**CBER SEND Pilot 3 – Standardized data validation review**

**I. Comments for CBER Nonclinical Reviewers:**

1. The data definition file (Define.xml) has various issues with regards to compliance with CDISC Define specification – see DataFit reports for details (not review related).

These are not review related.

SEND people has general question about how define is used, specification of the define seems fit for clinical, but might not make sense of non-clinical community. Improve the use of define can contribute to the standardization of define in non-clinical community.

1. The nSDRG section 4.3 says: “No supplemental qualifiers are used”. But there are MA and MI suppquals in the submission. By looking at the data contents in SuppMA and SuppMI, it seems that neither of the datasets is needed considering the captured suppqual information can be found in the primary datasets (MA and MI) – see the original result value variables.

What is “Result Modifiers” in suppMA? Are they duplicate information in MAORRES? We noticed it is listed in suppMA when “MASTRESC” is not “UNREMARKABLE”.

1. The dosage amount value is expressed in various forms (ex: 1.024e13, 1.024x1013, 10^13, 1E13). Ideally there should be a single correct expression to be used throughout the submission.

See below different display in different places:

1.024e13: Study Report;

1.024x1013: nSDRG; Study Report;

10^13: nSDRG; Define.xml; dataset (EX)

1E13: nSDRG; Define.xml; dataset (DM)

Check with IG to see if there is any recommendation in data or define

1. It appears that the dosage unit value uses two different expressions: VG (“vector genomes") and GC (“Genome Copies”); Ideally, a single term is used throughout the submissions and/or an explicit clarification about their equivalence is documented in the relevant files (i.e., study report, define.xml, nSDRG).

In SEND data, the unit is vg/kg, in study report it is listed (GC/kg), are they the same? Shall they be listed consistent?

Consistent units between report and data might be helpful

1. Should the below custom (i.e., non-CDISC terminologies) terminologies be added to the CDISC standard terminologies? Note: recommend adding to CDISC if a value is commonly seen in CBER studies.
2. In the nSDRG, the following “Tissue Comments” in MA explanation does not seem clear and understandable.

* Timing Variables:
  + - *“This dataset populates the MADTC with the of data collection dates, not the date/time”,* don’t understand what that means.

Time of observation might not be captured. nSDRG language should be more clear.

* + - *“Tissue comments in the microscopic pathology report which noted whether tissues were collected and did not addition interpretation, they were not included in the SEND dataset in the CO-Comments Domain.”* don’t understand what that means.

Maybe bad language.

* + - *“Measurements (dimensions) of lesions appear in –ORRES are not carried over to –STRESC or CO-Comments Domain.” Why?*

All information from ORRES should be parsed at standard locations and standard locations would be variables within the domain or supplement qualifiers, not in comments