Guidelines on the annotation of clinical trial abstracts.

1 Introduction

The present document describes the steps for annotating public clinical trial abstracts. We use the slot-filling-based annotation tool SANTO [1] configured to follow a conceptual schema derived from the Clinical Trial Ontology (C-TrO) [2]. The annotation task consists in identifying important clinical trial concepts/entities and their relationships. We focus on the cases of glaucoma and type 2 diabetes mellitus.

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 because the normal fluid in the eye is not properly drained. However, factors such as age, family history and other medical conditions such as diabetes, can increase the risk of glaucoma.

Type 2 Diabetes Mellitus (T2DM) is a chronic condition that affects the way the body metabolizes sugar (glucose). It is characterized by high blood sugar, insulin[1] resistance, and relative lack of insulin. People who suffer from this type of diabetes do not produce enough insulin to maintain the normal glucose levels. T2DM develops when the body becomes resistant to insulin or when the pancreas is unable to produce enough insulin. The exact causes of this disorder are unknown. Genetics and environmental factors, such as being overweight and inactive, seem to contribute to the development of this disorder.

Clinical trials

Clinical trials are studies in which two or more medical interventions are compared and evaluated. Important components of clinical trials are the PICO elements which answer to the following relevant questions in clinical research:

- **P**opulation/Problem: What are the characteristics of the population or patients? What is the Problem, condition or disease of interest?
- **I**: Which interventions are applied to the patients?
- **C**: What is the comparison or alternative to the intervention: placebo, a different drug, surgery, etc.?
- **O**: What are the possible outcomes of the study: reduce morbidity, death, complications, etc.?

An example of PICO elements for glaucoma is depicted in Figure 1.

[1] Insuline is the hormone that regulates the movement of sugar into the body cells.

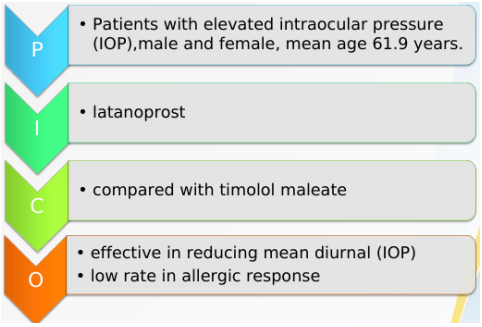

Figure 1: Example of the PICO elements for glaucoma

Outcomes

In our structure, the outcome of a medical intervention is composed by an *endpoint*, which is the name of an aspect or variable of interest and the *resulting measured value* that such variable reaches after the application of a given drug treatment. The measurements can be taken at different time points during the clinical trial. Some of the usual outcomes for glaucoma (an other types of glaucoma, like primary open-angle glaucoma (POAG) or chronic glaucoma, and ocular hypertension) and T2DM are described as follows.

Outcomes for glaucoma

Primary Outcomes Decrease of mean diurnal interocular (IOP) pressure. Eye pressure of 21 mmHg (millimeters of mercury) or higher, generally signifies ocular hypertension.

Secondary Outcomes Decrease of the excessive aqueous humor that provokes IOP.

Adverse effects Stinging, abnormal vision, bradycardia, cardiac failure, conjunctival hyperaemia, vision blurred, fatigue, hypotension, conjunctivitis, iris hyperpigmentation, etc.

Outcomes for T2DM

Primary Outcomes Decrease in HbA1c (*glycated haemoglobin*), also found as glycosylated haemoglobin, GHb, A(1c), haemoglobin A1c, or A1c. HbA1c decreases when *haemoglobin* – a protein within red blood cells that carries oxygen throughout the body– joins with glucose in the blood, becoming "glycated". Thus, measuring HbA1c can give an overall picture of the average blood sugar levels in the body. HbA1c can be expressed both as a percentage unit (%) or as a value in mmol/mol.

The HbA1c test that measures the average blood sugar level for the past two to three months considers: normal levels below 5.7%, prediabetes between 5.7% and 6.4%, and diabetes levels 6.5% or higher on two separate tests[2].

Adverse effects Weight gain or overweight, high-density lipoprotein (HDL), low-density lipoprotein (VLDL) and free fatty acid (FFA), kidney failure, high blood pressure, risk of heart attacks and cardiovascular diseases, other adverse effects like diarrhea, gastrointestinal disorder, nausea, vomiting, infection, headache, etc.

2 The annotation tool SANTO

To annotate clinical trial abstracts, we use a web-based annotation tool called SANTO [1] which has been configured according to the C-TrO structure.

After logging into SANTO, the annotators can select a file from a list of abstracts. Afterwards, the selected text is sentence-wise displayed.

The tool bar on the top-left of the web page depicted in Figure 2 has three annotation modes: *Curator / Annotator / Slotfilling*. The annotators have to choose only the Annotator and the Slotfilling modes. The *Annotator* mode allows to mark text with the corresponding annotation category. While *Slotfilling* allows to create templates and fill in their slots with annotated entities.

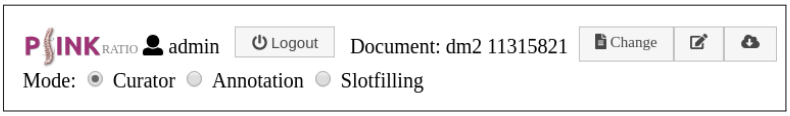

Figure 2: The mode/file tool bar

Annotation options

The options to annotate are found in the annotation toolbar shown in Figure 3:

- Add annotation: displays a catalog with the annotation categories (i.e. entity types).
- Delete annotation: deletes the selected annotation.
- Expand/Shrink annotation LEFT: the selected annotation is extended/shortened word by word on the left side.

[2]https://www.mayoclinic.org/diseases-conditions/type-2-diabetes/diagnosis-treatment/drc-20351199

- Expand/Shrink annotation RIGHT: the selected annotation is extended/shortened word by word on the right side.
- Done: indicates that the annotation of the whole document and the file is marked in the list of files.

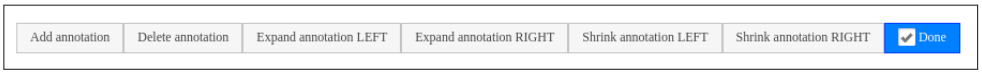

Figure 3: The annotation toolbar

3 Annotation

There are two types of annotations: single (or atomic) entities and composed entities (or complex templates). Here, we call the first ones single entities and the second ones just templates. The annotation of an entity consist of assigning an entity type (annotation category) to a span of text that represents such entity. For example, the text "1999" would correspond to the entity type `PublicationYear` and "Fox J" to the entity type `Author` . On the other hand, a template contains a set of slots (or fields) that can be filled with single entities or other templates names. In SANTO, the templates are predefined according to the underlying ontology schema. The templates type corresponds to a given class in the ontology. Templates can be populated by selecting the "Slotfilling" option in the menu bar. For example, a template called Publication_1 (an instance[3] of Publication) contains the slots for the entities author(s), publication year and title, and the slot for the *"describes"* relationship to a clinical trial template, so in this slot there can be the name of a clinical trial template: Publication_1 and ClinicalTrial_1 as *"Publication describes ClinicalTrial"*.

3.1 General annotation steps

1. Read carefully the whole text.
2. Identify the entities, as well as their entity types, templates and relations, in accordance to the C-TrO scheme [2] in Figure 7.
3. First, annotate the entities in the "Annotation" mode by marking a piece of text (e.g. "5.5") and then, assign the corresponding annotation (e.g. `ResultAbsoluteValue`). If the annotator needs to delete an annotation, he has to select the annotation to be deleted and then press key "Delete" in the annotation toolbar.

[3] An instance can be seen as a manifestation of a given class.

4. Once the entities are annotated, create templates by selecting the "Slotfilling" mode. In this mode the annotator can open a template and fill the slots with the corresponding annotated entities and relationships to other templates.

5. In case of being unsure, indicate the degree of certainty on the annotation by clicking on the circle next to the annotation (See Figure 4). If no degree of certainty is assigned, it means that the annotator is sure about the annotation. The degree of certainty does not have any effect on the annotation itself. In addition, the annotator should keep a record of the problems found during the annotation process in a text file, in order to discuss them with the annotation team.

6. Once the annotators have finished labeling all the entities and filling in the templates for an abstract, they have to tick the option "Done" on top of the Annotation toolbar and the Slotfilling panel in order to indicate in the list of files that both the entities and slots have been annotated in this abstract.

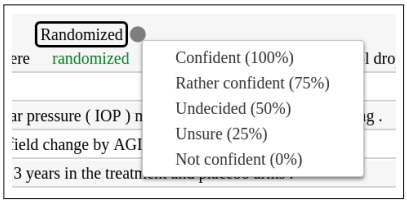

Figure 4: Selection of the degree of confidence of the annotation.

Annotation of entities

Single entities must be annotated with the most specific annotation. For example, the text "United States" has to be annotated as `USA` but not as `Country` since "USA" is more specific than "Country". In case that the most specific annotation of a concept does not appear in the catalog, the annotator should give notice to the administrators.

Annotation of templates

To create a template, select a template type and name it or leave the default name. Next, fill in the template with the adequate entities/template names. For example, in Figure 5, "Publication 1" is the name given to a "Publication" template that contains e.g. the entities "author(s)", "publication year" and "title", and the template "ClinicalTrial 1" in the field "describes" which relates the group "Publication 1" with the group "ClinicalTrial 1".

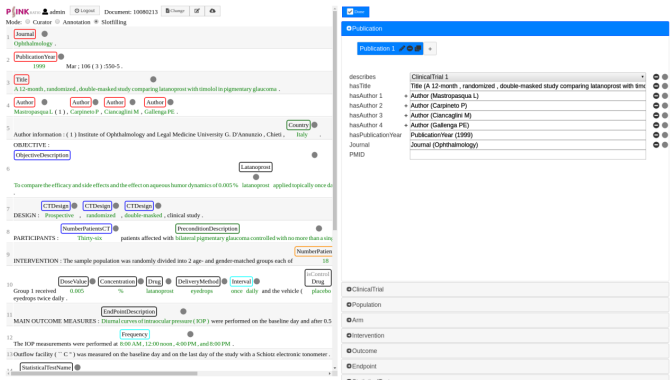

Figure 5: Template “Publication 1”

Annotation of multiple occurrences of an entity

All the occurrences of the same entity in a text have to be annotated, considering that an entity can be expressed in different ways as in lower and uppercases, acronymous, etc. For example, the drug Timolol could be found in different locations of a text as “timolol”, “Timolol” or “Timolol Maleate”.

When there are several occurrences of the same entity that are candidate to be included in a slot, only one of these occurrences has to be chosen. An example is depicted in Figure 6, where the drug “Insuline Glargine” is mentioned several times within the text but only one of these occurrences is included in the *hasDrug* slot in a Medication template.

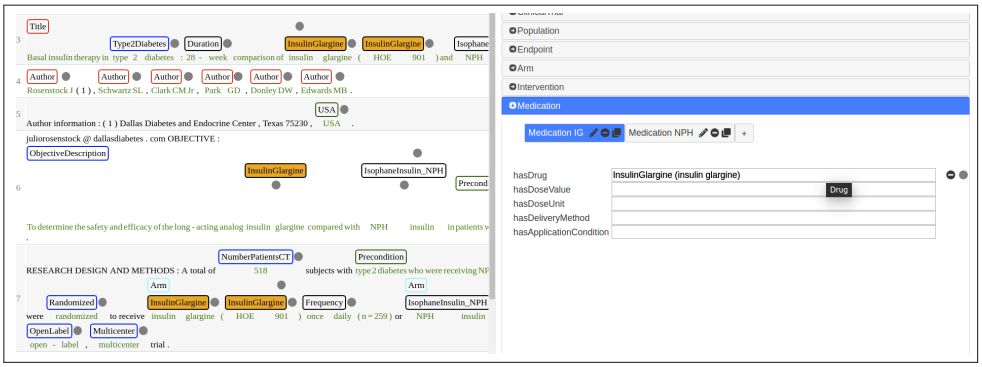

Figure 6: Annotation of different occurrences of “Insuline Glargine” and its inclusion in the *hasDrug* slot in Medication.

Notice that no more than one span of text has to be marked at the same type when the annotation category is assigned, otherwise all the marked spans will be concatenated into a single annotation. Only in some cases, the concatenation of several text spans is used. For example, when a title is too long.

4 Annotation of entities and template slots according to C-TrO

This section describes the slot types of the main templates in accordance to the annotation schema derived from C-TrO depicted in Figure 7.

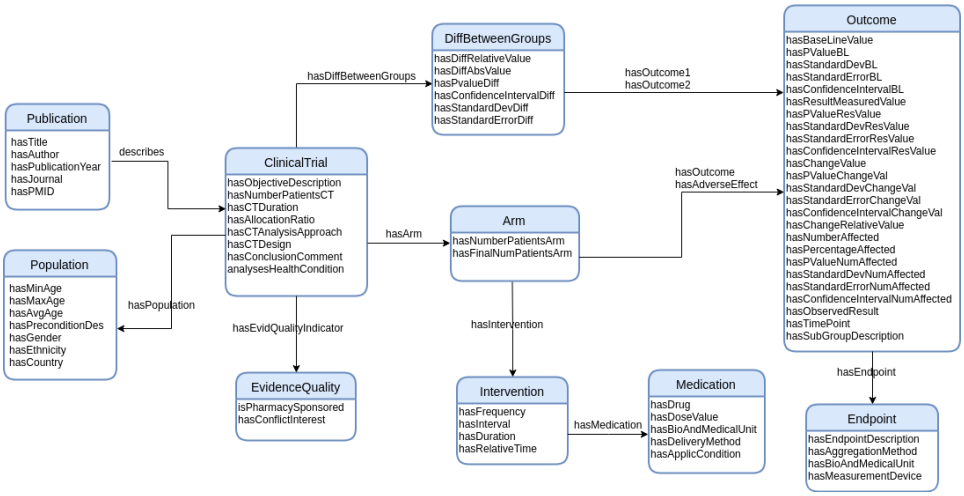

Figure 7: C-TrO Diagram

Arm

In the context of clinical trials, an arm is a group of participants that receives one or more interventions. These interventions can be of different type, but in this work we focus on medical ones (i.e., implying drug treatments). The selected participants comply an eligibility criteria or pre-conditions.

- NumberPatientsArm: initial number of patients in the arm at the begging of the study.
- FinalNumPatientsArm: final number of patients in the arm at the end of the study.
- hasIntervention: relates the arm to one or more interventions.
- hasOutcome: relates the arm with primary or secondary outcomes.
- hasAdverseEffect: links the arm to the outcomes of adverse effects.
- NumPatientsLeftArm: absolute number of patients who left the arm.
- RelNumPatientsLeftArm: relative number of patients who left the arm.

Outcome

An Outcome contains an endpoint description and the information relative to the results obtained for that endpoint in the clinical trial.

- *BaseLineValue*: initial measured value of the endpoint indicator at the beginning of the clinical trial. The corresponding p-value (PValueBL), standard deviation (StandardDevBL), standard error (StandardErrorBL), and confidence interval (ConfidenceIntervalBL) are included.

- *ResultMeasuredValue*: is the measured value at the last timepoint of the trial. The corresponding p-value (PValueResValue), standard deviation (StandardDevResValue), standard error (StandardErrorResValue), and confidence interval (ConfidenceIntervalResValue) are included.

- *ChangeValue*: indicates the amount of change (reduction/increment) in absolute values from the baseline of a given endpoint indicator (e.g. intraocular pressure, glucose level). This annotation contains the annotation subtypes "Reduction" and "Increment". The annotator has to select the appropriate annotation according to the direction of the change. In Figure 9 the corresponding reduction values of HbA1C for the insulin glargine and the NPH insuline treatments are annotated as `Reduction` . Notice that the sign "-" in "-0.46" and "-0.38" is not annotated since Reduction already implies a negative change direction.

- *ChangeRelativeValue*: indicates the amount of change (reduction/increment) in relative values (it is commonly a percentage) from baseline of an endpoint indicator (e.g. intraocular pressure, glucose level). This annotation contains the subtypes "Reduction" and "Increment". The annotator has to select the appropriate annotation according to the direction of the change.

- *NumberAffected*: refers to the number of people who reached an undesired (i.e, adverse effect) or a desired result for a given endpoint indicator (i.e. primary or secondary outcomes). For example, "**Ten** people reduced IOP in a significant amount" and "**Five** people suffer from headaches".

  The corresponding p-value (PValueNumAffected), standard deviation (StandardDevNumAffected), standard error (StandardErrorNumAffected), and confidence interval (ConfidenceIntervalNumAffected) are included.

- *PercentageAffected*: percentage of participants affected by adverse effects (e.g. "**10** percent of participants suffer from headaches.") or by a

desired value of the endpoint indicator. Figure 8 shows an example of number and percentage of patients affected by increased pigmentation.

- *ObservedResult*: is a textual description of the results (e.g. "Timolol was observed to be superior .").
- *TimePoint*: is the time point when the change from the baseline was measured. All the time points mentioned in the text have to be annotated. However, only those time points that specifically refer to the results of an endpoint indicator have to be included in the corresponding Outcome template.
- *SubGroupDescription*: describes either a population subgroup in the arm with certain characteristics, or special conditions from which the results are reported (e.g. "overweight patients").
- *hasEndpoint*: relates an Outcome template to an Endpoint template.

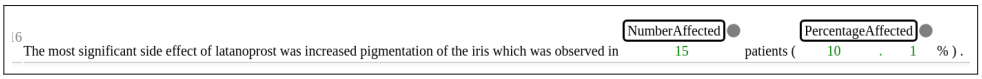

Figure 8: Annotation of "NumberAffected" and "PercentageAffected" (PMID 8628543).

Endpoint

The Endpoint template is formed by the slots:

- *EndpointDescription*: is the name of the endpoint indicator, which is the variable that is measured and compared in the study.
- *AggregationMethod*: is the method used to aggregate the measurements of an endpoint indicator (e.g. mean).
- *BioAndMedicalUnit*: unit in which the endpoint indicator is measured.
- *MeasurementDevice*: device used for measuring the endpoint indicator's values.

In the excerpt shown in Figure 9, we can see the annotations of endpoints, units, reduction values, etc.

Sometimes the endpoint unit is not reported since it is implied by the kind of endpoint in question, like in:

> "Reduction of IOP was 6.0 +/- 4.5 and 5.9 +/- 4.6 with latanoprost and 4.8 +/- 3.0 and 4.6 +/- 3.1 with timolol after 6 and 12 months, respectively."

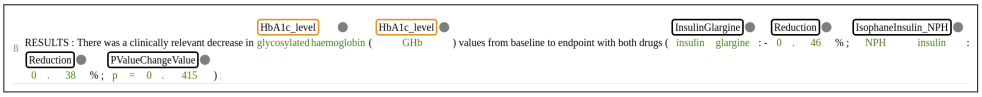

Figure 9: Annotation of absolute "Reduction" values from baseline for HbA1c and other entities (PMID 12734781).

Intervention

- *Frequency*: frequency in which the intervention is applied to the patients (e.g. "twice per day".)
- I*nterval*: interval of time in which the intervention is repeated (e.g. "every 12 hours".)
- *Duration*: duration of the intervention. Beware of its difference with the trial duration.
- *RelativeTime*: relative period of time such as *morning, afternoon and evening*. This is used to indicate when the drug was administrated (e.g. "Timolol was given once per day in the mornings".)
- *hasMedication*: relates the intervention to a Medication template.

Medication

- *Drug*: is the drug/substance given to the patients in an arm's intervention.
- *DoseValue*: value of the dose of the applied drug.
- *BioAndMedUnit*: unit of the dose.
- *DeliveryMethod*: way or means in which the drug was administrated to the patients.
- *ApplicCondition*: condition in which the drug had to be administrated (e.g. "without breakfast")

Clinical Trial

- *ObjectiveDescription*: is the aim of the clinical study. It is normally found in the section "Objective'/Goals" of the abstract.
- *NumberPatientsCT*: number of patients at the beginning of the clinical trial.
- *CTDuration*: duration of the clinical trial. Beware of its difference with the duration of the intervention.

- *AllocationRatio*: proportion in which the randomization was performed. Typically, this ratio is 1:1 (i.e., one to one person was assigned to each arm when there are two arms in the clinical trial).
- *CTAnalysisApproach*: is the approach used to determine results of the clinical trial. It can be either a Pre-Protocol or an Intention-To-Treat (ITT) approach. In the first approach, only those patients who complied with the initial protocol of the study are considered in the calculation of the results, while in the second approach, the whole population is considered.
- *CTDesign*: is the design of the clinical study (e.g "parallel", "randomized", "multi-center", etc.)
- *ConclusionComment*: is the conclusion of the clinical study. It is normally found in the section "Conclusions" of the abstract.
- *HealthCondition*: health condition, disorder or disease that is studied in the clinical trial (e.g. "T2DM".)
- *FinalNumPatientsCT*: final number of patients in the clinical trial. This information is not provided in every abstract.
- *NumPatientsLeftCT*: number of patients who left the clinical trial. This information is not provided in every abstract.
- *RelNumPatientsLeftCT*: percentage of patients who left the clinical trial. This information is not provided in every abstract.
- *hasPopulation*: joins this template to a Population template.
- *hasDiffBetweenGroups*: joins this template to a DiffBetweenGroups template.
- *hasEvidQualityIndicator*: joins this template to an EvidQualityIndicator template.

DiffeBetweenGroups

This template contains the information about the differences between the outcomes of two interventions.

- *DiffRelativeValue*: difference between the two results given as a percentage.
- *DiffAbsValue*: difference between the two results given as absolute value.

- PvalueDiff, ConfidenceIntervalDiff, StandardDevDiff, and StandardErrorDiff: are the corresponding p-value, confidence interval, standard deviation, and standard error.
- *hasOutcome1 and hasOutcome2*: indicate the respective Outcome groups of the two interventions whose difference is compared.

Figure 10 shows the annotation of a difference between groups value and its respective statistical measures (i.e., standard deviation and confidence interval) for the endpoint Diurnal IOP. The corresponding DifferenceBetweenGroups group is on the right.

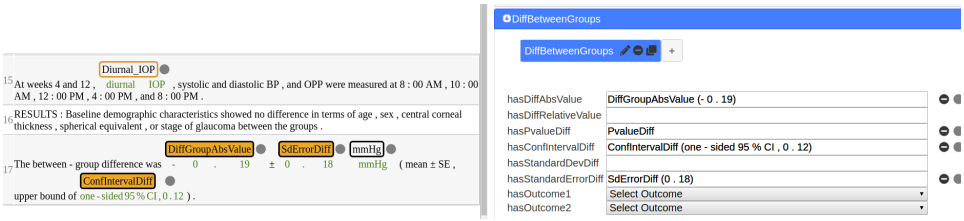

Figure 10: Group "DifferenceBetweenGroups" (PMID 26756747).

5 Remarks on the annotations

Different authors may express their methods and results in different ways. Here we give some guides of what to do in several cases observed. However, if there is not a case that is contemplated here, then contact the administrators.

- When the authors report both the resultValue and change value (i.e. reduction/increment). Both values have to be included in the Outcome template. If the respective PvalDiff for ResultValue and Reduction/Increment are reported, then in the DiffBetweenGroups template only the PvalDiff of Reduction/Increment has to be included.
- When only the differenceBetweenGroups values are provided, but the result or change values are not, then outcomes have to be created for each intervention group (i.e., arm) whose difference is being compared and these outcomes have to be included in the DiffBetweenGroups template. In this case, the outcome groups will remain empty.
- When the treatments that the participants undertook before randomization are described, such treatments do not have to be included as part of the drug interventions that are being compared in the trial, i.e, no Medication or /and Intervention template have to be formed. These pre-treatments have to be annotated as pre-conditions.

- Run-in period treatment: is a treatment given before randomization. So, for the annotation, its description and results are taken as a pre-condition, as it is currently annotated.
- In textual annotations, the period symbols have to be included.
- Well defined sections like title, objective, and conclusion have to be included in one or more annotations depending on their length.
- Those annotations that contain numeric values (e.g. duration, number of participants, etc.) are preferred to those that contain letters, to be included in the templates. For example, if there are two annotations for the duration of the clinical trial: "six-weeks" and "6 weeks". Then, the second one is preferred to the first one.
- Premixed insulins (in T2DM) and fixed combinations of drugs (e.g. "dorzolamide/timolol fixed combination" i.e. "TDFC") are considered single drugs. So, there is no need to include a Medication group for each of the components of these combined drugs.

References

[1] 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.

[2] Sanchez-Graillet, O., Cimiano, P., Witte, C., Ell, B.: C-TrO: An Ontology for Summarization and Aggregation of the Level of Evidence in Clinical Trials. *Proc. of the 5th Joint Ontology Workshops (JOWO): Ontologies and Data in the Life Sciences*, (2019)