Annotation Guidelines for outcome entity extraction in clinical trials articles

The goal of this task is to extract the different types of outcomes (Primary, Secondary, Other) in a clinical trial **article**, as well as their associated time frame and their (optional) description.

Each Highlight color corresponds to an entity. For Each entity example, only the useful entities for the current exemple are highlighted. Entities should avoid containing punctuation (commas and dot, parenthesis is ok)

Outcome Entities

PrimaryOutcome : A text span should be annotated as *Primary Outcome* if there is an explicit term referring to it saying it is a primary outcome / endpoint (or sometimes primary efficacy endpoint/ primary safety endpoint). Or if the section title explicitly states that the outcomes are primary.

Ex 1: *The primary outcome is time to clinical improvement at 28 days after enrollment.*

Ex 2: *Primary Outcomes*

Time to clinical improvement was evaluated at 28 days after enrollment.

You must take in account the way the sentence is formulated (plural or singular) if they declare a singular outcome as the combination of multiple measures, then it should be annotated as the same entity.

Ex : The primary outcome was a composite of mechanical ventilation, hospitalization >14 days or in-hospital death

SecondaryOutcome : Same as Primary outcome but there are usually more secondary outcomes than primary, so sometimes you have to be careful with the syntax of the outcome list, there can be commas and semicolons.

Ex: *secondary outcomes were the probabilities of alleviation of clinical symptoms; improvement of C reactive protein, erythrocyte sedimentation rate, tumor necrosis factor α, interleukin 6, and absolute blood lymphocyte count; improvement of lung lesions on chest radiology; all cause death; and disease progression in patients with mild to moderate disease*

→ In this example, the semicolons are the separators for the outcomes, and so the 2nd outcome here is the whole enumeration “improvement of C reactive protein, erythrocyte sedimentation rate, tumour necrosis factor α, interleukin 6, and absolute blood lymphocyte count”

If the outcome is measured on a subgroup of the patients, do not include the patient part in the outcome. And if there is a ‘or’ statement usually it is considered as the same entity so you should annotate the whole part including the ‘or’.

Ex: *Secondary outcomes were time to discharge from hospital and, among patients not on invasive mechanical ventilation at randomisation, post-enrolment use of invasive mechanical ventilation (including extracorporeal membrane oxygenation) or death.*

OtherOutcome : This annotation should be used for the declared outcomes without the mention of primary or secondary but it is still stated that it is an other/additional/… outcome/endpoint/analysis or if it comes right after the declaration of primary and secondary outcomes, following the same rules as the previous entities.

Ex : *The primary outcome was [...]. Secondary outcomes were [...]. In addition, we also analyzed the peripheral blood T lymphocyte subsets (count and proportion) before and after treatment.*

UnclearOutcome : This annotation should be used for all the other measures that look like an outcome but it is unclear in the text.

Ex: Lab investigations including total blood cell count, total leukocyte count, C-reactive protein, D-dimers, lactate dehydrogenase, oxygen saturation was performed on daily basis to monitor the status of patient

Outcome Related Entities

TimeFrame : Every time frame or time point linked to an outcome should be annotated as a TimeFrame and linked to it using a **MeasuredAt** relation link. The MeasuredAt link always goes from the outcome to the time frame. The TimeFrame text span should contain the more detailed TimeFrame that we can annotate.

Ex: The primary outcome was the clinical recovery rate at 7 days from the beginning of treatment.

One TimeFrame can be associated with multiple outcomes in some cases. But an outcome can only have one (or zero), in the case where there are multiple time frames, the closest one should be associated with the outcome. If there are multiple successive time points, they should be annotated as the same time frame (e.g. " … at 7, 21 and 28 days")

Ex: *Outcomes were assessed at 28 days after randomisation, with further analyses specified at 6 months. The primary outcome was 28-day all-cause mortality. Secondary outcomes were time to discharge from hospital and, among patients not on invasive mechanical ventilation at randomisation, post-enrolment use of invasive mechanical ventilation (including extracorporeal membrane oxygenation) or death.* → In this example, for the primary outcome "all-cause mortality" the link should be made with the 28-day Time Frame that is just before it. For the 2 secondary outcomes, the link should be made with the "at 28 days after randomisation, with further analyses specified at 6 months" , as they have no time frame specified near them and that the first timeframe refers to all outcomes as stated: "Outcomes were assessed …"

Special case if the Time Frame comes in the middle of the outcome declaration, then it should be part of the outcome entity and not be annotated as time frame

Ex : *The secondary outcomes were [...], all-cause 28-day mortality, [...]*

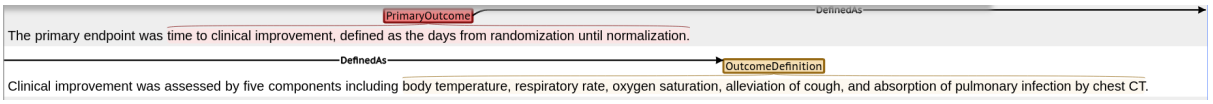
OutcomeDefinition : There are some cases when the outcome will be detailed later in the text. In this case the definition should be annotated as *Outcome Definition* and then we should use a relation link

DefinedAs between the short description of the outcome (annotated as PrimaryOutcome, SecondaryOutcome, or OtherOutcome) and the OutcomeDefinition.

An Outcome can have multiple definitions associated to it, but a definition should have only one outcome.

Ex 1: *The primary outcome was the clinical recovery rate at 7 days from the beginning of treatment. Clinical recovery was defined as continuous (>72 hours) recovery of body temperature, respiratory rate, oxygen saturation and cough relief after treatment, with following quantitative criteria: axillary temperature ≤36.6°C; respiratory frequency ≤24 times/min; Oxygen saturation ≥98% without oxygen inhalation; mild or no cough.* → *"DefinedAs" relation link between the PrimaryOutcome and the OutcomeDefinition*

Ex 2:



If the definition is cited next to the outcome then it should be integrated in the outcome (for example if the definition is in a parenthesis) unless it is stated "defined as", every time "defined as" appear after an outcome, it should be defined as an OutcomeDefinition

Ex : *Secondary outcomes were the improvement of clinical symptoms (cough, diarrhea, dyspnea, fever, myalgia)* → here we annotate the whole combination of outcome and its description in the parenthesis as SecondaryOutcome

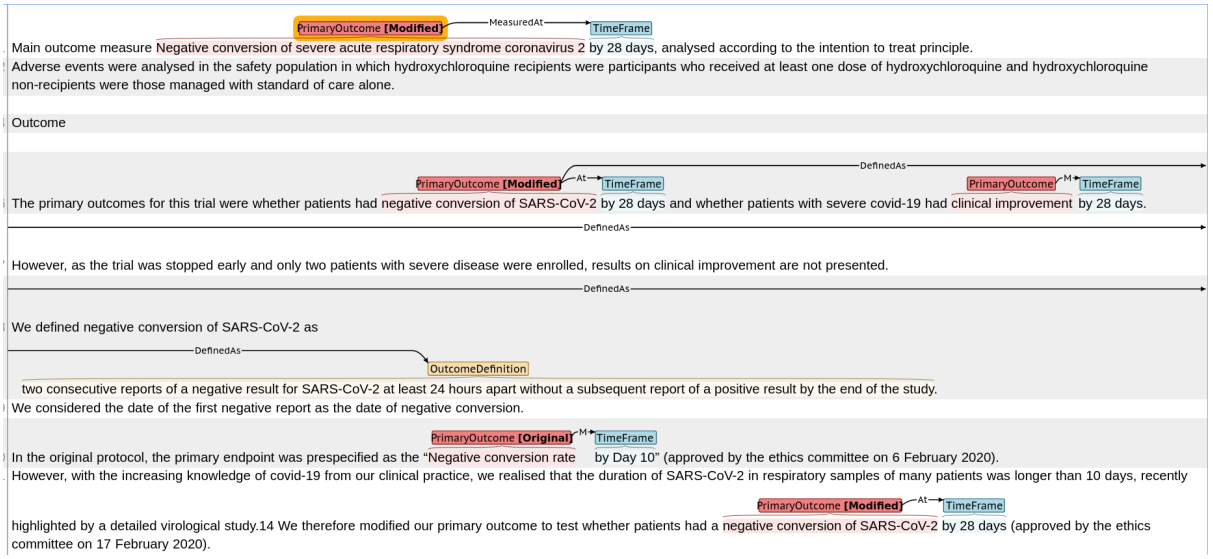
But if a TimeFrame or any other text part comes in the middle between the outcome and its description then we annotate the 3 different entities:

Ex : The secondary outcomes were the time until successful liberation from mechanical ventilation, the ventilator-free days (during the 28 days after inclusion in the study; numbers of days without mechanical ventilation), […]

Special Case of Declared modification of outcome : Modification attribute

This is the case when a modification of the **primary** outcome is stated : in this case an attribute should be added to the primary outcome to identify which is the original outcome and which is the modified outcome (even if it is only the time frame that changes, it should still be the outcome who has the attribute, as the timeframe is linked to the outcome anyway), if the modified comes multiple times in the article, then all the entities should have the "Modified" Attribute

Ex :


PrimaryOutcome [Modified]
MeasuredAt
TimeFrame
Main outcome measure Negative conversion of severe acute respiratory syndrome coronavirus 2 by 28 days, analysed according to the intention to treat principle.
Adverse events were analysed in the safety population in which hydroxychloroquine recipients were participants who received at least one dose of hydroxychloroquine and hydroxychloroquine non-recipients were those managed with standard of care alone.
Outcome
DefinedAs
PrimaryOutcome [Modified]
At
TimeFrame
PrimaryOutcome
TimeFrame
The primary outcomes for this trial were whether patients had negative conversion of SARS-CoV-2 by 28 days and whether patients with severe covid-19 had clinical improvement by 28 days.
DefinedAs
However, as the trial was stopped early and only two patients with severe disease were enrolled, results on clinical improvement are not presented.
DefinedAs
We defined negative conversion of SARS-CoV-2 as
DefinedAs
OutcomeDefinition
two consecutive reports of a negative result for SARS-CoV-2 at least 24 hours apart without a subsequent report of a positive result by the end of the study.
We considered the date of the first negative report as the date of negative conversion.
PrimaryOutcome [Original]
At
TimeFrame
In the original protocol, the primary endpoint was prespecified as the “Negative conversion rate by Day 10” (approved by the ethics committee on 6 February 2020).
However, with the increasing knowledge of covid-19 from our clinical practice, we realised that the duration of SARS-CoV-2 in respiratory samples of many patients was longer than 10 days, recently
PrimaryOutcome [Modified]
At
TimeFrame
highlighted by a detailed virological study.14 We therefore modified our primary outcome to test whether patients had a negative conversion of SARS-CoV-2 by 28 days (approved by the ethics committee on 17 February 2020).