

Treatments

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. For treatments, these study groups could refer to different treatments, different species, etc. Groups could be followed along multiple time points, and this field should be used to identify results belonging to the same group.
country	Categorical	Coded by EFSA. Collected in the form as a RADIO LIST.
year	Numerical	Year when the study was carried out. "-1" when missing in the paper.
agent	Categorical	Vector borne disease (VBD) agent. Coded by EFSA. Collected in the form as a RADIO LIST.
agentSuptypeDC1	Categorical	Only used in DACRAH1, this used to be a list of VBD subtypes, but the list was not comprehensive.
targetHost	Categorical	Coded by EFSA. Collected in the form as a RADIO LIST.
targetVector	Categorical	Collected in the form as a RADIO LIST. Only filled when the study specified the vector, which was not always the case here, since in DACRAH3 we separated vectors and treatments. Only papers which presented results from the perspective of a HOST (number of infections prevented, cured, etc) were eligible on treatment.
ageMonths	numerical	Checked for consistency and completeness. All values left blank were checked by a second reviewer against the original paper to confirm that missing. In case of a specific age range, we entered the <u>lowest</u> age value.
ageDC1	categorical	In DACRAH1 age was collected as categories (adult or young), but from DACRAH2 this was substituted by the field above.
sampStrategy	Categorical	Sampling strategy. Coded by EFSA. Collected in the form as a RADIO LIST.
sampUnit	Categorical	Sampling unit. Coded by EFSA. Collected in the form as a RADIO LIST.
sampUnitSize	Numerical	Checked that used consistently (corrected when needed during a data cleaning process).
route	Categorical	Checked for consistency and completeness. Only left blank in one paper (64655) where it really was not described. Collected in the form as a RADIO LIST, the options started from a list provided by EFSA, and were adjusted as needed when encountering papers with non listed routes of administration.
route_C	Free-text	Data collectors had an option to add a free-text description about the route, when the route of administration was not among the list provided in the field above. As much as possible, routes were categorized and harmonised during data cleaning. Any non-categorized descriptions were left here.
intervention	Categorical	Control, pharmaceutical treatment or insecticide treatment (possible here still, but only papers where the outcome was given form the perspective of the HOST were eligible for use).
testSubstance	Free-text	Insecticide or pharmaceutical substance. Added as free-text during DACRAH1, and then harmonised during data cleaning. From DACRAH2, we created the categorical field below, and re-enter all data using RADIO lists. See field below.

SubstanceCat	CATEGORICAL for DACRAH2, but see notes	<p>The number of substances and variants was too varied, so in DACRAH 1 we had chosen to leave it as free-text, but thoroughly standardized the naming during data cleaning.</p> <p>Looking at it now, after we have collected so many papers, we can see that part of why it was so difficult to standardize this, was because the way substance, dose and concentrations are reported, is so varied. We have added now a CATEGORICAL question for test substance, and a free-text question for the formulation % (for instance 5%). We have classified all papers collected in DACRAH2 into one of the categorical substance names. As highlighted above, the complexity of substance provided was too great to fully categorize, so we also left the free-text option. For instance a paper testing a combination of different oils as repellent was categorized as “oils”, and in the free-text we entered specific details. We also left a category “others”, because it was simply impossible to categorize the great variety of ways authors have chosen to report the substances and various combinations of them.</p> <p>As a results, at the moment:</p> <ul style="list-style-type: none"> - All data from DACRAH1 was collected as TEXT, but thoroughly standardized (field “testSubstance”) <p>All data in DACRAH2 has a CATEGORIAL “SubstanceCat” classification, PLUS the original free-text description entered by authors in “test substance”.</p>
testSubstance2	Free-text	During DACRAH1 we noticed that very often, studies used substances in combination, and it was impossible to capture the substance used with purely categorical options. This resulted in a lot of free-text, that was hard to make useful. In DACRAH2, we have set the form to be able to collect all substance and dosage information for up to 2 substances. This was left blank when only one substance was used.
Substance2Cat	Categorical	Same as “SubstanceCat”, used if a second substance was used in combination with the first (categorisation of information given in “testSubstance2”)
substancePerc and substance2Perc	Free-text	The percentage given for the formulation (for instance 2% when the test substance was permethrin 2%).
dosageFreq and dosageFreq2	numerical	Checked for consistency and completeness.
dosageInterval and dosageInterval2	numerical	Checked for consistency and completeness.
Dose and dose2	numerical	<p>Checked for consistency and completeness.</p> <p>It was very hard to standardise this, against the actual concentration of the substance. In many cases, a concentration/formulation was given, AND a dose (for instance, 10ml of a 2% substance). In some cases, only the concentration was given, so the reviewers entered that concentration as the dose.</p> <p>The best way we could find to deal with the variety of formats here was to: capture the substance categorically (“SubstanceCat”), have a field for formulation (percentage), a field for dose, AND free-text comments for any additional details.</p>
doseUnits and doseUnits2	Categorical	We have used the same categories from DACRAH1, which are partly provided by EFSA, and part constructed after reviewing all papers in DACRAH1.
substanceDose_C	Free-text	Any further comments from the paper regarding the dose and administration
matrix	Categorical	The material collected for testing (specimen). Coded by EFSA. Collected in the form as a RADIO LIST.
labTest	Categorical	The laboratory test used for confirmation of infection. Coded by EFSA. Collected in the form as a RADIO LIST.

labTest_Description	Free-text	Any further details added.
targetLab	Categorical	The target of the detection test/indicator of infection – for instance “antibody” or “nucleic acid”.
timePoint	Numerical	Indicate time point in days from the start of the study to which the following parameters (dead, clinical and tested) correspond. Checked for consistency and completeness. All values left blank were checked by a second reviewer against the original paper to confirm that missing, and if so, a -1 was entered (no blanks). In case of a specific age range, we entered the <u>lowest</u> age value.
nAnimals	Numerical	Number of animals in the group. Checked for consistency and completeness
nClinical	Numerical	The format of reporting the results varied, an only results given in each paper were filled in (the rest left blank). A second reviewer ensured that an outcome was reported in a numerical objective format for every paper.
titres		
nTested		
nPositive		
nNegative		
deadUnit		
efficacy		
LCI_efficacy	Numerical	Efficacy lower confidence interval.
UCI_efficacy	Numerical	Efficacy upper confidence interval.
Further outcome types		<p>One paper (60323) gave the result as “RT-PCR parasite loads”, expressed as “parasites (mean number) per ng of DNA.”. Unsure how to fit this in the form, and expecting other types of unique results to come up (as it did in vectors), we added the three additional fields as a “catch all” for any results not expressed as mortality of efficiency:</p> <ul style="list-style-type: none"> • otherOutcome: a categorical list of extra outcomes we encountered. • Outcome numerical: a numerical field to add the numerical part of the result given • Outcome units. <p>Categories were added as needed whenever an outcome was not efficacy, number of positive, or mortality.</p>
outcomes_C	Free-text	Treatments and outcomes can have very complex designs, and all details that did not fit the forms were left here.
rowID	numerical	Unique identifier of rows in the dataset.
uniqueID	numerical	Unique identifier of groups for the whole dataset – it is a combination of refID and groupID
interventionStudy	categorical	Created during data cleaning, this variables sets the intervention (Pharmaceutical treatment or Insecticide Treatment) for the entire study. This is to allow filtering all groups by type of study, without the “control” groups within each study being left out from the filtering.
mortalityNr	numerical	The number of animals dead. This is calculated based on a formula that takes the reported number of dead animals when the “deadUnit” is individual, or the number of animals dead versus the number of animals, when “deadUnit” is percentage.
mortalityPerc	numerical	Mortality calculated as percentage. This is taken directly form the reported mortality when “deadUnit” is percentage, and calculated based on sample size when the number of dead animal in reported in individual units.
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 or .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) Please make sure to see notes above regarding the data collection and coding for variable “SubstanceCat” and all variables related to the substance concentration and dose.