Experimental infections

STRUCTURE OF THE DATA AVAILABLE

Field name	Data type	Data format 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.
studyID	Numerical	A "studyID" field also exist, but it was used for different reasons in DACRAH1 and DACRAH2. It is studyGroupID that can be reliably used as a unique identification of groups within the same paper
year	Numerical	Year when the study was carried out. "-1" when missing in the paper.
agent	Categorical	VBD agent. Coded by EFSA. Collected in the form as a RADIO LIST.
agentSubtype Type	Categorical	Collected in the form as a RADIO LIST – so a fixed number of categories to choose from, but these category names are NOT CODED following any specific catalogue.
agentSubtypeType _comment	During data collection, the categories above had an option for "others", which allowed data collectors to enter additional comments in this free text field. All text entered was used for categorization during data cleaning, and no free-text was left. This column will show on the dataset because it existed on the forms, but should have no data after the data cleaning process.	
agentDetails	Free text	During data cleaning repeated subtypes were reviewed to ensure consistency and standardization. As much as possible, this was categorized.
agentSubtype	Categorical	This only existed in DACRAH1. In DACRAH1: Subtypes for some specific agents were entered in this variable as a RADIO LIST. This enforced categorization, but meant we could only collect the information for a handful of pre-set subtypes. Any additional information was entered in "agentDetails" as free-text, and as much as possible harmonized during data cleaned. In DACRAH2: we used instead "agentSubtype Type"+" agentDetails". The data collector would pick the subtype type (serotype, genotype, isolate, etc) in the categorical variable, and enter the details as free text in agentDetails. The data analysis scripts written during DACRAH3 account for that and merge the information.
targetSpecies	Categorical	Coded by EFSA initially, with more items later added by the Covetlab team (using codes from the same DCF catalogue). Collected in the form as a RADIO LIST.
studyTargetSpecies _comment	During data collection, the categories above had an option for "others", which allowed data collectors to enter additional comments in this free text field. All text entered was used for categorization during data cleaning, and no free-text was left. This column will show on the dataset because it existed on the forms, but should have no data after the data cleaning process.	
minAgeMonths	Numerical	Minimum age in the animal group being reported.
maxAgeMonths	Numerical	Maximum age in the animal group being reported.

ageMonths	Numerical	Only used in DACRAH1, and in DACRAH 2 substituted with min + max (above).
sampUnitSize	Numerical	Number of animals in the study group.
transplac	Binary	Evidence of transplacentary transmission: -1 when not given/not investigated, 0 when no evidence, 1 when confirmed.
hostHost	Binary	Evidence of direct host-to-host transmission: -1 when not given/not investigated, 0 when no evidence, 1 when confirmed
targetLabTest	Categorical	The target of the laboratory test (antigen, nucleic acid, etc). We have added "antibodies" and "not reported"
labTestUsed	Categorical	The specific test used to confirm infection. Coded by EFSA, except for three test types for which we could not find specific codes on EFSA catalogues: • Haemadsorption test (HAT) • Antigen capture ELISA (BTAC) • tests based on inoculation on eggs or day-old chicks ("inoculation assay").
labTestUsedC	Text	Any additional comments about the laboratory test above.
durationPI	Numerical	Duration of the Infection experiment. See data interpretation comments further below.
minIncub	Numerical	First day in which clinical signs were observed in any animal in the group. See data interpretation comments further below.
maxIncub	Numerical	Last day in which clinical signs were observed in any animal in the group. See data interpretation comments further below.
minDetect	Numerical	First day in which the VBD agent was detected in the specific listed matrix in any animal in the group. See data interpretation comments further below.
maxDetect	Numerical	Last day in which the VBD agent was detected in the specific listed matrix in any animal in the group. See data interpretation comments further below.
dead	Numerical	As per discussions with EFSA, several categories were added for the different reasons for death, and reviewers could put the number of dead for each specific category. No reported numbers mean NO DEAD animals, as "non reported" was specifically marked with a dedicated category. This field is categorical, but "checkbox", not radio, to allow multiple to be checked. As a result, in the cleaned dataset, this will show as several columns. Every reason for death listed is represented in the dataset as a pair of columns: a column to record whether that reason of death was reported or not, and one with the reported number of dead animals, if given. • deadInfection // nrDeadInfection • deadOther // nrDeadOther • eutInfection // nrEutInfection • eutOther // nrEutOther • eutProtocol // nrEutProtocol • eutEnd // nrEutEnd • mortNR (mortality not reported)
unitDead	categorical	Only on papers 202415 and 202559 percentages are used, all other numbers are absolute animal numbers. (please note that reviewers didn't always remember to check this to be "animal", but it is very clear from the data that it is). The data cleaning script will convert any mortality given to percentage when the unit is animal, and the sample size is given. These are calculated into the following columns: • percDeadInfection • percDeadOther • percEutInfection • percEutOther

		• percEutProtocol • percEutEnd
LCI_mortality	Numerical	Lower confidence interval for the mortality observed. Never provided among the papers reviewed.
UCI_mortality	Numerical	Upper confidence interval for the mortality observed. Never provided among the papers reviewed.
minMortalityTime	Numerical	Earliest day in which mortality was observed in any animal in the group.
maxMortalityTime	Numerical	Latest day in which mortality was observed in any animal in the group.
unitMortTime	Numerical	Reviewers didn't always remember to explicitly state that these were DAYS, but it's obvious from the data.
uniqueID	Numerical	A concatenation of refID and groupID, creating a unique identification of animal groups along the entire review.
ageCat	categorical	Created during data cleaning, based on the age in months given. Only for cattle (adults are those with age >=24 months) and small ruminants (adults when >=12)
route	Categorical	Route of infection of the animals. This field is categorical, but "checkbox", not radio, to allow multiple to be checked. Therefore, in the datasets each value becomes its own column. In the data cleaning process, we have concatenated all values using the separator "//". That is, multiple routes will show, as, for example, "routeIntravenous//routeSubcutaneous". We started from the categories coded by EFSA, but added more as needed according to the great variety of modes of infection reported in the paper. We leave at EFSA's discretion to code those, as it may be desirable to group some of these categories. We created a "not reported" category to make it explicit that the data is not available, rather than missed by reviewers. All forms therefore have some categorical data.
matrix	Categorical	Matrix sampled to test for the presence of the VBD agent. This field is categorical, but "checkbox", not radio, to allow multiple to be checked. Therefore, in the datasets each value becomes its own column. In the data cleaning process, we have concatenated all values using the separator "//". That is, multiple routes will show, as, for example, "routeIntravenous//routeSubcutaneous". We started from the categories coded by EFSA, but added more as needed according to the great variety of matrices reported in the paper. **We leave at EFSA's discretion to code those, as it may be desirable to group some of these categories**. We created a "not reported" category to make it explicit that the data is not available, rather than missed by reviewers. All forms therefore have some categorical data, except that, as mentioned the data users may want to further group categories.
ClinicalSigns	Checkbox + Numerical	As per discussions with EFSA, several categories of clinical signs were added. Reviewers could simply check a box to confirm that the clinical sign was observed/reported (variable "reportingCS"), and could also add the number of animals reported in each category (whenever numbers *were* reported, instead of just mentioning that the clinical signs were observed). In the cleaned dataset, every clinical sign is represented by a pair of columns. For fever, for instance: • CS_Fever – whether fever was reported or not

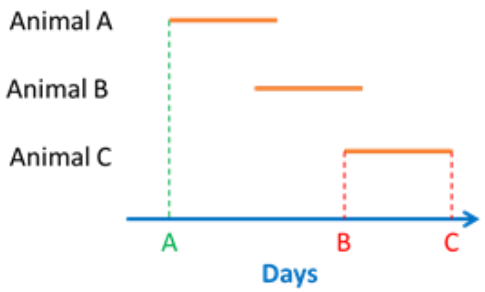
		• CSnr_Fever – the number of animals reported to have fever, when given.
rowID	numerical	A unique number for all rows in the dataset.
ShortBibliography	Free-text	Reference in the format "First author, et al. YEAR".
Author	Free-text	List of authors
Title	Free-text	Publication title
Abstract	Free-text	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 studyGroupID. 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. Say for instance that we have a group of animals, and clinical signs were observed for the first time in the group on day 5, and for the last time on day 8. Data collectors would enter minIncub=5, maxIncub=8. For the "detection period", however, the study authors reported two different viral detection tests: detection of virus in blood was positive from day 3 to day 10, and in feces from day 5 to day 9. There will be (at least) two rows of information for this study group, one stating minDetection=3, maxDetection=10, matrix=blood; and another one minDetection=5, maxDetection=9, matrix=feces. In BOTH forms, however, the values for window of observation of clinical signs would be reported as minIncub=5, maxIncub=8. *This is really important to consider when generating data summaries for specific parameters – the possibility of duplicated information within each study group needs to be carefully considered and accounted for (remove duplicates per group).*
5) It was not uncommon to have individual animals euthanized in the middle of the experiment to test matrices that can only be samples post mortem. This can cause the sample size to be different for multiple rows which were reported to refer to the same study group (the sample size to which each reported result refers to, at whatever time point they were recorded).
6) Particular attention should be paid to the following field definitions repeated from the table above:

durationPI	Numerical	Duration of the Infection experiment.
minIncub	Numerical	First day in which clinical signs were observed in any animal in the group.
maxIncub	Numerical	Last day in which clinical signs were observed in any animal in the group.
minDetect	Numerical	First day in which the VBD agent was detected in the specific listed matrix in any animal in the group.
maxDetect	Numerical	Last day in which the VBD agent was detected in the specific listed matrix in any animal in the group.

The specific way in which min and max were defined here is of crucial importance to interpreting results. Say we have an animal group with 3 animals, and the figure below represents the window of time during which each animal showed clinical signs. ***minIncub refers to time point A***, and ***maxIncub refers to time point C*** (NOT B). The same would apply for minDetect and maxDetect, if the figure below represented days positive to VBD detection.



Please consider the specific implications when analysing the data, among which:

A) The "min" (for clinical signs or detection) represents a distributions for the *earliest* time when positives can be expected to be observed, but are a poor representation of the "average" time to first observation, or of the maximum time an infection can go unnoticed.
B) The actual duration of the clinical signs period for each animal is not documented. Because the "max" is C (not B), maxIncubation does NOT refer to a true incubation period, but rather the end of a "clinical signs observed" period.
C) We can know when the value C is right censored, since we have recorded the duration of the experiment (DPI).
D) The day at which observations/detection *started* was always assumed to be 1, but not actually recorded. The value A can, in theory, be left censored, if observation/detection did not start until a certain time period. When that was obvious, however, we declared the "min" value to be "not given" (if for instance matrices were positive when first tested, but not tested until late in the study).
E) During data analyses, we re-checked EVERY case which had been reported to have min=max (for either incubation or detection). This would only happen if clinical signs/detection were only observed on a single day. Whenever this was a result of animals only being tested/observed on one time point, we have set the min to be unknown, the maximum to be the reported time point, and the DPI to also be that time point, so that it is clear that the value is censored.

META-ANALYSIS

Please refer to the document attached for a report of the meta-analysis explored and finally carried out with those data.