Draft Response to the US EPA Questions on Type Approval Data

USCG Received two questions stemming from the Supplemental Notice of Public Rule Making. The questions are listed below, following some background provided for context.

Background

For background, analyses are designed to demonstrate that concentrations are lower than the discharge standards. The analyses *are not* designed to resolve concentrations in below the discharge standard, nor is test data suited to compare BWMS performance evaluations conducted at different times and locations.

Test facilities have developed their sampling and analysis methods to meet the guidance in the US test protocols for land-based testing (US EPA 2010, hereafter, the "ETV")[1]. The guidance centers around the discharge limits (e.g., <10 per m^3). Both the sampling and analytical methods for organisms ≥50 µm and ≥10 and <50 µm, in general, are calculated per Eq. 1:

Eq. 1 $$P = \frac{ICD}{AS}$$

Where *P* is the population concentration (individuals per volume), *I* is the tally of living individuals in the target population, C is the concentrated sample volume, *D* is dilution (e.g., due to adding reagents), *A* is the aliquot volume, and *S* is the total sample volume (See First et al. 2022 for additional descriptions)[2]. Considering the minimum unit of detection is 1 living individual, the method MDL can be calculated as:

Eq. 1 $$MDL = \frac{1C}{AS}$$

Note, *D* is removed here, as the value is typically 1.

For your consideration, here is our response to the document prepared from ERG.

Please see the answers below. In general, we concur with the comment that the optimal approach to summarize data with left-censored data is to use a statistical approach, such as the Maximum Likelihood Estimation, as demonstrated in the article cited in the comment. However, we would also mention that the test facilities have designed their sampling and analysis protocols to demonstrate that concentrations are statistically <10 m^3 or mL.

- Regarding the low Method Detection Limits (MDLs) reported for organisms ≥10 and <50 µm:
  Can USGS investigate the validity of the 0.01 MDLs?
  Response: Typical MDLs for organisms ≥10 and <50 µm are ~0.3 per mL. This is reflective of the guidance in the ETV, where 3 L is concentrated to 1 L (a 3x concentration factor) and 1 mL is analyzed. It is also the MDL when 3 mL of whole water are analyzed (e.g., 1 ind. per 3 mL ≈ 0.3

[1] See https://www.dco.uscg.mil/Portals/9/ETV%20EPA%20Report.pdf?ver=2018-05-22-081043-607
[2] https://doi.org/10.3389/fmars.2022.1034386

ind. per mL). Some test facilities claim to concentrate 60 L into 1 L, then analyze 2 mL of the concentrate. This yields an MDL of 0.0083 mL^{-1} (rounded to 0.01 mL).

- Regarding the high MDLs reported for organisms ≥50 µm:
  Can USGS explain why these MDLs are high (especially 5 org/m3) or if they should decrease? Also, does the ETV protocol allow for only 0.2 m3 of water to be analyzed per trial, which would result in an MDL of 5 org/m3?
  In this case, the test facility concentrated 10 m^3 to 1 L and analyzed 20 mL. This is a higher MDL than most test facilities, but it is still valid under the ETV protocol.

- Regarding the Quality Control of the TA data entry:
  The summary of quality assurance and quality control procedures is be divided into three sections: Database structure, Data Entry, Data Analysis Plan, Query development. Details are available below. In summary, all data were manually entered by analysts familiar with the data and reports. The manual entries were 100% checked by a second analyst (our goal was >90%, but we checked 100%). Errors were tracked in a spreadsheet and resolved by the first and second analyst.

_____

More details about the TA data entry and QC procedures

I. Database structure:
   - An instructional PowerPoint was created, highlighting TA Database key components and formatting, which was then provided and discussed with all members involved with data entry and QAQC
   - An overview of data entry was described in the Power Point (with reference to the data entry template that was created)
     - Parameters divided among 3 NRL personnel
       - Parameters in the database were considered the 'Core Parameters' from the ETV Protocol
     - Open LB or SB Test Report
       - Typically located in searchable .pdf files
       - The test reports were generated by an Independent Laboratory's (IL) sub-laboratory
       - Test Reports were received from USCG for a specific TA and the majority of the data was for Version 0 of the TA.
     - Generate template of test cycles
       - This included a list of all biological efficacy tests including test facility (TF) nomenclature that we then standardized so each test cycle had a unique and relatable identifier

        - Determined if tests were considered valid/invalid and pass/fail by the IL; this usually was determined from the IL Summary Report
        - Other information such as dates of tests, hold times, locations (SB tests), etc. were included in the test cycles
    - Locate data for parameter of interest.
        - Replicate data entered where possible
        - Data typically in tables within LB Report
        - Manually transcribe data from report into its parameter specific Excel Worksheet (WS)
        - Copy data exactly as it was entered by Test Facility (i.e., special characters, terminology, units, etc.)
    - Relate each bit of data back to its test cycle and TA-ID which could then be used to relate it to the Type ApprovalStore data on network as it is entered
    - Post process: Terminology Standardization
- A Parallel operation was completed for administrative data entry related to the Type Approval and entered into the workbook titled 'Front Tables'
    - TA Certificate information was copied over including TA Certificate number, version, BWMS Manufacturer, model number, treatment type, limitations, dates, etc.
- An overview of Quality Assurance Checks:
    - Internal QA process
        - Manual data entries are rechecked row by row via QA lead (pre-determined) or person(s) selected by QA lead
        - Review included comparing the entered value within a worksheet back to the table from the LB or SB report in which it was generated
        - Mistakes are reconciled and recorded
        - >90% check of all data is goal
- Data mapping check
    - Check that data maps to appropriate Standardized Source Name
    - Check that data maps to appropriate Test Cycle
    - Check that the list of Test Cycles and their Valid/Invalid criteria was extracted correctly from the reports

- Instructions on which files to open (complete with examples) as well as the entry process, naming convention and archiving process to perform (for updated, in-process and completed worksheets)

II. Data entry

- A parameter data entry template was created to act as a 'working copy' for all data entry personnel to use. The template was discussed and agreed upon by the data entry team.
- An instructional PowerPoint was created to instruct personnel on the correct steps for entering parameter data into the TA Database. This step-by-step guide walked the user through the data entry template and detailed what each workbook tab represented; what each column represented and common pitfalls to avoid.

- As mentioned in the 'Database structure' section, all data entry worksheets are subject to an internal QA process that involves a row-by-row check and verification of all entered data. The personnel performing the QA checks have also been shown the 'working copy' template as well as the data entry user manual PowerPoint. The goal is to have the data entry and the QA performed by two, separate personnel. The data entry for parameters underwent a 100% QA check (i.e., all data was reviewed and verified).

III. Data Analysis Plan

- Once entered, Data analyses were performed using the built-in functions of Microsoft Excel.

IV. Query Development

The list below is an example of a data validation exercise we completed prior to releasing information to USCG/EPA for the EPA requested data queries. This exercise also provided quality control (QC) to determine the utility of the database in generating tables and figures for comparative and informative purposes.

1. Link test data together
    a. Data Queries were created using the MS Excel 'Power Query' feature which is a graphical interface for obtaining and linking data from the various structured tables within the parameter worksheets.
    b. This feature allowed data to be connected among the wide range of data sources that could be easily refreshed without altering the original data
    c. This also allowed for the ability to filter, transform, merge, group, append, etc., the data as needed.

The following set of steps provides the general process for 'filtering' the data and performing a set of QA checks for the EPA worksheets as they were generated

2. Exclude invalid data (in the "All_SB_Test_Cycles" tab, where "Valid/Invalid" field=Invalid)
3. Completeness check of SB data
    a. Check that all valid SB TA_Index have corresponding valid LB TA_Index and vice versa.
    b. Check that TA_Index for SB and LB share the same Treatment Type
4. After excluding invalid data, perform QC checks:
    a. Do any TrialID lack StandardizedSourceName=Treatment Discharge or Treatment Uptake?
    b. Do any TrialID/StandardizedSourceName combinations have missing 50, 10-50 organism data or TSS data?
    c. Do any TrialID/StandardizedSourceName combinations have replicate 50, 10-50-organism data or TSS data?
    d. Are there any extreme results (defined below) for StandardizedSourceName=Treatment Discharge?
        i. 50 organisms greater than 50
        ii. 10-50 organisms greater than 50
    e. Do any TrialID data fail the following logic checks for the two largest organism classes (50, 10-50)?
        i. Treatment Uptake < Treatment Discharge

ii. Treatment Uptake < 100