Pathogen Survival experiments

STRUCTURE OF THE DATA AVAILABLE

Field name	Data type	Cleaning notes
refID	Numerical	Unique identification of a reference.
FullReference	Free-text	Full reference in the format: "all authors, YEAR, publication title,journal,issue, pages".
groupID	Numerical	This field identify UNIQUE STUDY GROUPS within each paper. The format for this field has changed over the course of DACRAHs, but it remains UNIQUE within a paper (refID). It identifies experimental groups within the paper, so that during data analyses, it is possible to recognize multiple data forms that were filled to document different outcomes for the same group of animals, and differentiate results form the same study which are not the same group of animals.
agent	Categorical	The vector borne disease (VBD) agent. Data collected as a RADIO list.
agentSubtype	Free-text	Left as free-text, given the great variety of options possible. But during data cleaning, it was checked for consistency and standardized when possible.
agentDetails	Categorical	The category of subtype declared above, for instance serotype, genotype, strain and subspecies. Isolate was also in the list, but isolate was only given when other subtypes were not given.
agentSubtypeDC1	Categorical	In DACRAH1, only a fixed list of subtypes were provided as a radio list to data collectors. This field was later substituted by the 2 above.
targetSpecies	Categorical	The animal species from which infective material was derived. Options available to data collectors as a radio list.
sampUnitSize	numerical	This was checked during data cleaning. Note that there are a lot of "-1" (not given). This was double-checked, and it's indeed the case that many papers say, for instance, the number of animals in the total experiment, but not the number of samples per group or per time point.
matrix	Categorical	As much as possible, we tried to use **categories from EFSA's DCF** catalogues for matrix. Considering however how many different matrices were used, and the much varied amount of details given (for instance different concentrations of a chemical), **we chose to leave free-text comments in an additional field (see next).**
matrixDetails	Free-text	Used to give the matrix as provided in the paper, when deemed necessary given the categorization done during data cleaning (see above).
temperature	numerical	When not given, this was left blank - we avoided using "not provided" to enforce this field as numerical, but using "-1" or "99" didn't seem like a good option here, since these could be actual temperatures. We left blank when not provided, but during data quality it was double-checked that these were actually not given in the paper, rather than just not filled by the reviewer.
humidity	numerical	Checked for error.
labTest	Categorical	Laboratory test used for confirmation. Options available to data collectors as a radio list.
anMethText	Free-text	No data cleaning applied as this is meant to be a free description of the test details. This columns may not be available in the final cleaned dataset if data was never entered.
targetLab	Categorical	Target of the diagnostic tests, for instance nucleic acid, virus, antibody. Options available to data collectors as a radio list.
maxDetect	numerical	Latest day, from infection, in which the VBD was detected. Please note that this field can only be interpreted in light of the next field, which indicates if the experiment was truncated. When the experiment IS truncated, then the maxDetect listed here is only the end of experiment. A true maximum

		detection will be expressed as a number of days here, and a NO in the "truncated" question. Not all papers reported maximum detection, but instead, a calculated half-live. In these cases, a -1 value was added here (not given) and the half-live value was reported (see below).
truncated	Yes/no	Is the value of MaxDetect truncated? In other words: was the study finished before samples were negative. That is, is the reported "MaxDetect" equal to the last experiment day ?
halfLife	Numerical	This was added because some papers did not report the maximum detection day, but did provide an estimation of the half-life (in days). Confidence intervals were also added when provided.
LCI_halfLife	Numerical	Lower confidence interval (95% confidence) for the half-life, in days.
UCI_halfLife	Text	Upper confidence interval (95% confidence) for the half-life, in days. Left as text because some papers list as "Inf".
rowID	numerical	A unique number for all rows in the dataset.
uniqueID	numerical	A unique identification of all study groups in the dataset – a combination of refID + groupID.
ShortBibliography	Free-text	Reference in the format "First author, et al. YEAR".
Author	Free-text	List of authors
Title	Free-text	Publication title
Abstract	Free-text	Abstract
publicationYear	Free-text	Publication year.

NOTES AND WARNINGS ON DATA MEANING AND INTERPRETATION, ASSUMPTIONS AND SHORTCOMINGS

1) Data rows with the same refID are results reported from the same study
2) Individual study groups within these references receive the same groupID. These could be for instance a control and various treatment groups, groups of different species or age, or subjected to different experimental designs.
3) Combinations of refiD+ studyGroupID represent UNIQUE animal groups for which results are reported. These two fields should be used to identify multiple rows of outcomes that refer to the same animal group.
4) Data collection is performed in Distiller using "data collection forms". Each form results in one row when the data are looked in the tabular format (for instance in Excel of .CSV format). Every output can only be reported once in each form, therefore to report multiple values of the same type of outcome for the same group (say the detection window for different tests, or for different matrices), the entire form must be duplicated.

5) **Note that some studies were finished while VBDs were still detectable in the materials being studied. This censoring was explicitly recorded in the variable "truncated". However, this was not explicitly recorded in the first round of literature review. For this research objective, however, only 3 studies come from the first round of review (refID < 20000), and all other studies were reviewed in the second round of literature review.**