Polly Verified: v1.0

Curated Clinical Trial Data

Polly by Elucidata

Elucidata provides FAIR data from clinical trial publications and ensures that this data is machine-actionable and analysis-ready. This curated data enables seamless data exploration and analysis. This document covers the quality assurance and quality checks performed on each curated study besides those shared in the indication-specific specifications.

CLINICAL TRIAL STUDY-LEVEL INTEGRITY

- Cross verify information and total numbers (where applicable) for the following field categories: Source, Trial, Bias, Arm, Patient, Background and Treatment

DATA COMPLETENESS

- Ensure all reported background results are extracted from clinicaltrials.gov and article figures, tables, legends, text and supplementary information
- Ensure all reported patients are extracted from clinicaltrials.gov and article figures, tables, legends, text and supplementary information
- Ensure all reported endpoints and adverse events are extracted from clinicaltrials.gov and article figures, tables, legends, text and supplementary information
- Ensure all reported comparisons are extracted from clinicaltrials.gov and article figures, tables, legends, text and supplementary information
- Cross verify fields which are empty/NA

DATA HARMONIZATION AND VALIDITY

Data Harmonization Checks

- Ensure variable type for each field (integer, float, text, boolean
- Fields follow defined controlled vocabulary
- Fields follow defined format
- Ensure there are no duplicate rows

Data Validation Checks

- Fields values are within defined scientific range
- Fields that expect percentages, values in these columns should not be more than 101
- Columns randomized.drug, randomized.drug.class, drug1, and drug1.class are in concordance with each other
- Columns control.drug, control.drug.class, and control.treatment are in concordance with each other

114 Sansome Street
San Francisco, CA 94104

SCIENTIFIC ACCURACY

- If "endpoint.type" = Binary, then "n.event" should not be empty. Could be marked as NA, if only count of participants are not explicitly mentioned.
- trial.duration >= treatment.duration
- n.arm >= n.randomized >= n.treated
- n.arm >= n.randomized >= n.completed
- n.endpoint >= n.risk
- n.treated >= n.endpoint >=n.observed
- n.endpoint >= n.event
- if time is positive, then time >= time.last.dose
- Column time, exclude the timepoint of "0" for endpoints which have "baseline.X" column
- If trial.blinding is Open Label then adequate.blinding should be No
- If time == treatment.duration then primary.timepoint should be set to "Yes" else "No"
- If race.asian >=50, then asian.majority should be set to 'Yes', if its between 0 to <50 then No
- value in "baseline.hba1c" should be same for 'baseline' endpoint when endpoint == HbA1c
- value in "baseline.fpg" should be same for 'baseline' endpoint when endpoint == Fasting Blood Glucose or Fasting Plasma Glucose
- If value, change, percentchange or md has numeric value, then value.calc, change.calc, percentchange.calc and md.calc should not be empty
- If value.sd, change.sd, percentchange.sd or md.sd has numeric value, then value.sd.calc, change.sd.calc, percentchange.sd.calc, md.sd.calc should not be empty respectively
- If value.se, change.se, percentchange.se or md.se has numeric value, then value.se.calc, change.se.calc, percentchange.se.calc, md.se.calc should not be empty respectively
- X.ci.high > X.ci.low and X.iqr.high > X.iqr.low (where X = value, change, percentchange, md, md.change, md.percentchange, or, rr, hr, rr)
- If X.ci.low is numeric than X.ci.high should be numeric too (or vice-versa) and value.ci.calc should not be empty
- Ensure ci.range is not NA when X.ci.low and X.ci.high have a value
- If md, md.change, md.percentchange is numeric than value.control, change.control, percentchange.control should not be empty
- change.control, percentchange.control, md.change should be NA for Placeco or control arm
- if either of or, rd, rr, hr is numeric than ne.control should not be empty
- if "endpoint" field value has Yes/No in the name, then ne.control should not be empty
- if "endpoint" field value has Yes/No, or adverse event or endpoint with two possible outcome, then binary.endpoint should be Yes and endpoint.type should be Binary
- if "endpoint" field value does not have Yes/No, or adverse event or endpoint with more than two possible outcome then binary.endpoint should be No and endpoint.type should be Continuous
- If value of drug1.titration is forced or targeted, then drug1.titration.duration should not be empty nor NA
- If Selection : binary.endpoint is Yes, then endpoint.type = "Binary" ; AND If Selection : binary.endpoint is No, then endpoint.type = "Continuous" ;