Vector control

STRUCTURE OF THE DATA AVAILABLE

Please note that we have defined a ”vector control” study as a study in which the outcome is reported from the perspective of the VECTOR – for instance vector mortality, repellency, hatchability, etc. Papers which reported the outcome from the perspective of the HOST – generally infections prevented or cured, even if through vector control strategies – were classified as treatments.

Vector control studies are, therefore, expected to be inespecific for any VBD agent.

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”.
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	VBD agent. Coded by EFSA. Collected in the form as a RADIO LIST.
studyTarget_host	Categorical	Coded by EFSA. We have added the option “NOT HOST SPECIFIC”. Since in DACRAH2 the treatment and vector studies have been separated, this was the option that always applied here.
studyTarget_vector	Categorical	RADIO LIST: mosquitoes, midges, sandflies or ticks.
sampUnitType	Categorical	Coded by EFSA. It was difficult to categorize all the types of sampling units found in the papers. We added an option called “geo based unit”, and used it in case of villages or comparisons of vector mortality among different localities, when the use of “batch” or vector units was not possible form the information given in the paper. We enabled comments in this option, so that EFSA can review our use of this category if desired, and added as free-text the original description of units form the paper. When the study described, for example, using 3 replicated with 10 mosquitoes each, we reported the sampling units to be 3 replicate units. In the comments, we left the number of mosquitoes per replicate unit.
sampUnitType_C	Free-text	As explained above, we tried as much as possible to categorize the types of sampling units, but they were so varied, that we left extra comments in case a reclassification or further categorization is desired later.
sampUnitSize	numerical	The size of the sampling unit.
route	Categorical	We have used the same categories from DACRAH1, which are partly provided by EFSA, and part constructed after reviewing all papers in DACRAH1.
route_C	Free-text	We have categorized as many of the routes of administration as possible, but some individual cases were complex. We therefore created the route category “OTHER”, and in this free-text field, data collectors could enter additional comments/descriptions that we reviewed and harmonised during data cleaning.
intervention	Categorical	Control, insecticide treatment, or no intervention (the latter only used once, when results were not for a control group, but for a time point before treatment).
testSubstance1 testSubstance1_cat	CATEGORICAL for DACRAH2, but see notes	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. 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 <u>DACRAH2</u> 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. As a results, at the moment: - All data from DACRAH1 was collected as TEXT, but thoroughly standardized (field "test substance") - All data in DACRAH2 has a CATEGORIAL "test substance" classification, PLUS the original free-text description entered by authors in "test substance".
testSubstance2 testSubstance2_cat	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.
formulation1_Perc formulation2_Perc	Free-text	The percentage given for the formulation (for instance 2% when the test substance was permethrin 2%).
dose dose2	numerical	Checked for consistency and completeness. 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. We considered moving all the percentages to the field we created called "formulation", and leaving the dose empty. But then, we noticed that in some cases, the percentage does equally equate to the dose, for instance in WHO biassay protocols when the percentages tested are the doses. The best way we could find to deal with the variety of formats here was to: capture the substance categorically ("test substance CATEGORICAL"), have a field for formulation, a field for dose, AND free-text comments for any additional details.
doseUnits1 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.
dosageFreq1 dosageFreq2	numerical	How many doses were given.
dosageInterval1 dosageInterval2	numerical	The interval, in days, between doses.
substanceDose_C	Free-text	Any further comments from the paper regarding the dose and administration
sampledMatrix	Categorical	Material sampled for testing. Besides the categories provided by EFSA, we added the "Vector" category, which is always the case here, and for which we couldn't find an appropriate matrix code.
labTest	Categorical	The question regarding the laboratory test applied was added to all research objectives, but it is not actually relevant in the vector studies, where the outcome measured is mortality or presence, rather than testing subjects for presence of a VBD or an immune reaction.
labMethod_description	Free-text	It was obvious that the reviewers chose to report details of the assessment carried out here, since the actual list of laboratory tests is never applicable. This is free-text, and unlikely to be possible to categorize to use objectively. However, the type of assessment carried out should be obvious from the way we set out the fields to collect OUTCOMES – mortality, efficacy, and others reported below. When it wasn't, we added free-text here. See more below.
targetLabtest	categorical	For the reasons outlined above, never actually used.
timePoint	numerical	This was left blank when not reported or not relevant, but the mortality time (below) was always checked.
	Numerical	The type of results given were very varied. We describe further below how we capture outcomes different from mortality and efficacy.

dead (or mortality)		In the data collection form, a variables is available to report "number dead or mortality", with the units associated with the number reported in the next field. This field was used to report strictly mortality, but note that some papers evaluated a great quantity of concentrations or of times, and then, instead of reporting individual values, calculated the concentration that gave 50/90% mortality (LC50/LC90), or the time to mortality. These were always possible to capture in the current form because we have fields for mortality, time and concentration. But this should be kept in mind when looking at the results, so that it makes sense why, for instance, paper 70049 has a listed concentration of "17058.59ppm. That was the calculated concentration to reach 95% mortality after 24 hours. We believe the form format was appropriate and these details are captured, but those utilising the data should be aware of these details. During data cleaning, the reported mortality and mortality units were used to calculate the mortality in total numbers and in percentage. See below.
dead (or mortality) Units	categorical	Individuals or percentage.
mortNr	Numerical	Generated during data cleaning, this gives the mortality in numbers. This was copied from "dead (or mortality)" when the reported units was "individuals", and calculated from the sample size, when the reported units was "percentage".
mortPerc	Numerical	Generated during data cleaning, this gives the mortality in percentage. This was copied from "dead (or mortality)" when the reported units was "percentage", and calculated using the sample size as denominator, when the reported units was "individuals".
mortalityLCI	Numerical	Lower confidence interval for the mortality.
mortalityUCI	Numerical	Upper confidence interval for the mortality.
mortalityTime	numerical	Please note that "mortality time units" was also used to report mortality for other ouctomes (see below). So, for instance, if protection was reported as 100% after 4h, we added that percentage value to the outcome "efficacy", but added 4h **here**.
mortalityTimeUnits	categorical	Seconds, minutes, hours or days.
mortSec mortMin mortHours mortDays	Numerical	Calculated during data cleaning, based on the reported mortality time and units.
expHours exMinutes expDays	Numerical	Calculated during data cleaning, based on the reported exposure time and units.
efficacy	numerical	Checked for consistency. Whenever specific efficacy values were reported – for instance efficacy against mosquito landing, versus efficacy against biting, this was reported in the **labTest**, and the numerical values entered here.
efficacy_LCI	numerical	Lower confidence interval for the efficacy.
efficacy_UCI	numerical	Upper confidence interval for the efficacy.
outcomeType	categorical	Not all papers reported results as mortality, but they didn't necessarily calculate an efficacy form their results either. We therefore added variables to capture objectively other types of outcomes. These were: • Number of sampling units with vector presence (for instance, number of treated water jars which had larvae) • Number of vector units (when results were of the type counting vector presence, for instance in a hut with treated walls, and presented as absolute numbers, rather than efficacy in comparison to a control) • Hatchability
outcomeNumber	numerical	The numerical part of the outcome, for one of the outcomes listed in "outcomeType".
outcomeUnit	categorical	The units for the outcome, for one of the outcomes listed in "outcomeType".

nTested	numerical	This field is not applicable to most of the types of studies reviewed, in which vectors were not tested for the presence of the VBD, but observed for mortality, presence, hatchability, etc. However, some data collectors reported here number related to the units in the assays.
nNegative	numerical	This field is not applicable to most of the types of studies reviewed, in which vectors were not tested for the presence of the VBD.
nPositive	numerical	This field is not applicable to most of the types of studies reviewed, in which vectors were not tested for the presence of the VBD.
outcomes_C	Free-text	Any additional details and descriptions regarding the outcomes reported.
rowID	numerical	Unique id of all rows in the dataset
groupID	Numerical	This field identify UNIQUE STUDY GROUPS within each paper. For vector control, these study groups could refer to different treatments, different vector species, different geographical sites, etc. Groups could be followed along multiple time points, and this field should be used to identify results belonging to the same group (time point is recorded further below). Please note that the format of this field has changed over the projects. On DACRAH1, we had a dropdown list of numbers, but it became obvious that some papers had way more than 20 groups, so in DACRAH2 this is a numerical field. Distiller could not keep the information from DACRAH1 once we changed the type, so we created a new variable. The studyIDs for papers in DACRAH1 are actually recorded in a variable called "studyIDnumerical". For DACRAH2 and from now on, this information should be on studyID. The R scripts provided merge the information form the two columns into one.
uniqueID	numerical	Unique identifier of groups for the whole dataset – it is a combination of refID and 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 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

An important issue to document related to the vector studies, is that this research objective was NOT originally planned to be included in the literature review. In DACRAH1, EFSA set up data collection forms for a literature review focused on "preventive and curative treatments". As we started reviewing papers that were retrieved, it became clear that, for VBDs, a lot of preventive treatments were based on control of the vector. Some studies focused on the prevention of a specific VDB, while others were simply vector control studies. In DACRAH1, all these type of studies were reviewed within the same research objective, and data were collected using the same form.

In DACRAH2, it became clear that the forms originally designed for the literature review of treatments was not being able to capture the outcomes of the vector study controls. As a result, these two research objectives were separated, and a literature review focused specifically on vector control studies was triggered.

Unfortunately, a lot of the limitations of the data collection forms were not solved, as data collectors were mainly trying to fit the data into the forms originally designed by EFSA. Redesigning the forms form scratch was not considered a viable option, because of all the data already collected in DACRAH1 using the old forms.

More importantly, the complexity and variety of the study designs found, and in particular the lack of any reporting standardization, made it very hard to design a form that could capture all information reported in the studies reviewed.

As one example, the original form had a field “mortality time”, which was meant to capture the time between exposure of the vector, and the recording of the mortality. During data cleaning, we have added a “exposure time” field, which is the duration of the exposure. We then tried to enter this information retrospectively for as many of the papers as possible, either by reading all the extra comments that data collectors left on the forms, or very often by pulling back all the full-texts. It became clear, however, that the challenge wasn’t just redesigning the form and adding exposure time, but rather that in a lot of studies it was difficult to distinguish the “mortality time” from an “exposure time”.

Similarly, the field for formulation percentage was only created during data cleaning, and data collected retrospectively, trying to detangle the information entered in dose and dose comments.

To the best of our ability, we cleaned and harmonised all data collected in DACRAH1 and DACRAH2. But these limitations should be kept in mind when analysing the data.

One particular shortcoming of the data, resulting from the original design of the forms, is that the SPECIES of the vector and the LIFE STAGE (larvae, adults, etc) were NOT RECORDED.