied to an arm (e.g. statistical analysis and size of effect calculations) for the given outcome indicators (or endpoints), i.e., variables that pertain to the clinical trial objective. Outcomes can be primary or secondary outcomes. Primary outcomes refer to the most relevant endpoints that are expected to change as the result of the application of the intervention(s). While secondary outcomes are used for evaluating additional effects.

2.2. Methodology

Our ontology has to comply with both software engineering and text annotation aspects, as well as to serve as a basis for storing evidence from multiple clinical studies in a knowledge base. This information should be efficiently stored and retrieved. Thus, the information should be structured and logical, but also intuitive and coherent for the users (both annotators and clinicians). Consequently, the expressivity of the ontology should not have a high level of complexity and inference and should allow an efficient access to the information. Therefore, we decided to employ the macro-level development methodology *On-ToKnowledge* [7], which consists of five phases: feasibility study, kickoff, refinement, evaluation and application, and evolution. Unlike other methodologies, *On-ToKnowledge* is suitable for our work, since it considers the iterative refinement and evaluation of the ontology, which allows to obtain a most complete ontology structure. Besides, the development is driven by other processes that consider both software engineering and human aspects.

The specification of the ontology requirements of the kickoff phase includes the description of what the ontology should be competent to answer, the identification of relevant knowledge sources for the ontology description, the definition of concepts and relations, the hierarchical structure of the ontology, and the consideration of existing ontologies to be reused.

The competency questions specified in the kickoff phase are used along the development of the ontology as well as in the evaluation and refinement phases.

The ontology was implemented with *Protégé V. 5.2.0* [4] and was validated by executing queries on the knowledge base that answer the corresponding competency questions for a case study on glaucoma.

2.2.1. Knowledge Sources

In order to design the ontology, we analyzed abstracts and full articles that describe clinical trials and meta-analyses of different diseases and health conditions. More specifically, we concentrate on glaucoma and type 2 diabetes mellitus. We analyzed 40 abstracts of the glaucoma corpus [8] and 40 abstracts of diabetes and a pair of meta-analysis in glaucoma obtained with the PICO linguist tool[5]. Three systematic reviews on diabetes were provided by a medical practitioner expert in that field. It is possible to generalize to other diseases since the abstracts follow similar formats to report the trials and their results.

On the basis of the analysis of these publications, we identified and conceptualized the main entities mentioned in the studies together with their respective relationships. For example, we observed that an arm can have one or more interventions and that an intervention can have more than one medication or a combination of medications, or other types of therapies (e.g. psychotherapy). The assessment of the quality of the evidence was analyzed following the GRADE [9,10] criteria[6].

2.2.2. Competency Questions

Competency questions are of great importance for defining the scope of the conceptualization of the ontology [11]. They also help in the iterative evaluation of the ontology in the different development phases. The competency questions that our ontology should be able to answer are related to the aggregation of the available evidence across multiple clinical trials to determine superiority of the interventions.

We agreed upon the critical questions based on basic reasoning patterns for superiority (in terms of efficacy and safety) of the interventions that were derived from our analysis. In such patterns a set of premises yield to a conclusion. For example, in a reasoning pattern for superiority on efficacy, a major premise expresses the general objective of the primary outcome of the intervention, and a minor premise considers the magnitude of the differences between the intervention results. We also considered questions that would challenge the conclusion in the reasoning patterns. For example, questions about the statistical significance of the results, or the size of the population that receives the intervention. Some of the specified competency questions are the following:

- *CQ1* Which studies report that the intervention using drug1 is more effective than the one using drug2 in reducing (increasing) a given outcome indicator for a certain disease?
- *CQ2* In which studies, is the reduction (increment) of the output indicator statistically significant?

[4] https://protege.stanford.edu/

[5] https://babelmesh.nlm.nih.gov/pico.php

[6] http://www.gradeworkinggroup.org/

CQ3 In how many studies was it observed that a given drug intervention produced a given adverse effect?

CQ4 How reliable is the evidence from these studies on basis of their risk of bias or trial design?

Further competency questions can be formulated by combining the different elements of the ontology according to the preferences of the expert users. For example, the first competency question can become more specific if we add population's country, gender, age range and ethnicity. Thus, the resulting question would turn into:

> "Which studies report that the intervention using drug1 is more effective than the one using drug2 in reducing (increasing) a given outcome indicator for a certain disease in a population that resides in a certain country, composed by only men (women) with ages between x and z and a given ethnicity?"

2.3. Related Ontologies

The RCT Schema ontology [12], which was developed in an Ocelot frame-based format[7], models different phases of the clinical trial life cycle (e.g., trial design, protocol management and protocol execution). This ontology has a very comprehensive structure for capturing clinical trial information. However, it does not include some relevant properties for determining intervention superiority, such as the "direction of the outcome" which indicates whether the final value is a reduction or an increment from the baseline. This is an important information for the training of our IE system, since in published clinical trials it is uncommon that the baseline value is reported. Instead, it is mentioned whether the final value is a reduction or an increment from the baseline.

The Cochrane PICO ontology [13] is used for the annotation of three types of PICO models identified in the Cochrane[8] reviews. These annotations are meant to support the formulation of questions and search of clinical trials. This ontology does not include a more fine-grained information such as the frequency and interval of the treatments or the risk of bias and the quality of the evidence.

The OCRe ontology (Ontology of Clinical Research) [14,15] is a formal model that uses the OWL2 language to represent the different phases of clinical research, like the study design, the execution, and the interpretation of data. OCRe focuses on the abstract description of the studies rather than the quantitative representation of the results of the respective clinical studies.

The above mentioned ontologies have different granularity of representation. However, a higher level of detail would be needed in some aspects in order to compare the results of different studies and aggregate them into a coherent summary of the level of evidence. Table 1 presents a comparison of the characteristics useful for supporting the aggregation of the level of evidence of the described ontologies and C-TrO. It can be observed that C-TrO shares some similarities with the other ontologies, in particular with aspects of the PICO ontology that are concerned with the respective PICO elements.

[7] Frames are equivalent to classes and slots to attributes.

[8] The Cochrane Foundation: https://www.cochrane.org/

Table 1. Comparison of characteristics of clinical trial ontologies useful for supporting the aggregation of the level of evidence.

Ontology / Characteristics	**RCT Schema**	**PICO Ontology**	**OCRe**	**C-TrO**
Main use	Preparation of reports and analysis of RCTs.	Annotation of Cochrane Reviews according to its PICO models.	Indexing of research data across different clinical data resources.	Knowledge base and annotation schema for the aggregation and rationalization of the level of evidence of clinical trials.
Considered PICO elements (classes)	AnalyzedPopulation, Intervention-Arm: allows multiple interventions per arm, Outcome: different types of outcomes, including Side-effects.	InterventionGroup (arm), Interventions and Outcomes related through PICO_Comparison: only descriptive results. AdverseEffect class.	ArmPopulation, InterventionStudy Protocol: is divided into Arms (and Epocs), Outcome: is an analysis specification. Adverse effects are not modeled.	Arms are related to Population and allow multiple Interventions, which can have different Outcomes (primary, secondary and adverse effects).
Level of granularity	Detailed. However, the object properties are not clear.	Detailed. However, the outcome results are descriptive rather than numeric values.	Very-detailed. However, the outcome results are descriptive rather than numeric values.	Detailed. It includes numeric values for outcome results in different formats: absolute, relative or countable values.
DL expressivity (in Protégé)	Not specified	ALCO(D)[a]	ALCROIQ(D)	ALCOF(D)
Considered risk of bias/evidence quality aspects	Some aspects: in Trial-Details (e.g. fraud and stopping details).	No	No	Yes, in EvidenceQuality (e.g. GRADE rate, conflict of interest, etc.)
Availability	pont format[b] Restricted access to the formalized clinical trials.	Turtle format (ttl). Restricted access to the ontology.	Open access to the schema. No formalization of clinical trials is available.	Open access to the schema. The formalization of clinical trials is in progress.

[a] Aproximate expressivity considering the ontology scheme at https://linkeddata.cochrane.org/pico-ontology
[b] http://rctbank.ucsf.edu/home/trialreporting/rct-schema-1

2.4. Alignment with Other Ontologies

Even though the current version of C-TrO does not integrate any external ontology, C-TrO allows the alignment of relevant medical ontologies, (e.g. SNOMED[9]) for diseases

[9]https://www.nlm.nih.gov/healthit/snomedct/index.html

or health conditions and drugs. Other terminologies or classification models could also be integrated, such as the International Code for Diseases ICD-11[10] or the Anatomical Therapeutic Chemical (ATC) Classification System[11]. At the moment, the respective disease and drug Ids are added during the knowledge base curation phase since these identifiers are not included in the published studies. In future work, we will consider cross-linking C-TrO to the Evidence and Conclusion Ontology (ECO)[12] in which evidence is established as a type of information that supports an assertion (a statement about something). Such a procedure would be compatible with our work on reasoning patterns for superiority of interventions, in the sense that the conclusions would be equivalent to the ECO assertions and the premises would be similar to the ECO supporting evidence for the assertions.

3. Description of the Ontology Structure

In this section we describe the main classes and relations of C-TrO, whose structure is depicted in Figure 1.

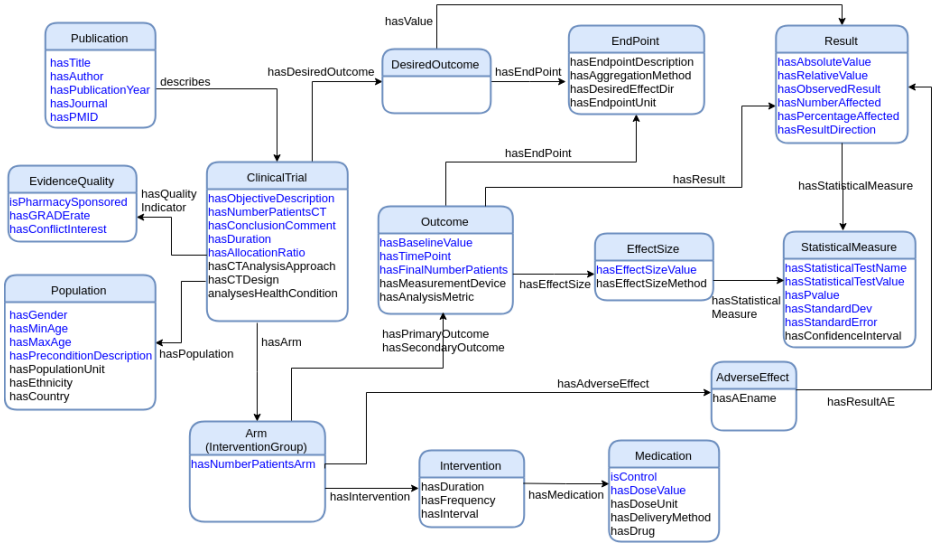

Figure 1. Diagram of the main classes of C-TrO: Data properties are in blue and Object properties in black. The arrows start in the domain classes and end in the range classes.

A *clinical trial* is described in terms of its objective, the number of patients involved, its duration, and its conclusion. A clinical trial analyses *health conditions* such as diseases, disorders, or syndromes and has a *desired outcome* that is an *endpoint* described in terms of a desired value for each measurement. The clinical trial is related to a certain *population* with attributes such as gender, ethnicity, country of residence, minimum age,

[10] https://www.who.int/classifications/icd/en/
[11] https://www.whocc.no/atc_ddd_index/
[12] http://evidenceontology.org/

maximum age, and description of preconditions. A clinical trial has a number of *arms* (Intervention groups). Each arm has a given number of patients who receive one or more interventions. Each arm defines one or more *interventions* such as the administration of a *medication* in a certain dose with a certain delivery method[13]. Arms have *outcomes* that can be primary or secondary. Outcomes are specified in terms of an analysis metric, baseline value, time points, measurement devices, and an ideal (desired) value of the outcome. An outcome has an *endpoint*, which is an outcome indicator with an aggregation method, measurement unit, a desired effect direction (e.g., whether a measure should be decreased by the intervention), and the final number of patients. Outcomes have *results* which can be qualitative or quantitative (e.g. absolute values, relative values and countable values), and the direction of the result (i.e., reduction or increment). Arms (Intervention groups) also have *adverse effects*, which are reported qualitatively or quantitatively in the results. Results have *statistical measurements* that include *p*-value for a given type of statistical test, a confidence interval, a standard deviation, a standard error, etc. Furthermore, the outcome can be described in terms of its *effect size*, which is in turn described in terms of an effect size method and value. A *publication* is described with publication meta-data and links to the clinical trial that it presents. A clinical trial may present *Evidence Quality* indicators, such as risk of bias, to be pharmacy sponsored, or to have a GRADE rate.

3.1. Completeness

Whilst C-TrO does cover the main aspects for modeling clinical trials and for rationalizing recommendation (i.e. those aspects relative to the PICO elements and the quantitative results of the interventions), its structure will remain open to modifications in order to be able to account for new methodologies, diseases and drugs or any other kind of emerging evidence.

Regarding the completeness of the information in the knowledge base, the fact that abstracts follow a similar reporting structure that the one proposed by the CONSORT standards[14] (e.g. title, authors, trial design, methods, results, conclusions, trial registration and funding), facilitates the identification of several important evidential information in clinical trial abstracts. However, it is important to mention that some information may not be included in the abstracts. For instance, information that denotes quality indicators, such as conflict of interest (e.g."The author(s) have no proprietary or commercial interest") and funding is rarely reported.

3.2. Upper Ontology

C-TrO does currently not rely on an upper ontology. However, in order to improve the interoperability, open use, and collaborative development of C-TrO, we will consider using an upper-ontology such as the Basic Formal Ontology (BFO)[15], which is part of the Open Biological and Biomedical Ontology (OBO) Foundry collection[16]. In this respect, further work towards the creation of adequate mappings will be necessary.

[13] Interventions are not restricted to drug medications. They could be of other types (e.g. physiotherapies) that could be added to C-TrO.

[14] http://www.consort-statement.org/

[15] http://basic-formal-ontology.org/

[16] http://obofoundry.org/

4. Case Study on Glaucoma

In order to validate the ability of C-TrO to answer the required competency questions, we decided to focus on the health condition "glaucoma", which has previously been studied by other evidence mining approaches [8,16] and which involves a clear primary outcome. Glaucoma is a disease that damages the optic nerve and that can lead to permanent visual loss. The damage of the optic nerve usually occurs when the internal pressure in the eye (IOP) increases[17]. Therefore, the reduction of IOP is a desired outcome in an intervention for glaucoma.

For this case study, we consider the RCTs compared in the meta-analysis carried out by Zhang et al. [17], in which two widely used drugs for treating glaucoma, latanoprost and timolol, are compared. The aim of the meta-analysis was to evaluate and compare the efficacy and the tolerance (or safety) of the two drugs. The quality of the evidence was assessed by considering the design characteristics of the clinical trials such as masking, randomization, etc. The main outcome measures studied were the percentage of IOP reduction for efficacy, and the relative risk for side effects (e.g. hyperaemia, conjunctivitis, etc.). The meta-analysis suggested that latanoprost was more effective than timolol in lowering IOP. However, it was observed that latanoprost often caused iris pigmentation. It was also found that latanoprost once daily evening regime was superior compared to latanoprost once daily morning regime.

We formalized these clinical trials in the knowledge base derived from C-TrO, which is encoded in RDF triples. Figure 3 in Appendix A shows an example of the RDF triples corresponding to a clinical trial on glaucoma, one of its arms and one of the arm's interventions.[18]

4.1. Answering Competency Questions

We used SPARQL[19] queries to retrieve information from the knowledge base for answering the corresponding competency questions. The competency question *CQ1* has been reformulated for our case study on glaucoma as:

"Which studies report that the intervention using latanoprost is more effective than the timolol treatment in reducing the diurnal mean IOP for glaucoma?"

The answer to this question is shown in Table 2 and the corresponding SPARQL query is presented in Appendix B.1. The average mean for latanorpost treatments is 7.3 mmHg and for timolol treatments 5.65 mmHg.

The retrieved information shows that in eleven comparable clinical trials, the *latanoprost* treatment reduced the *diurnal mean IOP* from the baseline in greater magnitude than the *timolol* treatment. This suggests that under the given circumstances[20], the latanoprost treatment is more effective compared to the timolol treatment in reducing the diurnal mean IOP.

[17] For more details about glaucoma visit https://www.glaucoma.org

[18] The formalized clinical trials used in this case study are available at: http://scdemo.techfak.uni-bielefeld.de/clintrials/clintrials.owl

[19] https://www.w3.org/TR/rdf-sparql-query/ C-TrO to an upper-ontology

[20] These circumstances can be changed by retrieving other information from C-TrO as required (e.g. dose, frequency, interval, etc.)

Table 2. Answer to the competency question CQ1 relative to the reduction of diurnal mean IOP (CT_n is the clinical trial identifier).

CT_Id	Reference	Mean IOP reduction by Latanoprost (mmHg)	Mean IOP reduction by Timolol (mmHg)
CT_1	Alm A et al,1995	7.8	6.7
CT_1	Alm A et al,1995	8.6	6.7
CT_10	Nicolela MT et al.,1996	6.8	5.3
CT_11	Drance SM et al.,1998	3.6	3.1
CT_2	Aquino MV et al.,1999	11.1	9.1
CT_3	Camras CB et al.,1996	6.7	4.9
CT_4	Diestelhorst M et al.,1998	4.9	2.1
CT_5	Mastropasqua L et al,1999	4.8	4.6
CT_6	Mishima HK et al.,1996	6.2	4.4
CT_7	Rulo AH et al.,1994	8.9	5.9
CT_8	Watson P et al,1996	8.5	8.3
CT_9	Diestelhorst M et al.,1997	9.8	6.7
Average Mean		7.3	5.7

Table 3 shows the answer to the competency question *CQ2* reformulated for the case of glaucoma as:

"In which studies is the diurnal mean IOP statistically significant?".

Only those *p*-values that were reported in the clinical studies are retrieved. In our example, only four studies reported the p-values of the respective interventions. The corresponding SPARQL query is shown in Appendix B.2.

Table 3. Answer to the competency question CQ2 relative to the statistical significance (p-values) of the reduction of diurnal mean IOP.

CT_Id	Latanoprost intervention	p-value	Timolol intervention	p-value
CT_2	CT2_Intervention1	< 0.001	CT2_Intervention2	< 0.001
CT_3	CT3_Intervention1	< 0.001	CT3_Intervention2	< 0.001
CT_5	CT5_Intervention1	< 0.001	CT5_Intervention2	< 0.001
CT_9	CT9_Intervention1	< 0.001	CT9_Intervention2	< 0.001

5. Example of the Use of C-TrO for Text Annotation

As mentioned before, one of the purposes of C-TrO is to guide the annotation of the evidence found in clinical trial publications in order to create a training corpus for an IE system. Therefore, hundred of abstracts of clinical studies will be manually annotated to form such corpus. Once the IE system is trained, it will automatically extract similar information (i.e., C-TrO concepts and relations) and populate the knowledge base with this information. Figure 2 shows an example of an abstract annotated with the *SANTO* annotation tool [18] that can be configured to follow the C-TrO structure. Thus, single annotated entities can be grouped into predefined slots (i.e. class instances) according to the C-TrO classes. The annotated text can be saved in an RDF format that formalizes

the annotated entities and their relationships (i.e. data properties and object properties), and in an annotation format that contains the span positions of the individual entities and indicates to which composed entities (i.e. instances) they belong. Figure 4 in Appendix C shows the annotation of group Publication_1 (i.e., an instance of Publication) in an "annotated format" file and in the corresponding RDF file it is indicated that Publication_1 *describes* ClinicalTrial_1 (i.e., an instance of ClinicalTrial).

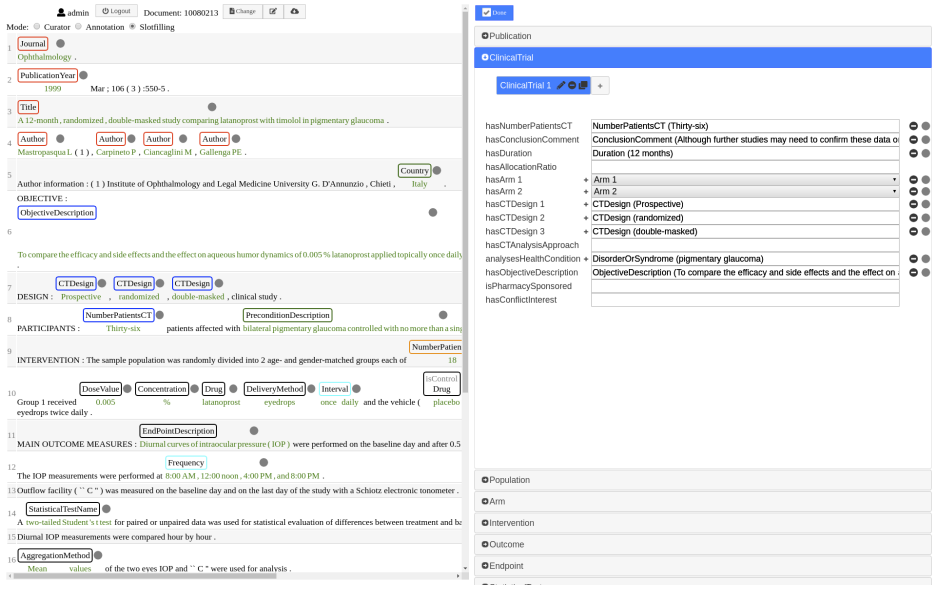

Figure 2. Example of the annotation of a clinical trial with the SANTO tool (The numbers on the left side of the text indicate the sentence number).

6. Conclusions

We have presented C-TrO, an ontology for modeling clinical trial information for the aggregation and rationalization of evidence. The development of C-TrO considered both human and system development aspects. Some concepts and relationships of C-TrO are similar to the ones of existing ontologies for clinical studies. However, C-TrO has a more detailed level of representational granularity in order to aggregate and rationalize the evidential information from multiple studies. We have shown that C-TrO is able to answer competency questions related to the aggregation of evidence to determine and rationalize the superiority of interventions in terms of efficacy and safety. In future work, we plan to continue using C-TrO for guiding the annotation of clinical trials abstracts. The set of annotated abstracts will form a training corpus for an IE system that identifies pieces of evidence and the relationships between them. We intend to improve the interoperability of C-TrO by means of an upper ontology. In due course, we expect C-TrO to constitute the backbone of a system that aggregates and summarizes the level of evidence from clinical trials, and that generates clinical recommendations. Our ontology, knowledge base and annotated corpus will be provided as open-access resources.

Acknowledgments

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG) within the project Rationalizing Recommendations (RecomRatio) as part of the Priority Program "Robust Argumentation Machines" (RATIO) (SPP-1999).

References

[1] Sackett, D. L., et al.: Evidence based medicine, *BMJ*, 313(7050), (1996), pp. 170.
[2] Sackett, David L., et al.: Evidence based medicine: what it is and what it isn't, *BMJ*, 312(71), (1996), pp. 71–72.
[3] Tsafnat, G., Dunn, A., Glasziou, P., Coiera, E.: The automation of systematic reviews, *BMJ*, 346(f139), (2013).
[4] O'Connor, A.M., Tsafnat, G., Gilbert, S.B., et al.: Moving toward the automation of the systematic review process: A summary of discussions at the second meeting of international collaboration for the automation of systematic reviews (ICASR), *Systematic Reviews*, 7(1), (2018).
[5] Murad, M.H., Montori, V.M.: Synthesizing evidence: Shifting the focus from individual studies to the body of evidence, *JAMA*, 309(21), (2013), pp. 2217–2218.
[6] Bastian, H., Glasziou, P., Chalmers, I.: Seventy-five trials and eleven systematic reviews a day: How will we ever keep up?, *PLoS Medicine*, 7(9), e1000326, (2010).
[7] Sure Y., Staab, S., Studer, R.: On-to-knowledge methodology (OTKM). *Handbook on ontologies*, Springer, Berlin, Heidelberg, (2004), pp. 117–132.
[8] Trenta, A., Hunter, A., Riedel, S.: Extraction of evidence tables from abstracts of randomized clinical trials using a maximum entropy classifier and global constraints. CoRR, abs/1509.05209, 2015.
[9] Guyatt, G.H., Oxman, A.D., Kunz, R., et al.: What is "quality of evidence" and why is it important to clinicians?, *BMJ*, 336(7651), (2008), pp. 995–998.
[10] Balshem, H., Helfand, M., Schünemann, H.J., et al.: GRADE Guidelines: 3. Rating the quality of evidence, *J Clin Epidemiol.*, 64(4), (2011), pp. 401–406.
[11] Uschold, M. ,King, M.: Towards a methodology for building ontologies. *Workshop on Basic Ontological Issues in Knowledge Sharing in IJCAI-95*, Montreal, Canada, (1995).
[12] Sim, I., Olasov, B., Carini, S.: An ontology of randomized controlled trials for evidence-based practice: Content specification and evaluation using the competency decomposition method, *J Biomed. Inform.* 37(2), (2004), pp. 108–119.
[13] Mavergames, C., Oliver, S., Becker, L.: Systematic reviews as an interface to the web of (trial) data: Using PICO as an ontology for knowledge synthesis in evidence-based healthcare research, *Proc. of the 3rd Workshop on Semantic Publishing (SePublica)*, (2013), pp. 22–26.
[14] Sim, I., Tu, S.W., Carini, S., et al.: The ontology of clinical research (OCRe): An informatics foundation for the science of clinical research, *J Biomed. Inform.*, 52, (2014), pp. 78–91.
[15] Tu, S.W., Carini, S., Rector, A., et al.: OCRe: an ontology of clinical research, *11th Intl. Protégé Conf.*, Amsterdam, Netherlands, (2009).
[16] Mayer, T. , Cabrio, E. and Villata, S.: Evidence type classification in randomized controlled trials, *Proc. of the 5th Workshop on Argumentation Mining*, (2018).
[17] Zhang, W.Y., Po, A.L.W., Dua, H.S., Azuara-Blanco, A.: Meta-analysis of randomised controlled trials comparing latanoprost with timolol in the treatment of patients with open angle glaucoma or ocular hypertension, *Br J Ophthalmol* 85(8), (2001), pp. 983–990.
[18] Hartung, M., ter Horst, H., Grimm, F., Diekmann, T., Klinger, R., Cimiano, P.: SANTO: a web-based annotation tool for ontology-driven slot filling, *Proc. of ACL 2018, System Demonstrations*, (2018), pp. 68–73.

Appendix A Knowledge Base

```
:CT_3 rdf:type ctro:ClinicalTrial ;
   :hasObjectiveDescription "Latanoprost, a new prostaglandin..." ;
   :hasConclusionComment "Latanoprost has the potential..." ;
   :hasAnalysisApproach PreProtocol ;   :hasArm Arm_31, Arm_32 ;
   :hasPopulation :CT3_Population ;   :hasCTDesign :DoubleBlind, :Randomized .
:Arm_31 rdf:type ctro:Arm ;
   :hasNumberPatients 134 ;    :hasIntervention :CT3_Intervention1 .
   :hasPrimaryOutcome :CT3_A1_OC1 ;   :hasAdverseEffect :CT3_A1_AE1 ;
:CT3_Population rdf:type ctro:Population ;
   :hasPreconditionDescription "Ocular hypertension and glaucoma" .
   :hasMinAge 30 ;       :hasMaxAge 90 ;
   :hasCountry :USA ;   :hasGender "Mixed" ;
:CT3_Intervention1 rdf:type ctro:Intervention ;
   :hasFrequency "Once_at_evening";
   :hasInterval "Daily" ;      :hasDuration "3 months";
   :hasAnalysisMetric "ChangeFromBaseLine" ;
   :hasDesiredEffectDirection "Reduction";   :hasMedication :CT3_I1_M1 .
:CT3_A1_OC1  rdf:type ctro:Outcome ;
   :hasEndpoint :EndPoint_CT3_A1_OC1 ;
   :hasAggregationMethod "Mean" ;
   :hasBaselineValue  25.3 ;     :hasBioAndMedUnit :mmHg ;
   :hasResult  :Result_CT3_I1_OC3 .
:EndPoint_CT3_A1_OC3 rdf:type ctro:EndPoint ;
    :hasEndpoint Description :Diurnal_IOP .
:Result_CT3_I1_OC3 rdf:type ctro:Result ;
    :hasResultValue 6.7 .
:CT3_I1_M1 rdf:type ctro:Medication;
    :hasDrug :Timolol; :hasDoseValue 005;
    :hasBioAndMedUnit "Percent";   :hasDeliveryMethod "Eyedrops".
```

Figure 3. Excerpt of an RDF representation of a clinical trial from C-TrO (Note: A dose of 0.005% timolol solution corresponds to one eyedrop.)

Appendix B SPARQL Queries

B.1 Query for Retrieving the IOP Mean Reduction

The following query retrieves pairwise comparisons of the diurnal IOP reduction values from baseline of the interventions in each clinical trial in the case study on glaucoma.

```
SELECT DISTINCT ?ct ?reference ?interv1 ?reduction1 ?interv2 ?reduction2
  WHERE{
{
  ?medic1 :hasDrug :Latanoprost.      ?medic2 :hasDrug :Timolol.
  ?interv1 :hasMedication ?medic1.    ?interv2 :hasMedication ?medic2.
  ?arm1 :hasPrimaryOutcome ?outcome1. ?arm2 :hasPrimaryOutcome ?outcome2.
  ?outcome1 :hasEndPoint ?endpoint1.  ?outcome2 :hasEndPoint ?endpoint2.
  ?endpoint1 :hasEndpointDescription  :Diurnal_IOP.
  ?endpoint2 :hasEndpointDescription  :Diurnal_IOP.
  ?outcome1 :hasResult ?res1.         ?outcome2 :hasResult ?res2.
  ?res1 :hasAbsoluteValue ?result1.   ?res2 :hasAbsoluteValue ?result2.
  bind(str(?result1) as ?reduction1)  bind(str(?result2) as ?reduction2)
```

```
  ?arm1 :hasIntervention ?interv1.    ?arm2 :hasIntervention ?interv2.
  ?ct :hasArm ?arm1.                  ?ct :hasArm ?arm2.
  ?pub :describes ?ct.                ?pub rdfs:label ?reference.
  FILTER (?result1 > ?result2)
}UNION{
  ?medic1 :hasDrug :Latanoprost.      ?medic2 :hasDrug :Timolol.
  ?interv1 :hasMedication ?medic1.    ?interv2 :hasMedication ?medic2.
  ?arm1 :hasPrimaryOutcome ?outcome1. ?arm2 :hasPrimaryOutcome ?outcome2.
  ?outcome1 :hasEndPoint ?endpoint1.  ?outcome2 :hasEndPoint ?endpoint2.
  ?endpoint1 :hasEndpointDescription  :Diurnal_IOP.
  ?endpoint2 :hasEndpointDescription  :Diurnal_IOP.
  ?outcome1 :hasResult ?res1.         ?outcome2 :hasResult ?res2.
  ?res1 :hasAbsoluteValue ?result1.   ?res2 :hasAbsoluteValue ?result2.
  bind(str(?result1) as ?reduction1)  bind(str(?result2) as ?reduction2)
  ?arm1 :hasIntervention ?interv1.    ?arm2 :hasIntervention ?interv2.
  ?ct :hasArm ?arm1.                  ?ct :hasArm ?arm2.
  ?pub :describes ?ct.                ?pub rdfs:label ?reference.
  FILTER (?result1 < ?result2)
}} ORDER BY ?ct
```

B.2 Query for Retrieving P-values

This query retrieves the statistical significance (p-values) of the diurnal IOP reductions of the interventions in the case study on glaucoma.

```
SELECT DISTINCT ?ct ?interv1 ?pvalue1 ?interv2 ?pvalue2
 WHERE{
{
  ?medic1 :hasDrug :Latanoprost.   ?medic2 :hasDrug :Timolol.
  ?interv1 :hasMedication ?medic1. ?interv2 :hasMedication ?medic2.
  ?arm1 :hasIntervention ?interv1. ?arm2 :hasIntervention ?interv2.
  ?arm1 :hasPrimaryOutcome ?outcome1. ?arm2 :hasPrimaryOutcome ?outcome2.
  ?outcome1 :hasEndPoint ?endpoint1.  ?outcome2 :hasEndPoint ?endpoint2.
  ?endpoint1 :hasEndpointDescription :Diurnal_IOP.
  ?endpoint2 :hasEndpointDescription :Diurnal_IOP.
  ?outcome1 :hasResult ?res1. ?outcome2 :hasResult ?res2.
  ?res1 :hasAbsoluteValue ?result1.  ?res2 :hasAbsoluteValue ?result2.
  ?res1 :hasStatisticalMeasure ?st1. ?res2 :hasStatisticalMeasure ?st2.
  ?st1 :hasPValue ?pvalue1.    ?st2 :hasPValue ?pvalue2.
  ?ct :hasArm ?arm1. ?ct :hasArm ?arm2.
  FILTER (?result1 > ?result2)
} UNION{
  ?medic1 :hasDrug :Latanoprost.
 ?medic2 :hasDrug :Timolol.
  ?interv1 :hasMedication ?medic1. ?interv2 :hasMedication ?medic2.
  ?arm1 :hasIntervention ?interv1. ?arm2 :hasIntervention ?interv2.
  ?arm1 :hasPrimaryOutcome ?outcome1. ?arm2 :hasPrimaryOutcome ?outcome2.
  ?outcome1 :hasEndPoint ?endpoint1.  ?outcome2 :hasEndPoint ?endpoint2.
  ?endpoint1 :hasEndpointDescription :Diurnal_IOP.
  ?endpoint2 :hasEndpointDescription :Diurnal_IOP.
  ?outcome1 :hasResult ?res1. ?outcome2 :hasResult ?res2.
  ?res1 :hasAbsoluteValue ?result1.  ?res2 :hasAbsoluteValue ?result2.
  ?res1 :hasStatisticalMeasure ?st1. ?res2 :hasStatisticalMeasure ?st2.
  ?st1 :hasPValue ?pvalue1.    ?st2 :hasPValue ?pvalue2.
  ?ct :hasArm ?arm1. ?ct :hasArm ?arm2.
FILTER (?result1 < ?result2)
}} ORDER BY ?ct
```

Appendix C Annotated File

Figure 4 shows an excerpt of an annotated file according to the SANTO's annotation schema, which is formed by the comma-separated fields indicated in the first row: *AnnotationID* is an id number assigned to each annotation, *ClassType* is the C-TrO class to which the annotated entity belongs, *DocCharOnset* and *DocCharOffset* are the first and last position of the annotated text (i.e., the annotation span), *Text* is the annotated text (i.e., entity), *Meta* is an optional comment added to the annotation, and *Instances* are the instances of the C-TrO classes containing individual annotations as properties. The RDF file contains the triples formed by the annotated groups which represent domains and ranges linked by the respective properties.

```
#AnnotationID,ClassType, DocCharOnset(incl),DocCharOffset(excl),Text,Meta,Instances
1, Journal, 0, 13, "Ophthalmology", "", "<http://ctro/data/Publication_1>
   <http://ctro/data/hasJournal > \"Ophthalmology\"."
2, PublicationYear, 16, 20, "1999", "", "<http://ctro/data/Publication_1>
   <http://ctro/data/hasPublicationYear> \"1999\"."
3, Title, 46, 149, "A 12-month , randomized , double-masked study comparing
    latanoprost with timolol in pigmentary glaucoma", "",
    "<http://ctro/data/Publication_1> <http://ctro/data/hasTitle > \ "A 12-month ,
    randomized , double-masked study comparing latanoprost with timolol in
    pigmentary glaucoma\"."
4, Author, 152, 166, "Mastropasqua L", "", "<http://ctro/data/Publication_1>
    <http://ctro/data/hasAuthor > \"Mastropasqua L\"."
5, Author, 175, 186, "Carpineto P", "", "<http://ctro/data/Publication_1>
   <http://ctro/data/hasAuthor > \"Carpineto P\"."
6, Author, 189, 202, "Ciancaglini M", "", "<http://ctro/data/Publication_1>
   <http://ctro/data/hasAuthor > \"Ciancaglini M\"."
7, Author, 205, 216, "Gallenga PE", "", "<http://ctro/data/Publication_1>
    <http://ctro/data/hasAuthor > \"Gallenga PE\"."
```

RDF File

```
<http://ctro/data/Publication_1> <http://ctro/data/describes>
<http://ctro/data/ClinicalTrial_1> .
```

Figure 4. Extract of an annotated file in "annotated format" showing a *Publication* group (instance of Publication) and the extract of the corresponding RDF file showing its relationship (describes) with ClinicalTrial_1 (instance of ClinicalTrial).