Datafiles

Filename1: 20230524_WoS_systematic.xlsx

Naming convention of Filename1: [Date of search]_[Database searched]_[Literature sampling method].xlsx

Data type: tabular data in xlsx format.

Filename2: ReferenceList.pdf

Naming convention of Filename2: [Content description].pdf

Data type: Text list of references from which tabular data in '20230524_WoS_systematic.xlsx' was extracted.

Data collection

Data collection methods described below, have been taken from the publication in which this dataset is used. See the full text at https://doi.org/10.1098/rspb.2023.2811.

1.1. Literature search

To capture all studies on imidacloprid lethality conducted on adult insects, we used the Web of Science as our primary search engine, with the following query string:

((((TS=("imidacloprid")) AND TS=(insect*)) AND TS=("survival" or "toxicity" or "mortality" or *lethal*)) AND TS=("LD50" or "LC50" or "LD(50)" or "dose")).

We conducted this search initially on 2022/02/14, which yielded 596 results and formed the basis for our initial analysis. We subsequentially updated the search on 2023/05/24 to cover the most recent publications and obtained 72 new results. In addition, we added any papers we were aware of from our own databases if they did not appear in the results of the search in Web of Science (*N=3*). Such omissions in papers picked up by our search query occur because lethality tests are not always the primary emphasis of the experiment, and are therefore occasionally omitted from keywords, titles, or abstracts. Instead, lethality trials may be utilized, for instance, to determine a *sublethal* dose where primary experiments aimed to study consequences of sublethal exposure.

1.2. Inclusion criteria and data extraction

We thoroughly examined each of the papers from our search results to ensure they met our minimum criteria of providing LC50 or LD50 data for adult insects exposed to imidacloprid. We also excluded studies conducted on sub-adult instars (including larvae of holometabolous species) or

studies that lacked clear information on the methods and could not be allocated to a methodological category. In some cases, we found potential errors in reported data (e.g. inconsistencies in units between different figures) and we contacted the authors to confirm the correct value. In two cases the authors didn't reply, and the possible mistake was not obvious, so we excluded those papers and their data points from our dataset. In some cases, the reported LC50 or LD50 values were outside the limits of their reported 95% confidence interval. These values were excluded from the database.

For papers that met our inclusion criteria, we created one or more entries in our dataset. Simple studies of a single species, route of exposure and assessment timepoint (figure 1) generated a single database entry ('data point' or rows in the database). But often a single paper would also provide additional entries. For example, if a paper reported LD50s for 2 different species, with an LD50 calculated for both oral and topical exposure for each species, and with lethality assessed both at 24 and 48 hours after exposure for all species and exposure methods, this would have led to 8 entries. Studies on multiple populations and strains, if reported, were also treated as multiple entries (rows). Each of these entries are subsequently referred to as 'data points' in the text and figures. For each data point, we extracted: the year of publication; the name of the species; family; insect order; the strain used (if stated); any additional specific adult characteristics (for example, sex, forager status, newly emerged etc.); method / route of exposure (see figure 1); the imidacloprid formulation used (e.g. commercial formula or pure compound/active ingredient); the duration of exposure & time of observation (see figure 1); the location sampled; and the reported or calculated LC50/LD50 values with confidence intervals (when reported). Our final dataset is available from the Swedish National Data Service for review purposes *(*DOI: https://doi.org/10.5878/w6ct-z602).

1.3. LD50 estimation

For each recorded data point (i.e., either LC50 or LD50), if the mass-specific lethality was not given in the original article as ng/unit of body mass, we estimated it. In many papers, LD50 was given as dose/insect without specifying the body mass. In such cases, we searched the literature for alternative reports for the fresh body mass of adults of the same species. Where possible we matched the specific sex, caste or post-emergence age to that of the collected data point. In just one case [16] we could not find a reliable estimate of the fresh body mass. However, the authors reported the dry mass of the experimental animals in the paper. The values from this paper were included in our study but marked in figure S1 with asterisks and in figure S4 as open circles/diamonds, since the lower body weight results in a higher LD50 compared to LD50 calculated for fresh body mass. In some additional cases where a study documented lethality as LC50 but specified a particular treatment volume (such as during topical application), we again calculated an estimated LD50 based on the specified concentration and literature estimates of fresh body mass. We do not distinguish among different subspecies in figures (e.g. figure S1 and S4) and analyses.

File organisation

All data are contained in a single .xlsx file where each row in the excel sheet corresponds to a single LC50 or LD50 value extracted from a study. One study can have multiple datapoints if they have

studied multiple species, measured toxicity at different timepoints in a single species or examined the effect of different methods of pesticide exposure.

Variable description

1. Paper index: A numerical value used to identify each unique article within the database.
2. DOI of data: The Digital Object Identifier of the study from which the data for this entry has been extracted. The DOI is in link format and can be clicked to lead directly to the relevant paper.
3. Year: Year of publication of the manuscript for this entry
4. Species: The scientific name of the species for this entry. This includes subspecies name when provided in the study.
5. Family: Family level classification of the species for this entry.
6. Order: Order level classification of the species for this entry.
7. Study-specific strain/grouping: This is a tag used to identify different populations of a species within a study. Tags can be broadly relevant when they refer to well-established strains (e.g. Canton-S, wild type *Drosophila melanogaster*) or study-specific when they refer to populations by a unique group name (e.g. lig5). Tags are often used to differentiate between resistant and susceptible strains of a species.
    a. **Data example 1:** lig5
        i. A grouping specific to the study used to differentiate between different groups of the study
    b. **Data example 2:** WHO-SRS
        i. An internationally recognised strain standard used over multiple studies, in this case, denotes a susceptible strain
    c. **Data example 3:** field collected / field population
        i. Indicates that insects used in the study have been field collected and is likely the only specification towards insect strain in the paper
    d. **Data example 4:** from *C. sativus*
        i. Sometimes authors specified only the growth substrate, we use from *C. sativus* to indicate that these are insects grown on the Autumn crocus, *C. sativus*.
    e. **Data example 5:** susceptible population
        i. Indicates that insects were collected from the field, but are likely to be relatively susceptible to pesticide due to a history of low or no pesticide use in that region.
    f. **Data example 6:** susceptible
        i. Indicates a relatively susceptible strain within the study
    g. **Data example 7:** resistant
        i. Indicates a relatively resistant strain within the study, resistance could be acquired in multiple ways including prior field exposure and prior laboratory exposure over multiple generations.
8. Locations/Regions of origin: Refers to the location where the insect was collected and can be used to infer previous exposure to pesticides for this entry. Points to geographic or laboratory origin of the specimens for this entry.
    a. String labels indicate insect origin with highest precision on the left and lowest precision on the right followed by the substrate of growth and the date of collection

if available. The amount of detail provided vary across data entries and mirrors the information provided in the published articles, below are some examples of highly specific and less specific categories and how they should be interpreted.

   b. **Data example 1:** Citrus Research Institute, Sargodha, Punjab, Pakistan, 2017
      i. [Institute], [city], [province],[country] and [date of collection]
   c. **Data example 2:** Haidian, Beijing (39°570 N, 116°190 E) 2011.06 Eggplant Q
      i. [District], [city],[coordinates],[date of collection yyyy.mm], [growth substrate]
   d. **Data example 3:** Germany III
      i. Indicates Hive 3 from Germany
   e. **Data example 3:** local hives
      i. Denotes locally acquired insects, with unspecified location, likely matches the geographic location where the study was conducted.
   f. **Data example 4:** F. Matsumura 1986
      i. Flies supplied by an individual and established since 1986, [individual][yyyy]
   g. **Data example 5:** Reared in the laboratory
      i. Insects used in the laboratory were reared in the laboratory and no additional detail about insect origin has been supplied

9. Stage: Refers to the age of the adult insect when specified or the caste of the insect when applicable.
10. Method of Exposure: How the insect was exposed to pesticides for this entry. The method of exposure reported in the studies varied a lot and have been re-classified into 7 non-overlapping classifications for this database.
11. Vehicle (e.g. acetone): The vehicle in which the pesticide was delivered. Allows one to track what other compounds have been administered along with the pesticide.
12. Time of observation (h): The time of observation of mortality used to calculate the LC50 or LD50 value in this entry. Time is presented in hours.
13. Time of exposure (h): Duration of exposure to the pesticide for this entry. For Topical applications where the pesticide was directly applied to the insect cuticle or for direct feeding where the pesticide was directly fed to the insect, the exposure duration was arbitrarily set to one second unless otherwise specified in the study. All other measures are as reported in the studies and provided here in hours.
14. Formulation: Whether a pure pesticide (active ingredient only), a commercial product, or an unknown formulation was used. Classified as pure, commercial and unknown, respectively.
15. Type of Correction: When mortality data has been corrected for control mortality in a study, the type of mortality correction used is specified here.
16. Mortality correction: A binary indication (yes/no) of whether mortality correction was performed or not.
17. LC50 Value: Numeric value of the LC50 reported in the study for this entry.
18. LC50 unit: units in which the value was reported.
19. Number of replicates, when s.d./s.e. was provided: Only filled when the standard deviation or the standard error of the mean was provided as an indication of data spread for LC50 or LD50 values. It indicates the number of replicates of the experiment that were used to calculate the variation.
20. Type of Limit: indicates what kind of upper and lower limits were reported with the LC50 or LD50 measure. Includes confidence limits/intervals, fiducial limits/intervals standard deviation and standard error of the mean.
21. Lower limit: Lower boundary of the limits

22. Upper limit: Upper boundary of the limits
23. LD50 Value: Numeric value of the LD50 reported in the study for this entry.
24. LD50 unit: units in which the value was reported.
25. Mass (mg): Body mass of the insect in milligrams.
26. Mass data DOI: The Digital Object Identifier of study from which the body mass of the insect was extracted. Where this was supplied in the LC50 or LD50 study, it was prioritised and reported here as 'measured in paper' and the DOI of data can be used to replace the Mass data DOI. Some values are listed as 'estimated from paper' when the body mass was not directly provided but could be calculated using the data supplied within the paper from which the LD50/LC50 value was extracted.
    a. **Data example 1:** 'measured in paper'
        i. Insect body mass supplied by authors of the study where the LC50/LD50 values were obtained.
    b. **Data example 2:** 'estimated in paper'
        i. Insect body mass not directly supplied in the LC50/LD50 study but derived from other values supplied in the study
    c. **Data example 3:** 'https://doi.org/10.1002/ps.4729'
        i. DOI link to the paper from which insect body mass was extracted
27. Estimated LD50 (ng/mg): This is the standardised LD50 of the insect for this entry. It is reported in ng of imidacloprid per mg of insect body mass. When reported in the paper this value was reported directly here otherwise it was estimated by divided the reported LD50 value by the body mass previously extracted from relevant studies.
28. Category: Indicates the kind of toxicity endpoint we were able to extract from studies and falls into five categories.
    a. **Data example 1:** LC50 added mass and transformed
        i. The study only reported the LC50 value, we managed to find an estimation of body mass in other literature and used the supplied insect to calculate an LD50 value in ng/mg.
    b. **Data example 2:** LC50 only
        i. The study only reported the LC50, it was impossible to estimate insect pesticide intake.
    c. **Data example 3:** LD50_insect
        i. LD50 values reported in units of imidacloprid per insect. We used weight estimated from other studies to calculate the LD50 value in ng/mg.
    d. **Data example 4:** LD50_insect + weight
        i. The study reported the LD50 in units of imidacloprid per insect and provided the weight of the insects which we used to calculate the LD50 in ng/mg.
    e. **Data example 5:** LD50_weight
        i. The study directly reported the LD50 in ng/mg and we used this value in our Estimated LD50 column.
29. Note: Additional notes that highlight special considerations for some studies such as the use of dry body mass instead of fresh body mass for one study (Paper Index 117).

Null Values

All null values in the tabular data are indicated by a dash (-). In this dataset, a NULL values is attributed to data we could not obtain or estimate.

References

BRUUS, M., RASMUSSEN, J. J., STRANDBERG, M., STRANDBERG, B., SØRENSEN, P. B., LARSEN, S. E., KJÆR, C., LORENZ, S. & WIBERG-LARSEN, P. 2020. Terrestrial adult stages of freshwater insects are sensitive to insecticides. *Chemosphere,* 239**,** 124799.

BRUUS, M., RASMUSSEN, J. J., STRANDBERG, M., STRANDBERG, B., SØRENSEN, P. B., LARSEN, S. E., KJÆR, C., LORENZ, S. & WIBERG-LARSEN, P. 2020. Terrestrial adult stages of freshwater insects are sensitive to insecticides. *Chemosphere,* 239**,** 124799.