



December 17, 2008

United States
Department of
Agriculture

VETERINARY SERVICES MEMORANDUM NO. 800.96

Animal and Plant
Health Inspection
Service

TO: VS Management Team
Directors, Center for Veterinary Biologics
Biologics Licensees, Permittees, and Applicants

Veterinary Services

Washington, DC
20250

FROM: John R. Clifford /s/ Brian J. McCluskey, for
Deputy Administrator
Veterinary Services

SUBJECT: Guidelines for Submitting Electronic Data Files for Statistical Analysis

I. PURPOSE

This memorandum provides guidelines for the submission of supporting data in the form of electronic files to the Center for Veterinary Biologics – Policy, Evaluation, and Licensing (CVB-PEL) and supplements the Guidelines for Submission of Materials in Support of Licensure specified in Veterinary Services Memorandum No. 800.50.

II. CANCELLATION

This Memorandum cancels Veterinary Services Memorandum No. 800.96 dated September 12, 2000.

III. BACKGROUND

In compliance with 9 CFR 102.3(b)(2)(ii), applications for United States Veterinary Biological Product Licenses must be supported by data as outlined in Veterinary Services Memorandum No. 800.50, section IV.C., Supporting Data for a Product License Application. License applicants commonly use electronic files to submit such data. To facilitate the timely and efficient review of data, APHIS recommends that data submitted in electronic form adhere to the guidelines described below.

IV. GUIDELINES

Guidelines for submitting electronic data files are appended to this memorandum.



Safeguarding American Agriculture

APHIS is an agency of USDA's Marketing and Regulatory Programs
An Equal Opportunity Provider and Employer

Federal Relay Service
(Voice/TTY/ASCII/Spanish)
1-800-877-8339

APPENDIX

GUIDELINES FOR SUBMITTING ELECTRONIC DATA FILES

- 1. Recording Data.** Submit the complete data in its raw unaltered form. Include all recorded data for each reported outcome. Indicate in the report if the analysis is based on a subset of the recorded data.
 - 1.1. Types of Data** – For the purpose of this document, data can be described as either quantitative (continuous measurements or integer counts) or categorical (assigned to one of a set of categories).
 - 1.2. Form Observed** – Record the raw data in the form observed before transformation, rounding, categorization, or other manipulation.
 - 1.3. Precision Observed** – Record the data with the precision observed. Example, record the number of animals as 7, not 7.0. In general, the number of digits should be neither too many (false precision) nor too few (rounding error.)
 - 1.4. Without Reduction** – Record the data without reduction, such as by summing, averaging, or taking percentages.
 - 1.5. Categorical Data** – Submit categorical data without collapsing across categories. Continuous outcomes which have been categorized should be submitted in raw form. For example, if white blood cell (WBC) counts have been recorded and then dichotomized into low (leucopenia) and normal, submit the raw WBC counts.
- 2. Organization of Data.** Present the data as a table.
 - 2.1. Content** – The table should contain only raw data. Do not enter data summaries, such as averages or totals. Do not include graphic enhancements such as lines, shading, or blank space. Do not merge cells.
 - 2.2. Headings** – Assign a single heading to each column and, if necessary, to each row.
 - 2.3. Entries** – Include a separate entry in each cell, even when entries for adjacent cells are identical.
 - 2.4. Missing Data** – Designate missing data by entering a single non-numeric code (such as ‘NA’ or ‘.’) in the appropriate position if using a text file. Cells may be left blank if using a spreadsheet.
 - 2.5 Text Data** – When inputting character data into electronic files, care should be taken to record the same value identically each time it is entered.

For example, if a technician is represented by the initials “jd”, there should be no entries such as “j.d.”, “JD”, or “ jd” (leading space), which will be treated as three additional different values by statistical software.

2.6 Redundant data – Avoid including the same data in more than one table, if possible.

Avoid submitting redundant data in the event an analysis was performed on a subset of the data or a joint analysis was performed by combining several datasets. Submit either the full dataset and provide subsetting information, or submit smaller datasets and indicate how they were combined for analysis.

3. Data Table Format. A data table is an array of observations. Each observation typically includes both a response and a set of covariates (explanatory variables or factors) associated with it.

3.1. Quantitative Response Data – Array the table so that each row contains a single unique set of covariate values associated with the response(s). To expedite timely and efficient review by CVB, applicants are **strongly encouraged** to submit data in this type of table.

A row may contain more than one response (e.g., lymphocytes, neutrophils, and platelets), but a row should not have varying values of covariates. For example, if white blood cell counts are measured daily over a series of days, there should be a separate row for each day. Do not put them on the same row. Similarly, for a series of optical densities measured over a sequence of dilutions, there should be a single row for each dilution.

3.2. Categorical Response Data – If there are only two factors including the outcome, the data may be submitted in the form of a contingency table or as a frequency table. If there are more than two factors, a frequency table should be used.

3.2.1. Contingency Table – A contingency table cross-classifies response counts.

In a contingency table, each cell shows the frequency of a unique combination of the categories which are identified on the margins of the table. They can be a useful visual display for categorical frequency data.

3.2.2. Frequency Table – A frequency table is a constant covariate table which has been collapsed over identical category combinations.

The explanatory variables in each row consist of a particular combination of the categories of the factors. The outcome in each row is the number of observations at that particular combination. To completely partition the data, the table should include a row for each possible unique combination of the categories of every factor. If any such combinations are omitted from the table, CVB will assume their frequencies are zero. There should be no more than a single row for every unique combination.

4. What to Submit

4.1. Data File - Submit the data table in an electronic file. Examples of file types received by CVB are SAS data files (*.sas7bdat), SPlus transport files (*.sdd), Excel spreadsheets (*.xls), formatted text files (*.txt), delimited text files (*.txt or *.csv), or text files containing character representations of R objects using the dump() function. Other file

types are also acceptable if they can be readily imported into Splus or SAS. If the data are submitted in an electronic file, it is not necessary to submit paper hardcopies.

4.2. *Data Descriptions* - Include a separate file indexing the data files and describing their contents. The description should state each variable's name (e.g., WBC), nature (e.g. white blood cell count), and measurement unit (e.g., 10^6 per mL). It should define all codes (e.g., A=Vaccine, B=Placebo) including the missing data code if applicable (e.g., missing data indicated by NA). It should state the number of observations and variables in the data set.

4.3. *Data Analysis* - Identify software used in the analysis. Review and processing of the submission will be expedited by including the programming code used to generate the output. For software that is entirely menu driven, include a session log. If a spreadsheet includes both data and analysis, they should be on separate worksheets so that the raw data are in accordance with section B.1. Include the cell formulae for the analysis.

5. Examples

Examples illustrating the types of data tables described in this document are available from CVB upon request.