

GENERAL FORMAT - INSTRUCTIONS

Tables

This format includes up to three tables, named *Variables*, *Individual*, and *Repeated*.

Variables – information about the variables in the other tables.

Individual – variables that appear only once, either because they are inherent to the individual unit, such as a subject's sex, or because they are measured only once, such as lesions at necropsy. The header of each column should have a unique variable name.

Repeated – repeated observations on each unit. The observations may be simple replicates, such as assays of replicate vials of a vaccine serial, or they may be sequential observations, such as daily blood cell counts over a two week period. The header of each column should have a unique variable name. One of the columns will contain the unit identifier linking the *individual* and *repeated* tables. Do not include other individual variables in the repeated table.

When creating when creating tables outside the template environment, always use column headers given in table descriptions.

Table: Variables

This table includes information about the variables in the other tables. It includes the following columns. Some of the columns are limited to specific values; if so, they are shown as an indented list.

Table – Which table is the variable in, *individual* or *repeated*.

Variable – The variable's name in the data set.

Label – A long name for the variable, suitable for use in tables and reports.

Data Role – The variable's role in the data set. Data types for each role are in parentheses.

Values of *Data Role* must be one of the following:

Identifier Unique identification given to each unit (or possibly subunit). The identifier also serves as the grouping factor for repeated observations made on the individual. (Data type: nominal)

Grouping Grouping factors are nominal categories important in study design, such as treatments or clusters. All efficacy studies have treatment groups, such as vaccine or placebo. Many have clusters, such as litter or pen. (Data type: nominal.)

- Observation* Direct observations or measurements. (Data types: continuous, count, ranked, dichotomous, ordinal, nominal.)
- Derived* Value calculated or derived from other variables. (Data types: continuous, count, ranked, dichotomous, ordinal, nominal.)
- Timing* Date or time. May be an actual date or time (Data type: date), or may be a count, such as the number of days in relation to an event such as challenge, or continuous, such as the number of minutes from inactivation. (Data type: count or continuous.)

Data Type – The variable’s type, roughly related to its distribution.

Values of *Data Type* must be one of the following:

- Continuous* A measurement on a continuous scale, such as body temperature. Large imprecise integer counts, such as blood cell counts, are often considered continuous. Interval censored continuous data, such as duration in days, may at times be treated as either continuous or integer.
- Count* Integer counts, such as frequencies.
- Ranked* A numerical value that reflects a valid ordering but a unit spacing that is not calibrated. For example, a valid clinical scoring scheme may produce integers from 0 to 16, where a higher number always indicates more severe disease than a lower number but the severity difference between any two adjacent scores is not necessary the same.
- Dichotomous* A binary response that can only take two values, such as dead or alive.
- Ordinal* Categories that are ordered, such as those applied to disease severity. The number of categories should be as few as possible and ideally would reflect distinct states, such as dead, sick, normal. (Excessive categories lead to issues of discreteness and sparseness.) Simple grading scales may generate ordinal categories.
- Nominal* Categories without any natural ordering, such as breed of cattle.
- Date* Date.

Units – The units of measurement. Units must be indicated for continuous and count data types. Leave blank otherwise.

Level Code – For categorical data types (dichotomous, ordinal, categorical), indicate the codes used for each category, one code per column. Thus the number of columns headed “Level

Code” will be the maximum number of categories in any categorical variable. (Example codes: 0, 1, 2; or IM, SC; or group1, group2, group3.)

Comments – Several different types of information may go under comments, depending on the data type or role. Use ‘Comments’ as the column header of all such columns.

For categorical data types, level labels corresponding to level codes may go here, indicated by ‘=’. (Example 0=none, 1=serous, 2=mucopurulent; or IM=intramuscular, SC=subcutaneous; each in a separate column.) If the level labels are the same as the level codes (e.g. BVD, IBR), they may be left blank.

For derived variables, indicate how they were derived. Use formula=something. (E.g. formula=width×length, or formula=see titration table.)

For the number of days in relation to an event, give the date corresponding to day zero. State it as origin=date.

Additional comments as necessary.