

STATEMENT

ADOPTED: 29 November 2021

doi: 10.2903/j.efsa.2022.7030

Statement on the active substance flupyradifurone

EFSA Panel on Plant Protection Products and their Residues (PPR),
Antonio Hernandez Jerez, Paulien Adriaanse, Philippe Berny, Tamara Coja, Sabine Duquesne,
Andreas Focks, Marina Marinovich, Maurice Millet, Olavi Pelkonen, Silvia Pieper, Aaldrik Tiktak,
Christopher Topping, Anneli Widenfalk, Martin Wilks, Gerrit Wolterink, Maj Rundlöf,
Alessio Ippolito, Alberto Linguadoca, Laura Martino, Martina Panzarea, Andrea Terron and
Annette Aldrich

Abstract

Flupyradifurone is a novel butenolide insecticide, first approved as an active substance for use in plant protection products by Commission Implementing Regulation (EU) 2015/2084. Following concerns that this substance may pose high risks to humans and the environment, the French authorities, in November 2020, asked the Commission to restrict its uses under Article 69 of Regulation (EC) No 1107/2009. To support this request, competent Authorities from France cited a series of literature papers investigating its hazards and/or exposure to humans and the environment. In addition, in June 2020, the Dutch Authorities notified the Commission, under Article 56 of Regulation (EC) No 1107/2009, of new information on flupyradifurone on the wild bee species *Megachile rotundata*. This notification is also referred to in the French notification on flupyradifurone. Consequently, the EFSA PPR Panel was mandated to quantify the likelihood of this body of evidence constituting proof of serious risks to humans or the environment. Therefore, the EFSA PPR Panel evaluated the likelihood of these studies indicating new or higher hazards and exposure to humans and the environment compared to previous EU assessments. A stepwise methodology was designed, including: (i) the initial screening; (ii) data extraction and critical appraisal based on the principles of OHAT/NTP; (iii) weight of evidence, including consideration of the previous EU assessments; (iv) uncertainty analysis, followed, whenever relevant, by an expert knowledge elicitation process. For the human health, only one study was considered relevant for the genotoxic potential of flupyradifurone *in vitro*. These data did not provide sufficient information to overrule the EU assessment, as *in vivo* studies already addressed the genotoxic potential of flupyradifurone. Environment: All available data investigated hazards in bee species. For honey bees, the likelihood of the new data indicating higher hazards than the previous EU assessment was considered low or moderate, with some uncertainties. However, among solitary bee species – which were not addressed in the previous EU assessment – there was evidence that *Megachile rotundata* may be disproportionately sensitive to flupyradifurone. This sensitivity, which may partially be explained by the low bodyweight of this species, was mechanistically linked to inadequate bodily metabolism processes.

© 2022 European Food Safety Authority. *EFSA Journal* published by John Wiley and Sons Ltd on behalf of European Food Safety Authority.

Keywords: flupyradifurone, butenolide, insecticides, genotoxicity, uncertainty analysis, environmental risk assessment, bees

Requestor: European Commission

Question number: EFSA-Q-2021-00159

Correspondence: pras.secretariat@efsa.europa.eu

Panel members: Paulien Adriaanse, Annette Aldrich, Philippe Berny, Tamara Coja Sabine Duquesne, Andreas Focks, Antonio Hernandez-Jerez, Marina Marinovich, Maurice Millet, Olavi Pelkonen, Silvia Pieper, Aaldrik Tiktak, Christopher Topping, Anneli Widenfalk, Martin Wilks and Gerrit Wolterink.

Declarations of interest: The declarations of interest of all scientific experts active in EFSA's work are available at <https://ess.efsa.europa.eu/doi/doiweb/doisearch>.

Acknowledgements: The Panel on Plant Protection Products and their Residues wishes to thank Emilio Benfenati and Thomas J Colgan for their support to the WG as hearing experts.

Competing interests: No competing interest was identified in line with EFSA's policy on declarations of interest.

Suggested citation: EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), Hernandez Jerez A, Adriaanse P, Berny P, Coja T, Duquesne S, Focks A, Marinovich M, Millet M, Pelkonen O, Pieper S, Tiktak A, Topping C, Widenfalk A, Wilks M, Wolterink G, Rundlöf M, Ippolito A, Linguadoca A, Martino L, Panzarea M, Terron A and Aldrich A, 2022. Statement on the active substance flupyradifurone. *EFSA Journal* 2022;20(1):7030, 55 pp. <https://doi.org/10.2903/j.efsa.2022.7030>

ISSN: 1831-4732

© 2022 European Food Safety Authority. *EFSA Journal* published by John Wiley and Sons Ltd on behalf of European Food Safety Authority.

This is an open access article under the terms of the [Creative Commons Attribution-NoDerivs](https://creativecommons.org/licenses/by/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited and no modifications or adaptations are made.



The EFSA Journal is a publication of the European Food Safety Authority, a European agency funded by the European Union.



Summary

Flupyradifurone is a novel butenolide insecticide, first approved as an active substance for use in plant protection products by Commission Implementing Regulation (EU) 2015/2084.

In November 2020, French Authorities asked the Commission to prohibit the sale and use of acetamiprid and flupyradifurone under Article 69 of Regulation (EC) No 1107/2009, in the light of potential concerns that these substances may pose high risks to humans and the environment. The French Authorities included, in their notification, scientific evidence to support this request, including references to published peer-reviewed studies. According to France, these studies indicate that, for acetamiprid and flupyradifurone, the approval criteria, referred to in Article 4 of Regulation (EC) No 1107/2009, are no longer fulfilled.

In addition, on 29 June 2020, the Dutch Authorities notified the Commission, under Article 56 of Regulation (EC) No 1107/2009, of new information on flupyradifurone on the wild bee species *Megachile rotundata*. This notification is also referred to in the French notification on flupyradifurone.

Consequently, the EFSA PPR Panel was mandated to advise the likelihood of this body of evidence constituting proof of serious risks to humans or the environment. Specifically, the EFSA PPR Panel evaluated the new studies aiming to quantify the likelihood of them indicating new or higher hazards and exposure to humans and the environment compared to previous EU assessments.

A total of 40 studies were referenced, which underwent an initial screening process based on predefined criteria. Upon screening, 16 studies were deemed relevant to the hazard assessment of flupyradifurone for humans ($n = 1$) or the environment ($n = 15$).

Among the bee studies, five references aimed to mechanistically explore differences in tolerance across bee species towards nicotinic acetylcholine receptor (nAChR) competitive modulators. These references were not entirely focused on acetamiprid or flupyradifurone, but were, nonetheless, retained in the assessment as supportive, read-across information.

All references retained after the screening underwent a full data extraction process, following which each measured endpoint was critically appraised following the principles of the Office of Health Assessment and Translation (OHAT)-NTP RoB assessment tool (NTP, 2019). For this purpose, *ad hoc* critical appraisal tools (CATs) were designed for the human health and environmental part, consisting of a series of questions aimed to quantify the relevance, reliability and precision of the assessments. For this purpose, each question was answered using a multiple-level scoring system. Upon appraisal, all endpoints and lines of evidence were summarised using heatmaps, where the overall classification of studies (i.e. the risk of bias, RoB) was calculated using predefined algorithms. Specifically, in these calculations, key questions for the assessment were given higher weight than others.

For the human health assessment, this step was followed by the quantification of uncertainties related to hazard identification (Step 1) and characterisation (Step 2). This was achieved by using a stepwise, hierarchical approach and a set of predefined factors/domains and related guiding questions tailored by lines of evidence. In a third step, experts were asked to compare the available evidence with the EU assessment by EFSA. Where deemed necessary this step was followed by an expert knowledge elicitation (EKE) process.

For the environment part, following appraisal similar data (i.e. assessment endpoints) were further collated into lines of evidence, where an additional indicator, the internal consistency, quantified how well these endpoints mapped together. Finally, the WG was asked to quantify i) the likelihood of each line of evidence indicating higher hazards than the EU assessment and ii) the uncertainty around this judgement.

The following key conclusions were drawn. For human health, only one relevant study was available for the genotoxic potential of flupyradifurone *in vitro*. These data did not provide sufficient information to overrule the EU assessment, as *in vivo* studies already addressed the genotoxic potential of flupyradifurone. Therefore, the PPR Panel concluded that the newly submitted evidence does not change the conclusion from EFSA on flupyradifurone and recommended no further action.

For environment, all available data investigated hazards in bee species. For honey bees, the likelihood of the new data indicating higher hazards than the previous EU assessment was considered low (acute and prolonged) or moderate (larvae), with some uncertainties that may need to be clarified. However, among solitary bee species – which were not addressed in the previous EU assessment – there was evidence that *Megachile rotundata* may be disproportionately sensitive to flupyradifurone. This sensitivity, which may partially be explained by the low bodyweight of this species, was mechanistically linked to bodily metabolism processes. Therefore, if a more up-to date risk assessment than SANCO (2002) will be used, the PPR Panel recommends that (i) for honey bees,

new laboratory experiments addressing chronic toxicity to adults and repeated exposure to larvae are carried out in accordance with the relevant OECD standards; (ii) the available higher tier honey bee studies are re-assessed against the principles of EFSA (2013); and (iii) for solitary bees, an appropriate specific risk assessment for the intended uses is performed considering the available data.

Finally, while acknowledging the purpose of this mandate, the PPR Panel considered that the elective selection of evidence may constitute an intrinsic bias to the assessment and, hence, to the conclusions reported above for both human health and the environment. Therefore, the PPR Panel recommends that systematic review approaches should be used in the future.

Table of contents

Abstract.....	1
Summary.....	3
1. Introduction.....	6
1.1. Background and Terms of Reference as provided by the requestor.....	6
1.2. Interpretation of the Terms of Reference.....	6
2. Human health.....	7
2.1. Data and methodologies.....	7
2.1.1. Data.....	7
2.1.2. Methodologies.....	7
2.1.2.1. Critical appraisal of the evidence (risk of bias).....	7
2.1.2.2. Data extraction.....	8
2.1.2.3. Uncertainty analysis and expert knowledge elicitation.....	8
2.1.3. Deviations from the protocol.....	10
2.2. Assessment.....	10
2.3. Data from the latest evaluation by EFSA.....	10
2.3.1. Critical appraisal results.....	10
2.3.2. Outcome of the uncertainty analysis and of the expert knowledge elicitation (EKE).....	11
2.4. Conclusion for human health part.....	12
2.5. Recommendation for human health part.....	12
3. Environment.....	13
3.1. Data.....	13
3.2. Methodology.....	13
3.2.1. Screening.....	13
3.2.2. Data extraction.....	13
3.2.3. Critical appraisal of the evidence (risk of bias and precision).....	14
3.2.4. Calibration.....	15
3.2.5. Weight of evidence and uncertainty analysis.....	15
3.2.6. Deviations from the protocol.....	16
3.3. Assessment.....	17
3.3.1. Results of the screening step.....	17
3.3.2. Bees.....	17
3.3.2.1. Data from previous peer review.....	17
3.3.2.2. Outline of the submitted hazard studies.....	18
3.3.2.3. Hazard characterisation and evaluation of the newly available data.....	21
3.3.2.4. Comparison of new data with previous hazard characterisation, weight of evidence and uncertainty analysis.....	28
3.3.2.5. Conclusion for bees.....	47
3.4. Conclusions for the environment part.....	49
3.5. Recommendation for the environment part.....	49
References.....	49
Abbreviations.....	51
Appendix A – Detailed results of the appraisal phase for hazard experiments (environment).....	52
Annex A – Protocol.....	55
Annex B – Outcome RoB flupyradifurone for Human Health.....	55
Annex C – RoB HeatMap flupyradifurone for Human Health.....	55
Annex D – Data extraction flupyradifurone for Human Health.....	55
Annex E – Uncertainty analysis table flupyradifurone.....	55
Annex F – Data extraction, weight of evidence and uncertainty analysis for mechanistic studies.....	55
Annex G – Detailed results of the appraisal for laboratory experiments with bees.....	55
Annex H – Detailed results of the appraisal for field experiments with bees.....	55

1. Introduction

1.1. Background and Terms of Reference as provided by the requestor

Acetamiprid is an active substance covered by the third batch of the renewal program for pesticides ('AIR3') in accordance with Commission Implementing Regulation (EU) No 844/2012. The active substance was first approved by Commission Directive 2004/99/EC and its approval was renewed by Commission Implementing Regulation (EU) 2018/113. A potential next renewal process needs to be initiated by 28 February 2031 at the latest.

Flupyradifurone is a novel butenolide insecticide, first approved as an active substance for use in plant protection products by Commission Implementing Regulation (EU) 2015/2084. To maintain the approval, a renewal process for this active substance needs to be initiated by interested applicants by 9 December 2022 at the latest.

On 30 November 2020, the French Authorities asked the Commission, under Article 69 of Regulation (EC) No 1107/2009, to prohibit the sale and use of these substances, taking into account the serious risks to health or the environment that their use may pose. Scientific evidence to support this request, including references to published peer-reviewed studies, were provided by France and the Netherlands.

By means of the mandate received on March 2021 from the European Commission, for flupyradifurone and acetamiprid, as foreseen in Article 69 of Regulation (EC) No 1107/2009, and for flupyradifurone under Article 56 of Regulation (EC) No 1107/2009 too, the Commission requested the EFSA PPR Panel to assess and explain whether:

- 1) based on the new information notified by France and the Netherlands and considering any other information available to the Panel from the recent evaluations by EFSA (2015a), including weight of evidence considerations, there are indications of a serious risk to human or animal health or the environment from the use of flupyradifurone;
- 2) based on the new information notified by France and considering any other information available to the Panel from the recent evaluations by EFSA (2016), and ECHA,¹ including weight of evidence considerations, there are indications of a serious risk to human or animal health or the environment from the use of acetamiprid.

1.2. Interpretation of the Terms of Reference

In line with the ToR, this EFSA statement aimed to assess the additional information provided by the French and Dutch competent authorities for the hazard identification and characterisation of pesticide active substance flupyradifurone. For the environmental part, the assessment is extended to the exposure characterisation, i.e. whether new routes of exposure to non-target organisms are identified and whether these are covered by the ones previously assessed.

This additional evidence complements the available one included in the latest evaluations conducted by EFSA to assess the impact on risk assessment.

In the human health part, it was first identified the toxicological assessment endpoints of interest in the area of genotoxicity.

In the environmental part, the working group (WG) firstly identified reliable tier-1 endpoints for most groups of non-target organisms from the previous peer review evaluation (summarised in the relevant EFSA conclusions). Any higher tier study available in the previous peer review was also considered, together with a mapping of the route of exposure/exposure scenarios previously deemed relevant for the risk assessment. In addition, situations where a high risk was concluded on the basis of the previous evaluations will be transparently reported in this statement.

For the studies newly submitted by France and the Netherlands, in both parts (i.e. human health and environment), an endpoint specific weight of the evidence was performed. Eventually, this culminated in an expert opinion on hazard identification and characterisation and impact on risk assessment, to support the decision making with regard to the application of Article 69 of Regulation (EC) No 1107/2009.

It should be pointed out that this statement is not based on a systematic review of all published and available information for the endpoints assessed, therefore, it is not excluded that additional work will be necessary outside the remit of this mandate.

¹ <https://echa.europa.eu/fr/information-on-chemicals/biocidal-active-substances/-/disas/factsheet/1235/PT18>

Working definitions

What is measured in experimental studies and the results of such measurements are often generically referred to as 'endpoints'. Other terms are also used, e.g. 'outcome', 'response', etc.

In order to make some clarity, working definitions are proposed here. These definitions should be interpreted as specific for this protocol. Similar, but slightly different definitions of the same terminology are reported elsewhere (e.g. U.S. EPA, 2003). This is not an attempt to overrule such existing definitions, but rather to make operative concepts that are relevant for the present project, and to ensure consistency between the assessment of human health and the environment.

Assessment endpoint: a parameter which is monitored and/or measured in one experiment. This may have a continuous, discrete, or dichotomic nature. Different assessment endpoints may be grouped in families of assessment endpoints when they refer to a common process (e.g. reproduction, development, DNA damage, apoptosis, oxidative stress, etc.)

Measured endpoint: the results of the measurements of the assessment endpoint. Depending on the nature of the endpoint, this may be expressed with a classification (e.g. positive/negative; present/absent) or with a quantification of an effect level by using a certain metric, often in comparison to a negative control. In some cases, the measured endpoint expresses the link between the effect level and the level of exposure triggering such effect.

2. Human health

2.1. Data and methodologies

2.1.1. Data

In support of the request to prohibit the sale and use of flupyradifurone in accordance with Article 69 of regulation (EC) No 1107/2009, the French and the Dutch authorities provided scientific evidence, including studies published in the open literature, on the potential serious risks that flupyradifurone may pose to human health and to environment.

For the evaluation of the human health data, two references mentioned in the mandate were screened for relevance for the human health risk assessment.

After a screening, only one of the two references (RefID 10 – Sekeroglu et al., 2018) was considered further in the assessment. The reference contained information on *in vitro* lines of evidence for the genotoxicity assessment endpoint category.

The other reference, Jeschke et al. (2015; RefID 11) reported structural considerations and docking models analyses for the assessment of insect specific site of metabolism where resistance to the pesticide was identified. Thus, this paper was considered as out of the scope for the current evaluation and therefore not in line with the ToRs.

2.1.2. Methodologies

Concerning the human health part, a predefined protocol was developed based on EFSA (2020) and reported in Annex A. The protocol includes both the problem formulation and the methodology planned for the assessment. Below only a brief summary of the methodology is reported for the sake of completeness. In addition, in Section 2.1.3, deviations from the original plan as described in the protocol are reported. The following steps were performed as part of the assessment: investigation of the internal validity using critical appraisal tools (risk of bias assessment); extraction of the relevant evidence; and data synthesis including uncertainty analysis.

2.1.2.1. Critical appraisal of the evidence (risk of bias)

Risk of bias (RoB) for the *in vivo* and for human observational studies was appraised using customised versions of the Office of Health Assessment and Translation (OHAT)-NTP RoB assessment tool (NTP, 2019). For *in vitro* studies, the tool used in the monograph on PFOS and PFAS (NTP, 2016) was adopted and adapted to fit the context of this assessment. Critical Appraisal Tools (CATs) were defined upfront and are described in the protocol (Annex A). Overall, the OHAT/NTP tool outlines 11 questions, grouped in six bias domains (selection, confounding, performance, attrition/exclusion, detection and selective reporting)

and one 'other sources of bias'. Table 1 shows the questions and domains appraised for the *in vivo*, *in vitro* and human lines of evidence with the agreed Key Questions for this specific assessment.

The evidence was appraised by at least two independent reviewers from the WG and EFSA staff using a 4-level scale. Answers were summarised at the level of individual studies and an algorithm was used to combine the answers to the appraisal question and to allocate the studies to the different classes: low (class 1), moderate (class 2) or high (class 3) RoB. Different weight was given to Key Questions as they are related to elements of the studies considered having a greater impact on the bias. Discrepancies in rating between assessors were solved through discussion with the WG members to reach the final recorded RoB rating for each question.

Eventually, the results of the appraisal were narratively reported in Annex B and graphically displayed in a heatmap (Annex C). The results were also contextualised in the uncertainty analysis step.

Table 1: Questions and domains appraised for the *in vivo*, *in vitro* and human lines of evidence with the agreed Key Questions for this specific assessment endpoints

Selection Bias	<i>In vitro</i>	<i>In vivo</i>	Human
Was administered dose or exposure level adequately randomised?	YES	YES	–
Was allocation to study groups adequately concealed?	–	YES	YES
Did selection of study participants result in appropriate comparison groups?	–	–	YES
Confounding Bias			
Did the study design or analysis account for important confounding and modifying variables?	–	–	Key Q
Performance Bias			
Were experimental conditions identical across study groups?	YES	YES	–
Were the <u>research personnel (cell maintenance and cell dosing)</u> blinded to the study group during the study?	YES	YES	–
Attrition/exclusion			
Were the measured endpoint data complete without attrition or exclusion from analysis?	YES	YES	YES
Detection bias			
Can we be confident in the exposure characterization?	Key Q	Key Q	Key Q
Can we be confident in the assessment of the results?	Key Q	Key Q	Key Q
Selective reporting			
Were all measured endpoints reported?	YES	YES	YES
Other bias			
Were there other potential threats to internal validity?	Key Q (cytotoxicity)	Systemic Toxicity	Statistics
Were there other potential threats to internal validity?	Replicates	–	–

2.1.2.2. Data extraction

Data were collected (i.e. extracted) from the provided studies by one EFSA staff and validated by another. A predefined form that comprises data on the characteristics of the study (study design, funding source, test system, species, ethnicity), the concentration/dose/exposure characteristics, the assessment endpoints and methods for measuring them, and the results was used to extract data at individual study level. The data model for extraction was tailored for each study type (i.e. *in vitro*, *in vivo*) and was provided (see Annex D). For flupyradifurone, only one specific assessment endpoint category dealing with *in vitro* lines of evidence was available.

It should be noted that the assessment endpoint category included in the uncertainty analysis was selected *a priori*, based on the endpoints measured and reported in the different studies (see Section 2.1.1 Data), while the specific assessment endpoints were selected as part of the appraisal step and not after the data extraction.

2.1.2.3. Uncertainty analysis and expert knowledge elicitation

For the only *in vitro* study retained for the assessment, the uncertainty analysis was performed within each hierarchical level (i.e. assessment endpoint category and specific assessment endpoint) to

support conclusions on hazard identification and hazard characterisation. The final purpose was to assess the impact of the additional evidence provided by the French and Dutch Authorities on the current assessments done by EFSA for flupyradifurone (EFSA, 2015a). A stepwise approach was used.

Differently from what was initially planned in the protocol, one additional question (Q3) was added to better reflect the aim of the assessment (see 2.1.3 protocol deviation n. 1). Moreover, the names of the active substances were no longer reported in Q1 and Q2. This is because there were many uncertainties in relation to the exposure characterisation (RoB class 3 for the majority of the studies) and therefore exposure reliability was considered a relevant uncertainty.

In step 1 and 2, the uncertainties related to hazard identification (Step 1) and characterisation (Step 2) were analysed. The uncertainty analysis was performed using a predefined list of factors/domains and related guiding questions tailored by lines of evidence. The factors/domains were assessed in two ways. First, potential explanations for the identified heterogeneity in the results (if any) were assessed. If inconsistencies could not be justified by any factor/domain, the unexplained inconsistencies were treated as a source of uncertainty. Second, the same factors/domains were appraised for adequateness in the body of evidence in relation to the specific endpoint/endpoint category/adverse outcome. Factors/domains considered not adequate were retained as sources of uncertainty. A detailed list of factors/domains by line of evidence is provided in Annex E (hereafter referred to as uncertainty tables). For both steps (assessment of the inconsistencies and of the potential sources of uncertainty), the judgement was achieved answering to domain and line of evidence specific 'guiding questions'. Synthetic answers (Yes/No/Not Relevant) and a narrative explanation for the rationale of the assessment were provided by EFSA Staff and checked by the WG.

The assessment was performed using a stepwise approach starting from the lower hierarchical levels and progressed at the higher levels (e.g. conclusions on the assessment endpoint category were based on those achieved for the specific assessment endpoints). Progression of the assessment towards a higher level (e.g. assessment endpoint category – genotoxicity) was carried out also if at the lower level (i.e. specific assessment endpoint) the measured endpoint was not affected in dose or concentration response relationship. This approach was taken to allow drawing conclusions on all the assessment endpoints categories identified in the scientific evidence provided by the French and the Dutch authorities.

Based on the answers to the 'guiding questions' a judgment was made on:

- specific endpoint being associated/affected in a dose/concentration–response relationship in the evaluated study (Q1 in Table 2).
- minimum dose/concentration at which the assessment endpoint is perturbed in the study evaluated (Q2 in Table 2).

Table 2: Assessment questions for the uncertainty analysis on hazard identification and characterisation

Line of evidence	Question 1. Hazard identification	Question 2. Hazard characterisation	Answer
<i>In vitro</i> experimental studies	Is the measured endpoint affected in a concentration-response relationship in the evaluated study?	What is the lowest concentration at which exposure affects the endpoint?	(Q1. Yes/No + Q2. Lowest concentration/dose) + summary of the uncertainties for the assessment endpoint category
<i>In vivo</i> experimental studies	Is the measured endpoint affected in a dose-response relationship in the evaluated study?	What is the lowest dose at which exposure affects the endpoint?	

In step 3, experts were asked to assess the contribution of the available evidence on the conclusions currently reached by EFSA for flupyradifurone (Q3 in Table 3).

Where necessary, and in line with the recommendation from the experts, Step 3 was followed by an expert knowledge elicitation (EKE) process (EFSA, 2014). If the experts' recommendation did not include an EKE, the process ended here. This was the case when all the available evidence in the updated data set (including the new evidence and the evidence already available in the EFSA conclusion) was already sufficient to conclude without the support of the EKE for the weight of evidence (WOE) analysis. The purpose of the EKE, when conducted, was to express the uncertainty using a quantitative WOE approach to address Q3. In this case, the uncertainty was quantified as probability (i.e. very low, low, moderate and high). For flupyradifurone, based on the results of the uncertainty analysis conducted, the process stopped here.

Table 3: Assessment questions for assessing the contribution of the available evidence on the conclusions currently reached by EFSA for flupyradifurone

Line of evidence	Question 3	Answer
<i>In vitro</i> experimental studies	Is the available evidence able to modify the conclusions currently reached by EFSA for flupyradifurone?	Yes/No + Recommendation on the assessed endpoint (including EKE where necessary)
<i>In vivo</i> experimental studies	Is the available evidence able to modify the conclusions currently reached by EFSA for flupyradifurone?	

EKE: expert knowledge elicitation.

2.1.3. Deviations from the protocol

- 1) For *in vivo* and *in vitro* studies, differently from what was initially planned in the protocol, one additional question (Q3) was added. The name of the active substance is no longer reported in Q1 and Q2 to better reflect the uncertainties in exposure (please, refer to Section 2.1.2.3 for more details).
- 2) The Roulette method proposed in the protocol was not applied. A customised version of the OHAT approach (NTP, 2015) was used instead to integrate the available evidence and to rate the certainty in a causal and positive association between exposure and health outcomes. This protocol deviation also accounted for lack of a quantitative estimation of the uncertainties as it was planned for the Roulette method.

2.2. Assessment

2.3. Data from the latest evaluation by EFSA

Flupyradifurone was first approved as an active substance for use in plant protection products by Commission Implementing Regulation (EU) 2015/2084. Its approval was renewed for a period of 10 years by Commission Implementing Regulation (EU) 2015/2084 and, to maintain the approval, a renewal process for this active substance needs to be initiated by interested applicants by 9 December 2022 at the latest.

The following is a summary of the peer review conducted by EFSA (2015a) for the toxicological assessment endpoints categories identified in the newly provided scientific evidence (i.e. genotoxicity).

Genotoxicity

Flupyradifurone was tested in both *in vitro* and *in vivo* test to assess the genotoxicity potential. *In vitro* the active substance did not induce gene mutation in the Ames test and in the mammalian cell study (CHO/HPRT). *In vivo*, flupyradifurone was found to not be genotoxic in two mouse bone marrow micronucleus tests; though, from the available data there was no evidence of bone marrow toxicity.

Based on these results, it was concluded that flupyradifurone is unlikely to be genotoxic.

No metabolites of toxicological concerns were found. Although, the difluoroethyl-amino-furanone (DFEAF) was found to be positive in an *in vitro* chromosomal aberration test with metabolic activation, this alert was not confirmed in the *in vivo* follow-up studies.

2.3.1. Critical Appraisal Results

For flupyradifurone the results of the appraisal were narratively reported in Annex B and graphically displayed in a heatmap (Annex C). A summary of the results is however included in the following lines and graphically displayed in Figure 1.

All the toxicological assessment endpoints were used for the evidence synthesis in line with the ToRs of the current mandate. The toxicological assessment endpoint categories were defined *a priori* based on the item proposed by the different studies.

IN VITRO – Flupyradifurone (4 specific assessment endpoints)	
Category	<p style="text-align: center;">GENOTOXICITY</p> <p style="text-align: center;">(4 specific assessment endpoints)</p>
RoB	Class 1 (n = 4)
Specific assessment endpoint	Chromosomal aberrations (CA) Micronucleus (MN) Mitotic index (MI) Nuclear division index (NDI)

Figure 1: Summary of the RoB conducted for the *in vitro* lines of evidence. The results were reported per assessment endpoint categories (i.e. genotoxicity) and per specific assessment endpoint

2.3.2. Outcome of the uncertainty analysis and of the expert knowledge elicitation

The contribution of the additional information provided by France and the Netherlands was assessed comparing the specific assessment endpoints and the overall body of evidence considering the evaluation conducted by EFSA, as required by the ToRs.

The uncertainty analysis table used to perform this evaluation includes information on the studies reported in the EFSA conclusions (EFSA, 2015a). The analysis provides a comparative assessment of the new data vs. the conclusion on the same toxicological assessment endpoint category used in the process of hazard identification and characterisation. In addition, a conclusive position of the PPR Panel on the impact of the new submitted studies on the current assessment (EFSA, 2015a) is also reported, which includes a recommendation on further steps necessary to fulfil the ToRs (Annex E).

The Annex E also includes an uncertainty analysis for endpoints used to define cytotoxicity and/or to establish the maximum concentration tested in the cell assay (i.e. mitotic index, nuclear division index). These endpoints were not further considered in the assessment and were included in the uncertainty analysis as complementary evidence to define how specific the observed effects for the toxicological assessment endpoints category were.

For the toxicological endpoint category **genotoxicity**, the PPR Panel concluded, following detailed assessment of the available evidence and uncertainties, that moving to the EKE was not necessary. The available evidence from the study by Sekeroglu et al. (2018) is not able to modify the conclusion reached by EFSA (2015a). The details of the uncertainty analysis showed that in this *in vitro* study there was a statistically significant increase in chromosomal aberrations without metabolic activation at the two highest concentrations tested and that the number of micronuclei was statistically significant increased at all the concentrations tested with metabolic activation and at the highest concentration tested without metabolic activation. Based on this evaluation, and considering the available database

on genotoxicity evaluated by EFSA (2015a), it was concluded that there was not sufficient evidence to move to the EKE for the assessment endpoint category genotoxicity and the current assessment provided by EFSA (2015a) was considered still valid.

It should be noted that following the uncertainty analysis of the newly submitted study, an overall assessment of the uncertainties associated with the full data set (including data from the newly notified study and EFSA conclusion) was conducted and described in a narrative way in the following lines (see Section 2.1.1 deviation n. 2):

- *In vivo* micronucleus (MN) tests (n = 2) conducted on mouse bone marrow were available in the EFSA conclusion (EFSA, 2015a). They were conducted using the intraperitoneal route of administration. This route of administration is expected to maximise exposure; however, a proper investigation of potential metabolites remains uncertain. Testing for aneugenicity after metabolic activation may therefore be limited. This was considered an uncertainty since in the *in vitro* MN test the results were observed after metabolic activation. However, in line with the EFSA guidance on aneugenicity (EFSA Scientific Committee, 2021) it is expected that the large majority of substances would induce aneugenicity without metabolic activation; nevertheless, an effect consequent to metabolic activation cannot be fully excluded.
- Evidence of bone marrow exposure was not properly described in the two *in vivo* MN bone marrow studies conducted in mouse; however, from the available data there is no evidence of bone marrow toxicity.
- There is uncertainty on the appropriateness of the high dose level selection and if the two *in vivo* MN bone marrow studies in mouse were conducted at the maximum tolerated dose. For both studies, clinical signs indicative of systemic toxicity, e.g. apathy, roughened fur, loss of weight, sternal recumbency, spasm, periodically stretching of body and difficulty in breathing, were observed at all doses in male mice. Similar changes were also observed in female mice but starting from mid-dose.
- Lack of fluorescent *in situ* hybridisation (FISH) or crest staining adds additional uncertainty on the nature of the results, i.e. clastogenicity vs. aneugenicity.

Although the study from Sekeroglu et al., (2018; RefID 10) was considered of good quality (class 1 in the RoB assessment), the outcome from *in vivo* studies included in the EFSA conclusion (EFSA, 2015a) was considered as sufficient evidence for addressing the genotoxicity potential of flupyradifurone in line with the Scientific Committee Opinions on genotoxicity (EFSA Scientific Committee, 2011b, 2017). Therefore, the additional study submitted does not trigger a new concern and does not provide any solution for addressing the uncertainties listed above.

2.4. Conclusion for Human Health part

In line with the ToRs, the contribution of the additional information notified by France and the Netherlands on the evaluation of flupyradifurone conducted by EFSA (2015a) was assessed in the current statement.

For genotoxicity, one study in which the genotoxic potential of flupyradifurone was tested *in vitro* was submitted. Considering that the results from *in vivo* studies conducted as part of the submission of the renewal dossier (EFSA, 2015a) provided sufficient evidence for addressing the genotoxic potential of the tested compound, it was concluded that the additional study provided by the authorities does not add additional concern and does not provide solutions for addressing the existing uncertainties.

Another study notified by the authorities (Jeschke et al., 2015) provided no data relevant for the risk assessment for human health concern as it reported structural considerations and docking models analysis for the assessment of insect specific site of metabolism where resistance to the pesticide is identified.

The PPR Panel concludes that the additional information notified by the authorities does not modify the conclusions reached in the evaluation by EFSA (2015a).

2.5. Recommendation for human health part

The only toxicological assessment endpoint category identified for flupyradifurone was genotoxicity. The current assessment was made on selected scientific evidence notified by French and the

Netherlands authorities. The PPR Panel concluded that the newly submitted evidence does not change the conclusion from EFSA on flupyradifurone and recommends that no further actions should be taken.

The PPR Panel recommends that elective selection of evidence, as it was done for this mandate, should be avoided and that a systematic review approach should be instead applied in the future.

3. Environment

3.1. Data

In support of the request to prohibit the sale and use of flupyradifurone in accordance with Article 69 of regulation (EC) No 1107/2009, the French authorities provided scientific evidence, including published studies, on the potential serious risks that the above-mentioned substances may pose to human health and to environment. In addition, the Dutch authorities submitted, under Article 56 of Regulation (EC) No 1107/2009, new information on flupyradifurone on wild bees. The mandate received from the EU Commission included also an assessment of the substance acetamiprid, for which data were submitted by the French authorities.

For the evaluation of the environmental data, all 40 references mentioned in the mandate were screened for relevance for the environmental risk assessment. After a first screening (see Section 3.3.1), information on flupyradifurone was available uniquely for bees.

3.2. Methodology

Concerning the environmental part, the full methodology used for the assessment is reported in the protocol (Annex A). Below only a brief summary of the methodology is reported for sake of completeness.

3.2.1. Screening

All documents submitted by France and the Netherlands underwent a screening phase to identify whether each document reported potentially useful information for the environmental risk assessment. Papers were considered relevant if they contained:

- **data** potentially informing the assessment/quantification of **hazard** and/or **exposure** for acetamiprid and flupyradifurone. In this case, papers were also classified on the basis of the type of experiments reported (e.g. laboratory, field effect, field exposure) and on the basis of the non-target group investigated.
- **mechanistic data** that support the explanation of the difference in tolerance between bee species, not necessarily related to acetamiprid and flupyradifurone. The focus of the available papers was mostly on the activity of specific enzymes belonging to the superfamily of cytochromes P450 (CYP). Some of these enzymes are known to play a role in the phase I detoxification pathways, and thus the presence/absence of some specific enzymes may drive the difference among experimental sensitivity. None of the assessment endpoints measured in these experiments can be used as input in any existing risk assessment model. Nonetheless, it is considered that these experiments may contribute to increase the mechanistic understanding behind the toxicity of some insecticides towards bees, and they may also be used as lines of evidence to aid the extrapolation of toxicity information from one species to another.

3.2.2. Data extraction

The data extraction process was performed differently for hazard/exposure experiments and mechanistic experiments.

Particularly for hazard data, the measured endpoints which can inform the environmental risk assessment for both flupyradifurone and acetamiprid were extracted using a structured data model. This step was implemented in the web-based tool DistillerSR. Extraction was performed by one reviewer, followed by a thorough checked by another reviewer (quality check). Extraction data models were tailored to the different study typologies, and in particular, they were different for laboratory and field studies.

For mechanistic data, the extraction was not performed following the same detailed structure used for hazard/exposure studies. The data extraction was on the contrary performed in a more narrative

way, also due to the difficulties in finding a common structure for summarising the findings of very diverse experiment types.

3.2.3. Critical appraisal of the evidence (risk of bias and precision)

In this step of the process, the Risk of internal and external Bias (RoB) and (im)precision was assessed separately in relation to each assessment endpoint.

Internal bias refers to any error in the conduct of the study that results in a conclusion which is different from the truth we are interested in. The method for measuring any assessment endpoint not being reliable/accurate is an example of source of internal bias in the studies relevant to this assessment. This term is often referred to as the intrinsic reliability of the assessment endpoint.

External bias affects the extent to which the study results are generalisable to the assessment question, e.g. when the study settings are not being representative of the reference population/conditions/landscape settings. This term is often referred to as the relevance of the assessment endpoint.

The third aspect next to internal and external bias that was assessed concerns the possible imprecision of the studies included in the assessment, which is related to random error and indicates the ability of a study to provide similar results when repeated under the same conditions. These aspects are mainly related to the sample size of the studies, which may not be large enough for providing a precise estimate of the assessment endpoint, resulting in an imprecise measured endpoint. Similarly, precision on the measured endpoint may depend on the number and the selection of the tested exposure levels.

For **hazard/exposure experiments**, internal and the external validity (or risk of internal and external bias) and (im)precision were appraised for each individual study using different critical appraisal tools (CAT). A 4-level rating was used for internal and external validity, in line with the OHAT/NTP tool for RoB assessment (NTP, 2015) and the human health assessment. Assessment of precision only used a 2-level scale as previous experiences (e.g. EFSA et al., 2020) demonstrated that establishing thresholds for intermediate categories can be extremely challenging for this part of the appraisal.

After a preliminary screening of the studies to be assessed, CATs were developed for different study typologies, which include:

- Laboratory studies investigating effects on bees
- Laboratory studies investigating effects on aquatic organisms
- Laboratory studies investigating effects on soil organisms
- Field studies investigating potential effects on bees
- Field studies providing information on exposure to bees (only external and internal validity)

A single study investigating the effects of acetamiprid on birds was also available. For this, no specific CAT has been developed, and the study was assessed following the principles included in the other CATs and elements of the standard OECD test guidelines for birds (e.g. OECD TG 206). The tools were translated in a digital form using Distiller SR. Appraisal for the only bird study was done outside of this tool. For each study, the appraisal was independently performed at assessment endpoint level by two reviewers. In agreement with the protocol, any disagreement was first discussed among the two reviewers and, if no solution was possible, the issue was discussed by the whole WG.

For each of the CATs, key questions and non-key questions were identified in order to assess internal and external validity and precision. Key and non-key questions were combined into a single scoring method, classifying each assessment endpoint from each study into a different class (from class 1 to class 3) reflecting the RoB.

Questions were considered key when a PH RoB or a DH RoB would immediately cause the assessment endpoint not to achieve the highest class. Key questions have also a higher weight in determining whether the assessment endpoint can achieve a class 2. Classification of questions in key and non-key was largely based on validity criteria from the most relevant OECD test guidelines, but it was also complemented by expert judgment and it considered the objectives highlighted in the most relevant guidance documents for the risk assessment.

It should be highlighted that a high RoB for key criteria did not translate in the dismissal of the assessment endpoint. All endpoints were considered in a final WOE (see Section 3.2.5), whether they were considered critical or not. This was done to provide a more transparent and comprehensive picture of the available information.

For **mechanistic experiments**, the appraisal was performed in a more narrative way. Since none of the assessment endpoints contained in those experiments will be directly used to quantify the hazard and/or the exposure, the need of classifying those into a specific 'risk of bias level' was deemed limited. Thus, while criteria guiding such appraisal were defined *a priori* (see Annex A) these were uniquely used as guiding principles, and no explicit categorisation of the RoB was performed. In this case the appraisal was done by one reviewer and later checked by a second reviewer.

3.2.4. Calibration

The full process involving screening, data extraction and appraisal underwent a calibration exercise involving a limited number of documents ($n = 3$). This was used to check the status of alignment among reviewers and to identify critical aspects that needed further clarifications and better definitions in order to avoid different interpretations of the same criteria.

3.2.5. Weight of evidence and uncertainty analysis

This part of the methodology was not fully detailed in the protocol, as this required an approach tailored to the available data, whose knowledge was limited before the full extraction and appraisal.

Initial assessment

The outcome of the critical appraisal was summarised using heatmaps. This data visualisation tool allowed to quantitatively synthesise precision, external and internal validity relative to each appraisal question (Appendix A). Additionally, the overall classification of precision, external and internal validity for each endpoint was first calculated using the algorithm described in Annex A, and then summarised using the same data visualisation tool described above (see figures under Section 3.3.2, as an example). The latter heatmaps were used to inform the evaluation of the available evidence, primarily, as a screening tool to identify the scores of reliability, relevance and precision. Additionally, heatmaps were used to support the identification and grouping of similarly relevant endpoints. These groups later defined the lines of evidence used in the final assessment.

Identification of the lines of evidence and comparison with previous endpoints

Given the heterogeneity of study designs and complexity of data, the identification of the lines of evidence required a certain degree of expert judgement and, therefore, could not be fully standardised across studies. Nonetheless, a significant effort was made to harmonise the approaches used across non-target organisms and study types. To facilitate the synthesis of endpoints, results of the data extraction were arranged by study typology, exposure regime and assessment endpoint type. Then, the resulting endpoint groups were graphically plotted using standard data visualisation tools (Wickham, 2016; R Core Team, 2021). A limitation of this approach is that data visualisation tools are intrinsically limited by the number of aesthetics which can be assigned to given variables. Therefore, a careful choice of the type of data aggregation was required on a case-by-case. This is particularly relevant, since studies were heterogeneously designed and not standardised. However, this should not be considered a major limitation, given that data visualisation was used as a tool for, and not the outcome of the WOE. Additionally, because of the nature and heterogeneity of data, and consequent to the data extraction process, standard research synthesis methods (e.g. meta-analytical approaches) were not deemed practical.

Plots were standardised in the following aspects:

- i) The x axis (continuous) represented the exposure level.
- ii) The y axis (factor) identified specific combinations of study and experiment ID
- iii) The aesthetics (i.e. dot size, shape and colour) were assigned to the most relevant combination of grouping variables for a given line of evidence (i.e. species; exposure route; internal validity; effect level and assessment endpoint type)
- iv) Where plots were arranged in multiple panels, the latter were used to display and sort endpoints by external validity or assessment endpoint type
- v) Whenever the exposure regime used in the studies under assessment was comparable (or could be approximated) to the standard regimes used across studies of the EU assessment, the relevant EU agreed endpoints were also plotted in the same graphs as vertical dashed lines.

Only comparable exposure units were used in a single line of evidence. Whenever possible, concentrations were converted accordingly. Whenever conversion to the same exposure units used in

risk assessment was not possible, endpoints were discarded. Indeed, harmonising exposure units to those used in the EU risk assessment was considered key to this mandate.

Our methodology did not exclude any data *a priori* but, rather, gave higher consideration to endpoints characterised by the highest scores of relevance, validity and precision. For this purpose, heatmaps were used as screening tool to identify – and therefore, focus on – those endpoints characterised by the highest validity and precision. Endpoints with the lowest score of internal validity were given low weight in the final assessment, but were still described, summarised and discussed in each line of evidence.

Weighing the evidence and the uncertainty by expert judgment

Upon assessment, different lines of evidence were collated in individual tables following a categorisation by study type and assessment endpoint group. These tables summarised the WOE and uncertainty analysis with a structured approach. For this purpose, the strength of each line of evidence was defined by its overall scores of validity and precision. Additionally, a new, 3-level (i.e. low; moderate and high) quality score named 'internal consistency' was introduced. The purpose of this indicator was to quantify the coherence across endpoints characterising each line of evidence. Finally, a 3-level (i.e. low; moderate and high) judgement was assigned to the potential of each line of evidence to indicate a higher hazard compared to the data considered in the previous peer review EFSA (2015a). Paired to this judgement, a threefold qualitative indicator of the uncertainty of such judgement was introduced indicating the level of certainty of the assessment. Specifically, the uncertainty – whose quantification required a certain degree of expert judgement – was defined as the link between external validity, internal validity, precision and internal consistency. Additionally, a text column was used to further justify the rationale behind the judgement.

Mechanistic studies

As reported in Annex A, a series of ad hoc criteria were developed for the data extraction and appraisal of mechanistic studies. Briefly, because of the different nature of the mechanistic data, it was decided to extract and appraise the endpoints with a descriptive approach and not to assign quantitative (validity and precision) indicators to each endpoint. Particularly, the data extraction was initially performed narratively, along with the appraisal. Nonetheless, upon later reconsideration, an additional schematic and more structured data extraction was deemed useful to collate the different lines of evidence (Annex F).

It should be noted the references including mechanistic data also included description of experiments with standard laboratory designs, which were considered directly and highly relevant to the scope of this mandate (e.g. Hayward et al., 2019; RefID 32). These endpoints underwent a full, separate assessment, using the CATs developed for bee laboratory studies.

The resulting mechanistic endpoints were collated into a single WOE and uncertainty analysis (i.e. including consideration of both acetamiprid and flupyradifurone), which – similar to the appraisal – were done in a more descriptive way than other designs. The reason behind this choice is that a considerable proportion of mechanistic data were not specifically linked to any pesticide (i.e. phylogenetic studies and expression profiling). Furthermore, other endpoints were used as read-across information (i.e. linked to substances other than acetamiprid and flupyradifurone, but still indirectly informative of their assessment). Consequently, the proportion of mechanistic data specifically linked to either acetamiprid or flupyradifurone was low. Therefore, the same evaluation of the mechanistic experiments was reported in both the statements.

3.2.6. Deviations from the protocol

CATs for hazard/exposure studies

Some modifications of the CATs were considered necessary after the evaluation process started. These were needed as the original formulations of the different RoB categories for some criteria and for specific situations not tested in the calibration exercise, resulted in contradictory interpretations between the assessors. These deviations were transparently reported in yellow-highlighted cells directly in the protocol description in Annex A.

Weight of evidence and uncertainty analysis for hazard data and mechanistic studies

The methodology for the WOE and the uncertainty analysis was not fully detailed in the protocol. Hence, the methodology outlined in Section 3.2.5 is considered a deviation from the original plan.

3.3. Assessment

3.3.1. Results of the screening step

For flupyradifurone, hazard and mechanistic data were only available for bees. Relevant exposure data were on the contrary not available.

Apart from the two references considered in the human health assessment, there were other references which reported environmental data, which, nonetheless, were not considered relevant for the present assessment.

Tang et al. (2019; RefID 9) focuses on the effects of flupyradifurone on an aphid, which is a target species. So, this study can inform the efficacy analysis of this substance, but not the environmental risk assessment for non-target species.

Traynor et al. (2016; RefID 12) measured residues from alive in-hive bees, stored pollen and wax in migratory colonies over time and compared exposure to colony health. However, no residues of flupyradifurone were reported. Thus, the study cannot inform the exposure assessment for the active substance under investigation.

O'Neill and O'Neill (2011; RefID 35) analysed the pollen load composition and size in *Megachile rotundata*. The study does not provide any direct information about exposure to any of the two substances considered in this mandate. In principle, if information on the uses of these substances were defined, pollen preferences might be qualitatively used to predict the relevance of the exposure to these two substances in conditions comparable to the ones of the study. The study was carried out in Montana (US) in an area characterised by alfalfa monoculture, which - because of the location - is not so relevant for EU. The predominant pollen types both by count and by volume were alfalfa, mustard and sweet clover. The landscape was dominated by alfalfa, so it is not surprising that this was dominant in the pollen loads. High abundance of mustard confirms attractiveness of brassica flowers. The proportion of crop/non-crop flowers in the area is not known, so it is difficult to extrapolate these findings to other contexts. However, the authors do mention that 'The relative densities [...] other flowering plants at the same site was assessed in an earlier study, in which we showed that the proportion of pollen types extracted from females correlated with the relative density of different plant species within 50 m of nest boxes (O'Neill et al., 2004)'. Hence, this study as such does not provide specific information that allows dismissing foraging on crops in general nor on crops other than alfalfa and Brassicaceae. Overall, the paper does not provide usable exposure information for the risk assessment.

Sinu and Bronstein (2018; RefID 36) reported foraging preferences of leafcutter bees regarding leaf discs used as nesting materials. This source of contact exposure, while possibly relevant, is not considered in the current risk assessment scheme. Preference for nesting materials may be completely different compared to preference for pollen and nectar foraging, which is the main route of exposure currently considered. In addition, the study reports about investigations carried out mainly in non-agricultural crop (most were in urban areas) and hence the relevance of the findings for agricultural areas is disputable.

Of the 40 references available, 19 (10 for human health, and 9 for the environment) reported data for acetamiprid (EFSA PPR Panel, 2021) and are therefore not further considered in this statement. Fifteen references were further considered for the environmental part of this statement.

3.3.2. Bees

3.3.2.1. Data from previous peer review

A summary of bee laboratory endpoints available from the previous peer review (EFSA, 2015a) is reported in Table 4. Such data include oral and contact acute toxicity assays with both the active substance and the representative formulation Flupyradifurone SL 200 G, all highlighting greater oral toxicity compared to contact. The first-tier risk assessment, carried out to SANCO (2002), relied uniquely on these toxicity data.

However, additional laboratory data were available, which were considered only qualitatively in the risk assessment, as they were not included in the SANCO (2002) scheme. These included:

- A chronic oral honey bee toxicity test with the active substance, which presented some uncertainties linked to the lack of a chronic standard. It must be noted that the relevant OECD 245 (OECD, 2017) was not yet published at the time.

- A repeated exposure test on honey bee larvae with the active substance (NOED equivalent to the highest tested dose). Also in this case, the relevant standard guideline (OECD GD 239; OECD, 2016) was not available at that time.
- An acute contact assay with *Bombus terrestris* and the representative formulation Flupyradifurone SL 200 G.

Table 4: Summary of bee laboratory endpoints from the previous peer review (EFSA, 2015a)

Species	Test item	Test type	Endpoint
<i>Apis mellifera</i>	Flupyradifurone	Acute oral	LD ₅₀ = 1.2 µg a.s./bee
	Flupyradifurone	Acute contact	LD ₅₀ = 122.8 µg a.s./bee
	Flupyradifurone SL 200 G	Acute oral	LD ₅₀ = 3.2 µg a.s./bee
	Flupyradifurone SL 200 G	Acute contact	LD ₅₀ = 15.7 µg a.s./bee
	Flupyradifurone	Chronic oral	LDD ₅₀ = 1.83 µg a.s./bee per day NOED = 0.79 µg a.s./bee per day
	Flupyradifurone	Repeated exposure larvae	NOED = 1.32 µg a.s./larva per dev. period
<i>Bombus terrestris</i>	Flupyradifurone SL 200 G	Acute contact	LD ₅₀ > 100 µg a.s./bee

a.s.: active substance; LD₅₀: lethal dose, median; LDD₅₀: lethal daily dose, median; NOED: no observed effect dose.

Additionally, higher tier studies were available.

In five semi-field (tunnel) studies carried out in Germany and Italy, which were considered reliable during the peer review, BYI 02960 formulations were applied to *Phacelia tanacetifolia* during bee flight. Different application regimes were tested. Assessment endpoints measured in these experiments included: mortality in front of the hive and in the tunnel, foraging activity, weight of the hives, number of bees on tent walls, food stores, brood production and development, presence of a healthy queen, colony strength and behavioural anomalies. In some of these experiments, exposure was confirmed by means of residue analysis of pollen and nectar.

In two field studies, carried out in Germany and France, BYI 02960 formulations were applied on oil seed rape with honey bees actively foraging on the crop (i.e. during bee flight). Different application regimes were tested (one spray and one seed treatment + spray application) Assessment endpoints measured in these experiments included: mortality in front of the hive (bee traps and linen sheets), foraging activity, weight of the hives, food stores, brood production and development (including overwintering), colony strength (including overwintering), and behavioural anomalies. Exposure was confirmed by means of residue and palynological analyses.

Lastly, in a long-term (6 weeks) feeding study colonies were forced to feed on up to 10 mg/kg diet. While several parameters were measured in this experiment, only overwintering success was considered reliable enough during the peer review.

Overall, in the higher tier studies, some deviations from the control were observed for forager mortality, flight intensity, brood development or hive weight. However, these observations were considered as indicative of 'slight, transient treatment-related effects'. However, it was concluded that the data set does not indicate any 'adverse acute or long-term effects to honey bee colonies including assessments for overwintering'.

3.3.2.2. Outline of the submitted hazard studies

Overall, 10 references were submitted reporting on pure laboratory experiments, while 1 reference reported on experiments with a field phase.

Among the laboratory experiments, many did not follow standard test guideline, and often the exposure duration is in between the standard duration for 'acute' and 'chronic' tests. In the present analysis, we considered 'acute' only those experiments where bees were exposed to one contamination event, i.e. one contaminated meal or one contact event with the substance. Everything else was considered as 'prolonged exposure' and more useful for informing chronic toxicity.

Acute exposure laboratory experiments

Six references reported on acute experiments.

Tosi and Nieh (2019; RefID 2) carried out several experiments to investigate synergistic effects of flupyradifurone and propiconazole on *Apis mellifera*. However, synergism with other active substances is not to be addressed in the context of the present mandate. Hence, only the four acute oral experiments with technical flupyradifurone were retained in the present analysis. Such experiments were carried out on spring bees (experiments 1 and 2) and summer bees (experiments 3 and 4), focussing on foragers (1 and 3) and in-hive bees (2 and 4). Assessment endpoints included survival and abnormal behaviour including motion coordination deficits, hyperactivity, apathy, curved-down abdomen or moribund (behaviour group).

Hesselbach and Scheiner (2019; RefID 3) performed two acute oral experiments with technical flupyradifurone and *Apis mellifera*. One with summer bees (experiment 1) and one with winter bees (experiment 2). Assessment endpoints included survival and locomotor activity (behaviour group).

Chakrabarti et al. (2020; RefID 7) reported on two acute contact exposure with *Apis mellifera* and the product Sivanto Prime (content of active flupyradifurone 17.09%). In both cases, bees received a single dose of pesticide via a Potter spray tower. In one experiment (experiment 1), they were monitored for 6 h, while in the other (experiment 2) they were monitored for 10 days. In both, assessment endpoints included survival, water and sugar consumption (behaviour group), oxidative stress and caspase-3 protein activity as a proxy of cell apoptosis (subindividual group).

Finally, three unpublished documents from Bayer (Bayer, 2017a,b,c; RefID 1001, 1002, 1003 respectively) reported on three acute contact tests with *Osmia bicornis*, *Osmia rufa* and *Megachile rotundata*. The three experiments were carried out with formulation Flupyradifurone SL 200 and, being close to the standard test with honey bees, focussed uniquely on survival.

Prolonged exposure laboratory experiments

Five references reported on prolonged exposure experiments.

Hesselbach and Scheiner (2019; RefID 3) apart for the aforementioned acute experiments, also performed two oral experiments where honey bees were exposed to technical flupyradifurone for 24 h. Similarly, to the acute part, one experiment was with summer bees (experiment 3) and one with winter bees (experiment 4). Assessment endpoints included survival and locomotor activity (behaviour group).

Tan et al. (2017; RefID 4) performed one experiment on adult *Apis cerana* bees (experiment 2), by exposing them via spiked food each 12 h for 3 days to technical flupyradifurone. Assessment endpoints included survival, learning and memory (behaviour group) via proboscis extension response (PER).

Al Nagggar and Baer (2019; RefID 5) performed two experiments on adult *Apis mellifera* bees. In one experiment, bees were exposed to flupyradifurone technical (experiment 3), while in another, bees were exposed to Sivanto 200SL (experiment 4). In both cases, exposure occurred via diet and lasted for 6 days. A subset of bees was later inoculated with *Nosema ceranae* in a crossed treatment design. Assessment endpoints included survival, sugar consumption (behaviour group), *Nosema* infection intensity and expression of two set of genes (subindividual alteration group) linked to detoxication (SODH2, CYP51A, CYP6A3, CYP6A4, GSTD1) and to immune response (chitinase, hymenoptecin, defensin1, apismim, Lys-1 and PGRPS2).

Hesselbach et al. (2020; RefID 8) performed three laboratory experiments exposing *Apis mellifera* adults to technical flupyradifurone. In all three experiments, exposure was via diet and lasted 10 days. One of those (experiment 1) tested survival of winter bees in January/February 2018. Another (experiment 2) tested survival of newly emerged bees in May/June 2018. Finally, the last laboratory experiment (experiment 3) assessed alterations in the brain histology (subindividual alteration group).

Tong et al. (2019; RefID 10) reported two experiments testing interaction of dietary exposure to flupyradifurone technical and nutritional stress (i.e. nectar with low sugar content). Experiment 1 focussed on summer honey bees (*A. mellifera*), while experiment 2 focussed on winter honey bees. In both cases, exposure lasted for 3 days. Assessment endpoints included survival, thermoregulation (subindividual alteration group) and several others belonging to the behaviour group, such as sugar consumption and several flying ability endpoints (i.e. flight average and maximum velocity, flight duration and flight distance), all measured in flight mills.

Larvae laboratory experiments

Two references reported on experiments where bees were exposed during the larval stage. In all these experiments, bees were exposed as larvae, but some assessment endpoints were measured in the adult phase.

Tan et al. (2017; RefID 4) exposed *Apis cerana* larvae (experiment 1), by exposing them during 6 days to technical flupyradifurone. Exposure started when larvae were 1 day old. Survival was measured at cell sealing and at emergence. In addition, once adult have emerged, learning and memory (behaviour group) were measured via proboscis extension response (PER).

Al Naggar and Baer (2019; RefID 5) performed two experiments by exposing *Apis mellifera* larvae over 3 days. In one experiment, larvae were exposed to flupyradifurone technical (experiment 1), while in another, they were exposed to Sivanto 200SL (experiment 2). Once into adulthood, a subset of bees was later inoculated with *Nosema ceranae* in a crossed treatment design. Survival was measured twice, at emergence and after the inoculation with *Nosema* spores. Other assessment endpoints measured during the adult stage are the same listed in the description of experiments 3 and 4 of the same reference reported in the previous section. These included sugar consumption (behaviour group), *Nosema* infection intensity, and expression of two set of genes (subindividual alteration group) linked to detoxication (SODH2, CYP514, CYPQ3, CYPD1, GSTD1) and to immune response (chitinase, hymenoptecin, defensin1, apismín, Lys-1 and PGRPS2).

Effect field experiments

Only Hesselbach et al. (2020; RefID 8) reported on two experiments carried out in the field, in the form of a feeding study. In both experiments, newly emerged bees were equipped with a transponder and colour-marked. The tagged bees were placed in two cages per treatment group on top the original hive frames. Bees were exposed to flupyradifurone technical via artificial feeding for 7 days. After exposure, bees were released into the hive and tracked for 40 days using RFID technology. The two experiments differed in the timing, with experiment 1 being performed in July/August and experiment 2 being performed in September/October. Assessment endpoints included the onset/end of foraging, the number of trips and trip duration. All of these belong to the behaviour group, although it can be argued that the end of foraging is also informative for longevity and hence survival.

No field studies using application of the test item on or through plants as the mean of exposure were submitted.

Mechanistic experiments

A series of studies included lethal and subindividual assessments aimed to investigate the genetic and molecular basis of the inter-species sensitivity of bees towards nicotinic acetylcholine receptor (nAChR) competitive modulators, including neonicotinoids and the butenolide insecticide flupyradifurone.

These studies included standard toxicity experiments, which were mainly used as ground base to further explore the molecular basis of bee sensitivity to neonicotinoid exposure. Because of this reason, these were identified (and are hereby referred to as) mechanistic studies.

Below is the list of these studies:

- RefID: 31 – Beadle et al. (2019)
- RefID: 32 – Hayward et al. (2019)
- RefID: 33 – Johnson et al. (2018)
- RefID: 34 – Manjon et al. (2018)
- RefID: 37 – Troczka et al. (2019)

Across the studies listed above, Johnson et al. (2018) looked at the phylogeny of cytochrome P450s in 10 bee species, to search for footprints of eusociality in phytochemical detoxification. As such, and because not specifically focusing on (nAChR) competitive modulators, this reference was deemed outside the scope of this mandate and was therefore excluded from the WOE.

Despite not necessarily focusing on flupyradifurone and acetamiprid, all other references were deemed informative of the assessment of flupyradifurone. Indeed, upon more careful evaluation, it became apparent that mechanistic studies could have been used as supportive (i.e. read across) evidence on the mode of action and metabolism of the pesticides under assessment. Additionally, they may be used as lines of evidence to aid the extrapolation of toxicity information from one species to another.

For evaluation purpose, the mechanistic experiments were allocated to one of the following categories: (i) bee survival; (ii) phylogenetic analyses (including consideration of genome assembly); (iii) pharmacokinetics (i.e. pesticide uptake upon topical exposure); (iv) receptor binding studies; (v) pesticide metabolism; (vi) gene expression profiling; (vii) survival of recombinant *Drosophila melanogaster*.

Across experiment categories, a total of 79 endpoints were extracted, which are briefly listed below:

- 1) Sixteen **survival** endpoints characterised the effects of thiacloprid and imidacloprid, alone or in combination with a P450 inhibitor, on *Apis mellifera* (n = 4), *Osmia bicornis* (n = 4), *Megachile rotundata* (n = 2) and *Bombus terrestris* (n = 6).
- 2) The **phylogeny** of P450 genes was investigated across 4 studies. As previously mentioned, one additional reference including a phylogenetic analysis was not deemed directly relevant to this assessment (Johnson et al., 2018; RefID 33).
- 3) Two **pharmacokinetic** studies investigated the speed of cuticular penetration of radiolabelled ¹⁴C-imidacloprid and ¹⁴C-thiacloprid in *Osmia bicornis*.
- 4) Ten endpoints provided information **on receptor (radioligand) binding affinity** of imidacloprid (n = 4), thiacloprid (n = 4) and flupyradifurone (n = 2) in *Osmia bicornis* (n = 2), *Megachile rotundata* (n = 3), *Apis mellifera* (n = 3) and *Bombus terrestris* (n = 2).
- 5) Seventeen **metabolism** endpoints provided information on the ability of microsomal preparation (7) or cell lines (10) expressing P450s from *Osmia bicornis* (n = 3), *Megachile rotundata* (n = 2) and *Apis mellifera* (n = 2) to metabolise thiacloprid (n = 6), imidacloprid (n = 4), flupyradifurone (n = 1), acetamiprid (n = 4), tau fluvalinate (n = 1) and nicotine (n = 1).
- 6) Seven **expression profiling** endpoints provided information on the whole-body (n = 1) or tissue-specific (n = 6) expression of P450 genes involved in the neonicotinoids detoxification in *Osmia bicornis* (n = 3), *Megachile rotundata* (n = 2) and *Apis mellifera* (n = 2).
- 7) Sixteen survival endpoints investigated if and how the functional, *in vivo* expression of key **recombinant P450 genes in *Drosophila melanogaster*** induced increased tolerance to imidacloprid (n = 7), thiacloprid (n = 8) and acetamiprid (n = 1).

3.3.2.3. Hazard characterisation and evaluation of the newly available data

Acute exposure laboratory experiments

For acute exposure laboratory experiments, the available assessment endpoints belong to survival, behaviour and subindividual alteration groups. Survival data are available for four bee species, while behaviour and subindividual endpoints were only available for *Apis mellifera*. A summary of the appraisal is presented in the form of heatmap in Figure 2. A more detailed presentation can be found in Appendix A and in Annex G.

Survival

Chakrabarti et al. (2020; RefID 7) reported about 20% mortality after 6 h and about 55% mortality after 10 days (control = 30% mortality) at the only tested exposure level. Such exposure level lowest observed effect concentration (LOEC) corresponds to a spray concentration of about 2664 mg/L of water. Both measured endpoints were considered to present moderate RoB for external validity (class 2), while they differed in the evaluation for internal validity, mainly due to the control performances. Low RoB (class 1) was assigned to mortality after 6 h and high RoB (class 3) after 10 days. Both presented low precision. The impossibility to convert the LOEC into a dose prevented the use of this data in the following steps of the analysis.

Tosi and Nieh (2019; RefID 2) reported oral LD₅₀ values between 1.9 and 6.8 µg a.s./bee. A low RoB (class 1) for external validity was concluded for these endpoints. RoB for internal validity was low (class 1) in two cases and high (class 3) in other two, mainly due to control mortality being higher than the corresponding OECD criterion. Precision was considered high. All in all, these endpoints were generally quite consistent among each other, despite the difference in the internal validity classification.

The Bayer experiments (Bayer, 2017a,b,c; RefID 1001, 1002, 1003) reported that the acute contact LD₅₀ for *Osmia bicornis*, *Osmia rufa* and *Megachile rotundata* were 28.96, 14.13 and 0.09 µg a.s./bee, respectively. All experiments were conducted in a close resemblance with the standard guidelines and the resulting assessment endpoints were considered to present low RoB (class 1) for both external and internal validity. Precision was considered high except for the experiment with *Osmia rufa*, due to a lower number of tested bees.

Behaviour

Tosi and Nieh (2019; RefID 2) reported that the frequency of honey bees exhibiting abnormal behaviours increased for pooled in-hive and foragers summer bees (merged experiment 3 and 4) in a dose-dependent manner. Significant effects compared to control were observed at all tested doses 1 h after the exposure (lowest observed effect dose (LOED) = 0.75 µg a.s./bee) and at all doses except the lowest 2 and 4 h after the exposure (NOED = 0.75 µg a.s./bee). Results for experiments 1 and 2 were not reported with a sufficient level of detail to identify a NOED or a LOED and were therefore not further considered. All these assessment endpoints were considered to have high RoB (class 3) concerning external validity, due to the nature of the assessment endpoint. However, they were considered to have low RoB (class 1) for internal validity and high precision.

Hesselbach and Scheiner (2019; RefID 3) reported abnormal behaviour linked to locomotor ability when honey bees were exposed to the higher of the two tested doses (1.2 µg a.s./bee for both experiments with summer and winter bees), but not to the lower one (NOED = 0.12 µg a.s./bee for both experiments). These assessment endpoints were assigned high RoB (class 3) for both external and internal validity, and low precision.

Chakrabarti et al. (2020; RefID 7) reported no significant alteration of sugar and water consumption 6 h after exposure. Nevertheless, 10 days after the exposure, sugar consumption significantly increased in honey bees oversprayed with a water solution concentration of about 2664 mg a.s./L. These assessment endpoints were assigned high RoB (class 3) for both external and internal validity, and low precision. As discussed for survival, the impossibility to convert the measured endpoints into a dose, prevented the use of this data in the following steps of the analysis.

Subindividual alteration

Chakrabarti et al. (2020; RefID 7) reported significant alteration of oxidative stress for honey bees 6 h after being sprayed with a water solution of about 2664 mg a.s./L. A significant increase of the activity of caspase-3 protein (proxy of cell apoptosis) was observed 10 days after the exposure, but not 6 h after the exposure. These assessment endpoints were assigned high RoB (class 3) for external validity, due to the nature of the assessment endpoint. A low RoB (class 1) was assigned to all assessment endpoint with the only exception of cell apoptosis after 6 h, as the lack of effects observed together with a lack of a positive control created uncertainty about the sensitivity of the system. A low precision was assigned to all endpoints, as only one dose was tested. As discussed for survival and behavioural endpoints, the impossibility to convert the measured endpoints into a dose, prevented the use of this data in the following steps of the analysis. Since this was the only reference reporting on subindividual alteration, this assessment endpoint group was not further considered in the context of acute exposure.

	ExtVal	IntVal	Prec
2 1-Survival	Green	Red	Green
2 2-Survival	Green	Green	Green
2 3-AbnBehaviour	Red	Red	Green
2 3-Survival	Green	Red	Green
2 4-Survival	Green	Green	Green
3 1-LocomotrActivity	Red	Red	Red
3 2-LocomotrActivity	Red	Red	Red
7 1-CellApop	Red	Green	Red
7 1-OxydativeStress	Red	Green	Red
7 1-SugarCons	Red	Red	Red
7 1-Survival	Yellow	Green	Red
7 1-WaterCons	Red	Green	Red
7 2-CellApop	Red	Green	Red
7 2-OxydativeStress	Red	Green	Red
7 2-SugarCons	Red	Green	Red
7 2-Survival	Yellow	Red	Red
7 2-WaterCons	Red	Red	Red
1001 1-Survival	Green	Green	Green
1002 1-Survival	Green	Green	Red
1003 1-Survival	Green	Green	Green

Figure 2: Summary of the appraisal done on the assessment endpoints for acute exposure laboratory experiments with bees. The outcome takes into account the risk of bias and the precision for several criteria combined with a predefined algorithm (see Annex A). Green indicates low risk of bias or high precision (class 1), yellow moderate risk of bias (class 2 for external and internal validity), while red indicates high risk of bias (class 3) or low precision (class 2)

Prolonged exposure laboratory experiments

For prolonged exposure laboratory experiments the available assessment endpoints belong to survival, behaviour and subindividual alteration groups. Also, data related to *Nosema ceranae* infection (other assessment endpoint group) were available. All data relate to honey bees, mainly *Apis mellifera*, but some also to *Apis cerana*. All experiments relate to oral exposure. A summary of the appraisal is presented in the form of heatmap in Figure 3. A more detailed presentation can be found in Appendix A and in Annex G.

Survival

Hesselbach and Scheiner (2019; RefID 3) reported no effect on survival at both tested doses in both experiments with summer and winter honey bees (*Apis mellifera*) when these were exposed for 24 h (NOED = 1.04 and 1.75 µg a.s./bee for summer and winter bees, respectively). These assessment endpoints presented moderate RoB (class 2) for external validity, due to the non-standard exposure length. They presented high RoB (class 3, mainly because of reporting drawbacks) and low precision (only two tested doses).

Tan et al. (2017; RefID 4) reported significant effect at both tested doses (LOED = 0.066 µg a.s./bee per day) after *Apis cerana* worker bees were exposed for 3 days (12 h on each day). Mortality was about 20% at the LOED and close to 50% at the higher dose (0.66 µg a.s./bee per day). This assessment endpoint presented moderate RoB (class 2) for both external and internal validity. The low number of tested doses led to a low precision classification.

Al Naggari and Baer (2019; RefID 5) reported a significant decrease of survival for *Apis mellifera* bees exposed for 6 days to both flupyradifurone technical and the formulation Sivanto 200SL. In particular, significant effects were observed in both experiments after 10 days (LOED = 0.645 µg a.s./bee). However, it should be noted that the additional mortality caused by the treatment was very low when expressed in absolute terms (i.e. about 1% and 5% for flupyradifurone technical and the formulation Sivanto 200SL, respectively). After this, a subset of bees was infected with *Nosema ceranae*. Survival was then measured again after 6 days. Treatment groups infected with *Nosema* were not considered further for survival. The other groups showed a significant decrease for bees initially exposed to formulation Sivanto 200SL. In this case, the effect size could not be precisely derived, but it is certainly < 20%. A moderate RoB (class 2) was concluded for these assessment endpoints related to the non-standard exposure length. A low RoB (class 1) was assigned to internal validity. The presence of a single dose led to a classification of low precision.

Hesselbach et al. (2020; RefID 8) reported that survival was decreased for both summer bees and winter bees after 10 exposure days at the intermediate tested dose (LOED = 0.907 µg a.s./bee per day and 0.999 µg a.s./bee per day for winter and summer bee, respectively). Although not precisely quantified, mortality across replicates was close to 20% in both experiments. On the contrary, at the lower tested doses (0.086 and 0.102 µg a.s./bee per day for winter and summer bee, respectively) no increased mortality was recorded. Finally, at the highest tested dose (4.949 and 9.108 µg a.s./bee per day) mortality was 100% well before the end of the experiments. The derived endpoints were assigned low RoB for external validity (class 1). However, a high RoB (class 3) was identified mainly due to concerns related to the performed statistical analysis. A low precision was assigned due to the low number of tested doses.

Tong et al. (2019; RefID 10) reported that exposure to flupyradifurone sustained for 3 days impaired survival of forager bees in summer (LOED = 0.241 µg a.s./bee per day, 14% mortality increased), while no impact was recorded for winter bees (NOED = 0.241 µg a.s./bee per day). A moderate RoB (class 2) for external validity was granted for these assessment endpoints. The only issue was related to the non-standard length of the exposure. A high RoB (class 3) was instead concluded due to the high control mortality. Low precision was given due to the presence of a single tested dose.

Behaviour

Hesselbach and Scheiner (2019; RefID 3) reported no significant effect on locomotor activities for summer honey bees (*Apis mellifera*) exposed for 24 h (NOED = 1.04 µg a.s./bee). For winter honey bees, the count of fallings per time walking was increased after exposure to the higher tested dose, but not to the lower (NOED = 0.12 µg a.s./bee). The biological meaning of the proposed categorisation of locomotor activities is unclear, which together with the non-standard nature of the exposure length, resulted in a high of bias (class 3) for external validity for these assessment endpoints. They presented also high RoB (class 3, mainly because of reporting drawbacks) and low precision (only two tested doses).

Tan et al. (2017; RefID 4) reported significant effect on learning and memory (PER) at both tested doses (LOED = 0.066 µg a.s./bee per day) after *Apis cerana* worker bees were exposed for 3 days (12 h on each day). These assessment endpoints presented a high risk of bias (class 3) for external validity, due to the difficulties of linking the observed effect to a colony-level impairment. A moderate risk of bias (class 2) was concluded for internal validity. The low number of tested doses and the lack of finding for a no-effect threshold led to a low precision classification.

Al Nagggar and Baer (2019; RefID 5) found no effect on sugar consumption when *Apis mellifera* bees were exposed for 6 days to both flupyradifurone technical and the formulation Sivanto 200SL (NOED = 0.645 µg a.s./bee). A high RoB (class 3) for both external and internal validity was concluded together with a low precision.

Tong et al. (2019; RefID 10) reported that a 3-day exposure to flupyradifurone did not impact average flight velocity, flight duration and flight distance in either summer or winter bees (NOED = 0.266 µg a.s./bee per day). Summer bees were also not affected for flight success and maximum velocity (NOED = 0.266 µg a.s./bee per day). Nevertheless, sugar consumption was decreased by 14% (LOED = 0.213 µg a.s./bee per day) in summer bees fed with high quality nutrition, but not in winter bees (NOED = 0.213 µg a.s./bee per day). In addition, flight success was decreased by 19% in nutritional stressed winter bees (LOED = 0.266 µg a.s./bee per day). On the contrary, exposure to flupyradifurone increased maximum flight velocity of winter bees under poor nutrition by 13% (LOED = 0.266 µg a.s./bee per day). All behavioural endpoints in this reference were considered to have high RoB (class 3) for external validity. Most were also assigned high RoB for internal validity (class 3) with the exception of sugar consumption (experiment 1 with summer bees), flight success and maximum velocity (experiment 2 with winter bees) for which a low RoB (class 1) was concluded. The main difference in the RoB classification for internal validity is related to the lack of a positive control to verify the sensitivity of the system, which was considered less of an issue for those assessment endpoints for which a significant effect of the treatment was recorded. Since only one dose was tested, all measured endpoints were considered to have low precision.

Subindividual alteration

Al Nagggar and Baer (2019; RefID 5) reported on subindividual effects of *Apis mellifera* bees when they were exposed for 6 days to both flupyradifurone technical and the formulation Sivanto 200SL. Results of the statistical analysis are not presented per test item and consider exposure to flupyradifurone (in both forms) as a single factor. The expression of a set of genes linked to detoxification (SODH2, CYP514, CYPQ3, CYPD1, GSTD1) was not significantly impacted by exposure alone (NOED = 0.645 µg a.s./bee). On the contrary, two genes (apismn and Lys-1) in the set of linked to immune response (containing also chitinase, hymenoptecin, defensin1 and PGRPS2) presented a significant increase compared to the control. For these assessment endpoints a high RoB (class 3) for both external and internal validity was concluded together with a low precision.

Hesselbach et al. (2020; RefID 8) reported no alteration of brain histology in newly emerged bees exposed for 10 days to a single dose of flupyradifurone (NOED = 0.999 µg a.s./bee per day). A high RoB (class 3) for both external and internal validity was concluded together with a low precision.

Tong et al. (2019; RefID 10) measured the thoracic surface temperatures of the bees after 3 days of exposure, before and after flight, in order to investigate the effect of the treatment on thermoregulation. This was not significantly altered in winter bees (NOED = 0.213 µg a.s./bee per day), but it was (−4%) for exposed bees fed with a nutritional rich diet (LOED = 0.213 µg a.s./bee per day). Assessment endpoints were considered to have high RoB (class 3) for external validity. A high RoB (class 3) for internal validity was concluded for the experiment on winter bees, while a low RoB (class 1) was concluded for the experiment on summer bees. The main difference is related to the lack of a positive control to verify the sensitivity of the system, which was considered less important for the experiment for which a significant effect was recorded.

Other assessment endpoints

The only additional assessment endpoint which could not be included in any of the categories listed above is the effect of exposure to flupyradifurone on the infection intensity of *Nosema ceranae*. This was measured by Al Nagggar and Baer (2019; RefID 5). Results of the statistical analysis are not presented per test item and consider exposure to flupyradifurone (in both forms) as a single factor. They reported that exposure for 6 days to the pesticide caused a significant increase in the infection intensity following inoculation (LOED = 0.645 µg a.s./bee), especially in bees from two out of four experimental colonies. For these assessment endpoints, a high RoB (class 3) for both external and internal validity was concluded together with a low precision.

	ExtVal	IntVal	Prec
3 3-LocomotrActivity	Red	Red	Red
3 3-Survival	Yellow	Red	Red
3 4-LocomotrActivity	Red	Red	Red
3 4-Survival	Yellow	Red	Red
4 2-Learning_PER	Red	Yellow	Red
4 2-Memory_PER	Red	Yellow	Red
4 2-Survival	Yellow	Yellow	Red
5 3-Detox_geneExp	Red	Red	Red
5 3-Immune_geneExp	Red	Red	Red
5 3-Nosema	Red	Red	Red
5 3-SugarCons	Red	Red	Red
5 3-Survival	Yellow	Green	Red
5 4-Nosema	Red	Red	Red
5 4-SugarCons	Red	Red	Red
5 4-Survival	Yellow	Green	Red
8 1-Survival	Green	Red	Red
8 2-Survival	Green	Red	Red
8 3-BrainHist	Red	Red	Red
10 1-AvgVelocity	Red	Red	Red
10 1-FlightDistance	Red	Red	Red
10 1-FlightDuration	Red	Red	Red
10 1-FlightSuccess	Red	Red	Red
10 1-MaxVelocity	Red	Red	Red
10 1-SugarCons	Red	Green	Red
10 1-Survival	Yellow	Red	Red
10 1-Thermoreg	Red	Green	Red
10 2-FlightDistance	Red	Red	Red
10 2-FlightDuration	Red	Red	Red
10 2-FlightSuccess	Red	Green	Red
10 2-MaxVelocity	Red	Green	Red
10 2-SugarCons	Red	Red	Red
10 2-Survival	Yellow	Red	Red
10 2-Thermoreg	Red	Red	Red

Figure 3: Summary of the appraisal done on the assessment endpoints for prolonged exposure laboratory experiments with bees. The outcome takes into account the risk of bias and the precision for several criteria combined with a predefined algorithm (see Annex A). Green indicates low risk of bias or high precision (class 1), yellow moderate risk of bias (class 2 for external and internal validity), while red indicates high risk of bias (class 3) or low precision (class 2)

Larvae laboratory experiments

For experiments where bees were exposed during the larval stage, the available assessment endpoints belong to survival, behaviour and subindividual alteration groups. Also, data related to *Nosema ceranae* infection ('other' assessment endpoint group) were available. All data relate to honey bees, either *Apis mellifera* or *Apis cerana*. A summary of the appraisal is presented in the form of heatmap in Figure 4. A more detailed presentation can be found in Appendix A and in Annex G.

Two references reported on experiments where bees were exposed during the larval stage. In all these experiments, bees were exposed as larvae, but some assessment endpoints were measured in the adult phase.

Survival

Tan et al. (2017; RefID 4) reported that *Apis cerana* larvae showed a significantly decreased survival at the emergence when they were exposed for 6 days (24 h on each day) to technical flupyradifurone at 0.33 µg/larvae per day,² but not at 0.033 µg/larvae per day. Nevertheless, the data

² It should be highlighted that the two tested daily doses reported in the original paper are 0.33 µg a.s./larva per day and 0.033 µg a.s./larva per day. However, these are a factor of 10 higher than the concentrations resulting from the procedures described in the same paper, which entail giving 2 µL of sugar solution to the larvae with concentrations of 1.65 and 16.5 mg a.s./L. For the time being, the reported doses were considered correct and the mismatch was considered due to a misreporting the concentration.

set was re-analysed by using the same statistical test performed by the author (Fisher's exact tests) and also an alternative method (Cochran-Armitage test) finding in both cases a significant difference with the control. Hence, a LOED was set at the lower dose (0.033 µg a.s./larvae per day). Survival was decreased by < 10% at the lower dose and by about 35% at the higher dose. These assessment endpoints were assigned moderate risk of external bias (class 2) due to the non-standard length of the exposure, which started when larvae were 1 day old. A low RoB (class 1) was concluded for internal validity, while a low precision was given due to the low number of tested doses.

Al Nagggar and Baer (2019; RefID 5) also reported on survival at emergence after a 3-day exposure during the larval stage. No effects were recorded when exposure was to the formulation Sivanto (experiment 2; NOED = 0.025 µg a.s./larvae per dev. period), while significant effects were seen when exposure was to technical flupyradifurone (experiment 1; LOED = 0.025 µg a.s./larvae per dev. period). Nevertheless, even in this second case, the effect size was very small, i.e. about 2% difference in survival compared to the control. Once into adulthood, a subset of bees was later inoculated with *Nosema ceranae* in a crossed treatment design. Survival was measured again after inoculation with *Nosema* spores. In the exposed (non-inoculated) bees, survival was not impacted compared to the control (NOED = 0.025 µg a.s./larvae per dev. period for both experiments with the active and the formulation). A low RoB (class 1) was concluded for both external and internal validity, while a low precision was assigned due to the presence of a single dose tested.

Behaviour

Tan et al. (2017; RefID 4) reported a significant effect on learning and memory (PER) on adult worker bees after *Apis cerana* larvae were exposed for 6 days (24 h on each day) at both doses (LOED = 0.03 µg a.s./larva per day). These assessment endpoints presented a high RoB (class 3) for external validity, due to the difficulties of linking the observed effect to a colony-level impairment. A low RoB (class 1) was concluded for internal validity. The low number of tested doses led to a low precision classification.

Al Nagggar and Baer (2019; RefID 5) found no effect on sugar consumption of *Apis mellifera* adult bees when larvae were exposed for 3 days to both flupyradifurone technical and the formulation Sivanto 200SL (NOED = 0.025 µg a.s./larvae per dev. period). A high RoB (class 3) for both external and internal validity was concluded together with a low precision.

Subindividual alteration

Al Nagggar and Baer (2019; RefID 5) reported on subindividual effects of adult *Apis mellifera* bees when they were exposed as larvae for 3 days to both flupyradifurone technical and the formulation Sivanto 200SL. Results of the statistical analysis are not presented per test item and consider exposure to flupyradifurone (in both forms) as a single factor. The expression of one gene (CYPD1) with a set linked to detoxication (containing also SODH2, CYP51A4, CYP3 and GSTD1) was significantly increased by exposure alone (LOED = 0.025 µg a.s./larvae per dev. period). On the contrary, one gene (apismn) in the set of those linked to immune response (containing also chitinase, hymenoptacin, defensin1, Lys-1 and PGRPS2) presented a significant increase compared to the control (LOED = 0.025 µg a.s./larvae per dev. period). However, in both cases, the alteration presented a different pattern in bees which were inoculated with *Nosema*. For these assessment endpoints a high RoB (class 3) for both external and internal validity was concluded together with a low precision.

Other assessment endpoints

The only additional assessment endpoint which could not be included in any of the categories listed above is the effect of exposure to flupyradifurone on the infection intensity of *Nosema ceranae*. This was measured by Al Nagggar and Baer (2019; RefID 5). Results of the statistical analysis are not presented per test item and consider exposure to flupyradifurone (in both forms) as a single factor. They reported that exposure as larvae for 3 days caused a significant increase in the infection intensity following inoculation (LOED = 0.025 µg a.s./larvae per dev. period), especially in bees from two out of three experimental colonies. For these assessment endpoints, a high RoB (class 3) for both external and internal validity was concluded together with a low precision.

	ExtVal	IntVal	Prec
4 1-Learning_PER	Red	Green	Red
4 1-Memory_PER	Red	Green	Red
4 1-Survival	Yellow	Green	Red
5 1-Detox_geneExp	Red	Red	Red
5 1-Immune_geneExp	Red	Red	Red
5 1-Nosema	Red	Red	Red
5 1-SugarCons	Red	Red	Red
5 1-Survival	Green	Green	Red
5 2-Nosema	Red	Red	Red
5 2-SugarCons	Red	Red	Red
5 2-Survival	Green	Green	Red

Figure 4: Summary of the appraisal done on the assessment endpoints for laboratory experiments with bee larvae. The outcome takes into account the risk of bias and the precision for several criteria combined with a predefined algorithm (see Annex A). Green indicates low risk of bias or high precision (class 1), yellow moderate risk of bias (class 2 for external and internal validity), while red indicates high risk of bias (class 3) or low precision (class 2)

Effect field experiments

A summary of the appraisal for effect field experiments is presented in the form of heatmap in Figure 5. A more detailed presentation can be found in Appendix A and in Annex H. Data were available uniquely from one reference (Hesselbach et al., 2020; RefID 8) and only behavioural endpoints. In the experiments described therein, bees were exposed for 7 days to a unique concentration level (1.4×10^{-5} mol/L). Based on average consumption data this was translated into a dose of 0.102 µg/bee per day.

Behaviour

In neither experiment (one carried out in summer and one in autumn), there was detection of significant effect for end of foraging (NOED = 0.102 µg/bee per day). The authors clarified that, in their experiments, end of foraging did not necessarily overlap with death, as they found many live bees at the end of the experiment whose last day of foraging had been much earlier. No significant effects were also recorded for the number of trips and trips duration for the summer experiment (experiment 1).

On the contrary, significant effects were recorded for start of foraging (both experiments), trip number and trip duration for the autumn experiment (experiment 2). For those, LOED = 0.102 µg/bee per day.

All these assessment endpoints were considered to have high RoB for external validity, as none of these assessment endpoints can be quantitatively link with the attribute to protect (colony strength). All were also assigned low RoB (class 1) for internal validity, with the only exceptions of start of foraging for the autumn experiment (class 3). All assessment endpoints were considered to present low precision (see Figure 5).

	ExtVal	IntVal	Prec
8 1-EndForag	Red	Green	Red
8 1-StartForag	Red	Green	Red
8 1-TripDuration	Red	Green	Red
8 1-TripNumber	Red	Green	Red
8 2-EndForag	Red	Green	Red
8 2-StartForag	Red	Red	Red
8 2-TripDuration	Red	Green	Red
8 2-TripNumber	Red	Green	Red

Figure 5: Summary of the appraisal done on the assessment endpoints for effect field experiments with bees. The outcome takes into account the risk of bias and the precision for several criteria combined with a predefined algorithm (see Annex A). Green indicates low risk of bias or high precision (class 1), yellow moderate risk of bias (class 2 for external and internal validity), while red indicates high risk of bias (class 3) or low precision (class 2)

3.3.2.4. Comparison of new data with previous hazard characterisation, weight of evidence and uncertainty analysis

Acute exposure laboratory experiments

Among the studies submitted by France and the Netherlands, acute survival endpoints for honey bees are only available from Tosi and Nieh (2019; RefID 2, see Table 5). These acute oral endpoints vary in the RoB classification for internal validity, and highlight different sensitivities of honey bees depending on the moment of the year and on their age/role. Nevertheless, when taken altogether, they provide a consistent picture, with an overall limited variability in the recorded LD₅₀s (1.9–6.8 µg a.s./bee, i.e. within a factor of 4). All of these LD₅₀ are consistent with the acute oral LD₅₀ for honey bees considered in the previous peer review (1.2–3.2 µg a.s./bee), being only slightly above the lower value used for the risk assessment (see Figure 6).

No data are available among the studies submitted by France and the Netherlands to confirm the apparent difference between oral and contact acute toxicity which was recorded in the honey bee data package considered during the peer review. Indeed, the dossier data presented a 100-fold difference between oral and contact endpoints for the active substance and a more limited fivefold difference for the representative formulation.

Contact LD₅₀ for two *Osmia* species (28.96 and 14.13 µg a.s./bee) were very similar to the lower contact endpoint for honey bees considered in the previous peer review (15.7 µg a.s./bee, obtained with the representative formulation).

However, the LD₅₀ for *Megachile rotundata* (0.09 µg a.s./bee) indicates that this species is considerably more sensitive than honey bees, *Osmia spp.* and *Bombus terrestris*. While the available evidence is somehow limited, the previously available risk assessment, uniquely performed for *Apis mellifera*, cannot cover for this species as well.

The observed ranking of the available contact LD₅₀ reflects a rather clear size gradient, with the heavier bee (*Bombus terrestris*) presenting the higher value and the lighter bee (*Megachile rotundata*) presenting the lower. Apart from considerations linked to the bee size, more mechanistic considerations were made on the basis of other references submitted by France and the Netherlands.

Behavioural assessment endpoints for honey bees suggests that negative impacts start to occur at doses around 0.7–1.2 µg a.s./bee, which is only slightly below the oral LD₅₀ previously used in the risk assessment. This seems logical, as sublethal effects, if present, are expected to occur at lower doses than those triggering lethal effects. However, considering the small difference between onset of sublethal effects and 50% mortality, no particular concern emerges from the present data set for honey bees (Table 5).

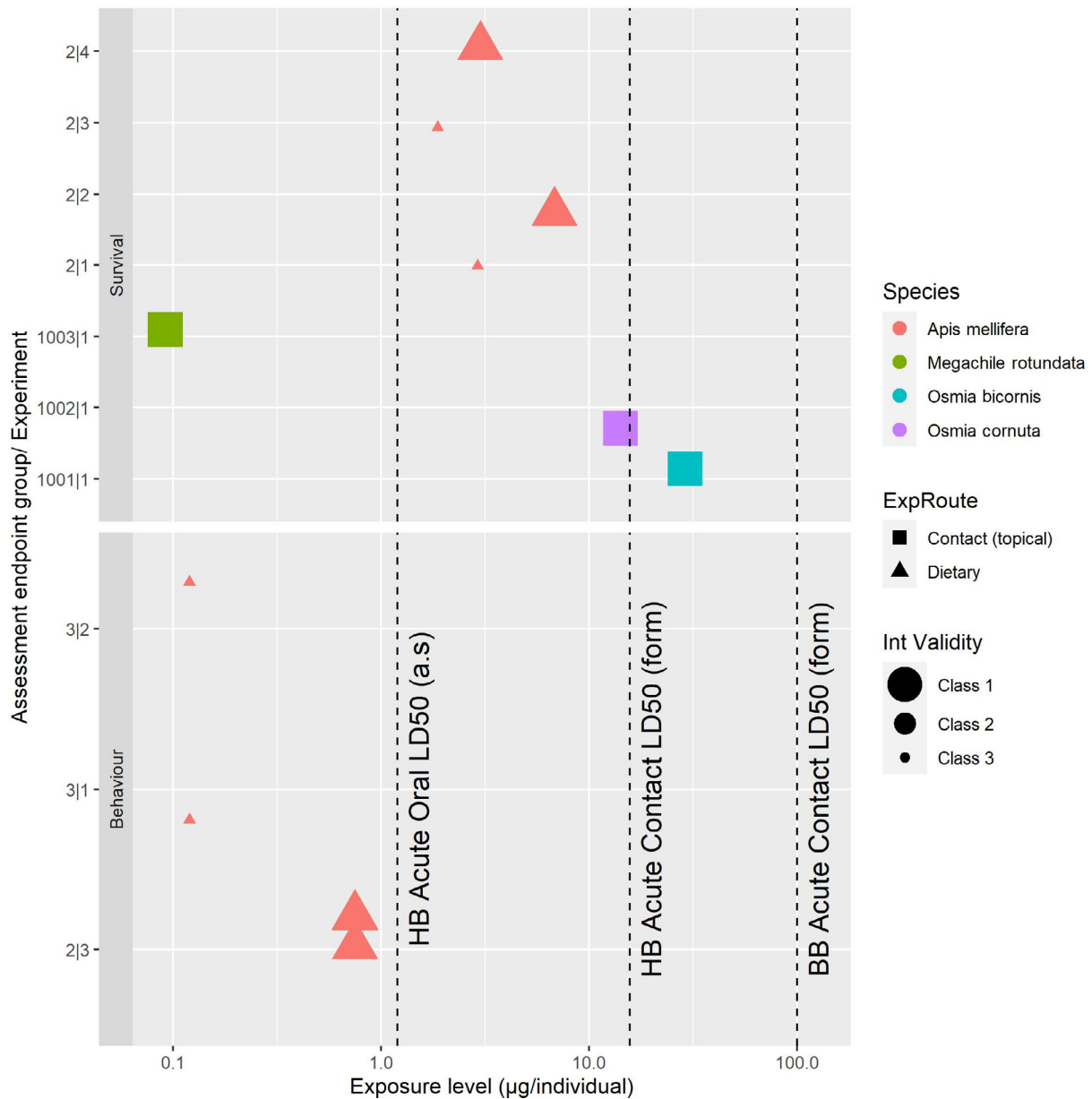


Figure 6: Summary plot of the acute bee data available for flupyradifurone. Each line on the y-axis represents an experiment within a reference (e.g. XX|Y indicate experiment Y within reference XX), organised by assessment endpoint group. Colours identify the tested species, shapes the exposure route group and size of the markers identify the internal validity class (class 1 representing low risk of bias). All measured endpoints for survival are LD₅₀, while all measured endpoints for behaviour are all NOED, with the only exception of one LOED at 0.75 µg a.s./bee for experiment 2|3. Vertical dashed lines highlight the endpoints available in the EU peer review (EFSA, 2015a)

Table 5: Weight of evidence and uncertainty analysis of the available acute exposure laboratory data for bees. EV = external validity; IV = internal validity; Prec = precision; IC = internal consistency

Assessment endpoint group	RefID exp	Strength of the line of evidence	Potential to indicate a higher hazard compared to EFSA (2015a)	
			Judgment	Rationale
Survival	<i>Apis mellifera</i> (oral exposure) 2 1 2 2 2 3 2 4	EV: low RoB IV: low to high RoB Prec: high IC: high	Low with high certainty	The data are of primary relevance for the risk assessment of honey bees. Overall, the acute oral data seem robust, consistent and in line with the previously available endpoints
	<i>Megachile rotundata</i> (contact exposure) 1003 1	EV: Low RoB IV: Variable RoB Prec: high IC: NA	High with high certainty	The data are of primary relevance for the risk assessment of solitary bees, which was not considered in the previous peer review (EFSA, 2015a). The availability of a single experiment does not allow to address the consistency of the line of evidence, but the data were considered to be reliable. <i>Megachile rotundata</i> seems to be considerably more sensitive than other bee species under acute contact exposure. Since for honey bees the oral LD ₅₀ was considerably lower than the contact one, there is additional uncertainty about whether the available endpoint is also protective for other routes of exposure
	<i>Osmia spp.</i> (contact exposure) 1001 2 1001 3	EV: low RoB IV: low RoB Prec: low to high IC: high	Low with moderate certainty	The data are of primary relevance for the risk assessment of solitary bees, which was not considered in the previous peer review (EFSA, 2015a). Overall, the acute contact data seem robust and consistent between species of the same genus. The data do not indicate an increased toxicity compared to honey bees under contact exposure. Nevertheless, since for honey bees the oral LD ₅₀ was considerably lower than the contact one, there are indications that the available endpoints for <i>Osmia spp.</i> are not protective for other routes of exposure. Overall, it is possible that the acute oral endpoint of honey bees is still protective for <i>Osmia spp.</i> , but this extrapolation presents some uncertainties
Behaviour	<i>Apis mellifera</i> (oral exposure) 2 3 + 2 4	EV: high RoB IV: low to high RoB	Low with high certainty	The data are of limited relevance for the risk assessment, due to the impossibility to link behavioural

Assessment endpoint group	RefID exp	Strength of the line of evidence	Potential to indicate a higher hazard compared to EFSA (2015a)	
			Judgment	Rationale
	3 1 3 2	Prec: low to high IC: high		alteration to effects at the colony level for honey bees. The available endpoints present a diverse level of internal validity and precision, but overall suggests that negative impacts start to occur at doses slightly below the oral LD50 previously used in the risk assessment. Considering the small difference between onset of sublethal effects and 50% mortality, no particular concern emerges from the present data set for honey bees

Prolonged exposure laboratory experiments

Among the studies submitted by France and the Netherlands, survival endpoints for honey bees in prolonged exposure experiments are available from five references (see Table 6). These endpoints vary in the RoB classification for internal validity, and most importantly in the exposure duration, from 1 to 10 days. During the data extraction phase, most were classified as NOED or LOED, either expressed as daily dose or as total dose over the entire exposure period (see Figure 7). Nevertheless, in order to assess the internal consistency of this line of evidence, it is important to 1) rescale the doses as either total or daily; 2) assess the actual level of observed mortality. This was done in Figure 8, where also the results of the only chronic test available in the dossier considered in the peer review for flupyradifurone (EFSA, 2015a) are represented.

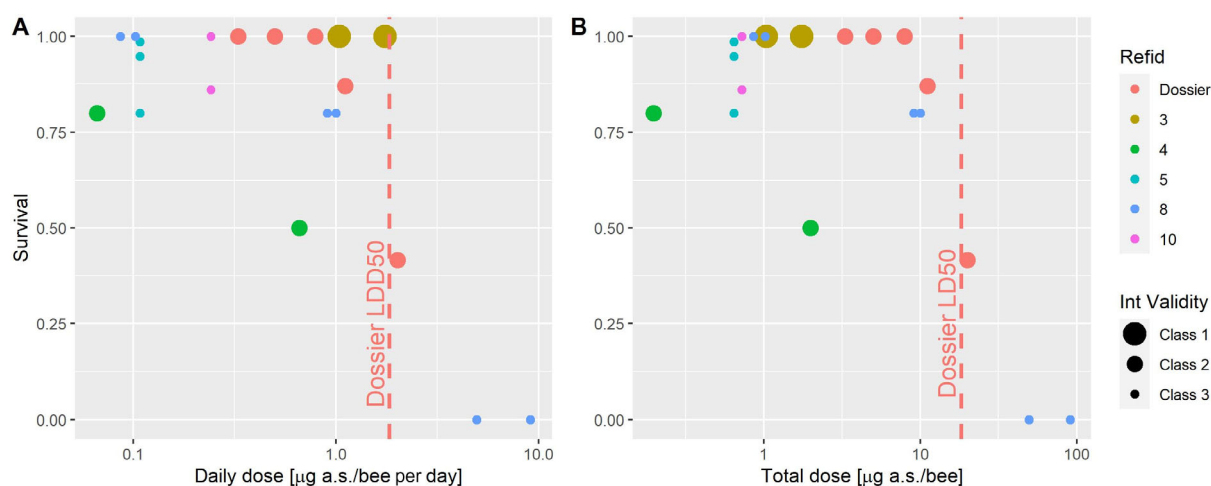


Figure 7: Survival data from the prolonged exposure experiments with honey bees available in the data package. The value from the dossier study is also reported with an arbitrary internal validity of 2, as it was considered acceptable but with some uncertainties in the previous peer review (EFSA, 2015a). (A) exposure as daily dose; (B) exposure as total dose. Vertical dashed lines indicate the 50% effect thresholds (as LDD₅₀ in (A) and as LD₅₀ in (B)) estimated in the available dossier study

All in all, the available data do not provide a very consistent picture, but the overall tendency is that large effects (> 50%) are expected at daily doses > 2 µg a.s./bee per day. EFSA (2013) makes use of LDD₅₀ as primary endpoint for the chronic risk assessment of bees, meaning that the interest is on the daily dose causing 50% effect. While a quantitative chronic risk assessment was not carried out in the peer review of flupyradifurone (EFSA, 2015a), the endpoint available from the dossier could have been used (i.e. LDD₅₀ = 1.83 µg a.s./bee per day). None of the available datapoints suggests

that a mortality higher than 50% is expected at lower daily doses (Figure 7A). The only doubt is represented by the data from Tan et al. (2017; RefID 4), which reported a mortality of 50% at a lower daily dose (0.066 µg a.s./bee per day) administered for only 3 days. This 'odd' response is even more evident when considering the total dose (Figure 7B). Nevertheless, it must be considered that Tan et al. (2017; RefID 4) reports on the only experiment carried out on *Apis cerana*, which is generally smaller than *Apis mellifera* and for which a lower measured endpoint is to be expected.

A considerable number of measured endpoints concerning behaviour alteration, subindividual alteration and *Nosema* infection (group 'other' in Figure 8, see also Table 6) were also available.

Effects on many of these assessment endpoints were observed at doses comparable to the ones that triggered some (minor) effects on survival. These endpoints are normally not used to characterise the hazard properties in the first tier of the risk assessment and, therefore, trying to establish a relationship between a certain effect level and its impact at the colony level is extremely challenging.

Perhaps the only exception in this sense is represented by *Nosema ceranae* infection intensity. Previous studies (Emsen et al., 2020) showed that such intensity has quantifiable impacts on honey bee colony health. In their study, Emsen et al. (2020) used thresholds of intensity at < 1 million spores/bee (low intensity) and > 2 million spores per bee (high intensity), although they reported infections at least as high as 16 million spores per bee. Results from Al Naggari and Baer (2019; RefID 5) showed that honey bees previously exposed to flupyradifurone and later inoculated with *Nosema* spores can present infection intensities slightly higher than 2 million spores per bee. Nevertheless, it must be noted that this only happened in bees from one out of the 4 colonies used in the experiment, where also control bees presented infection intensities above 0.5 million spores per bee.

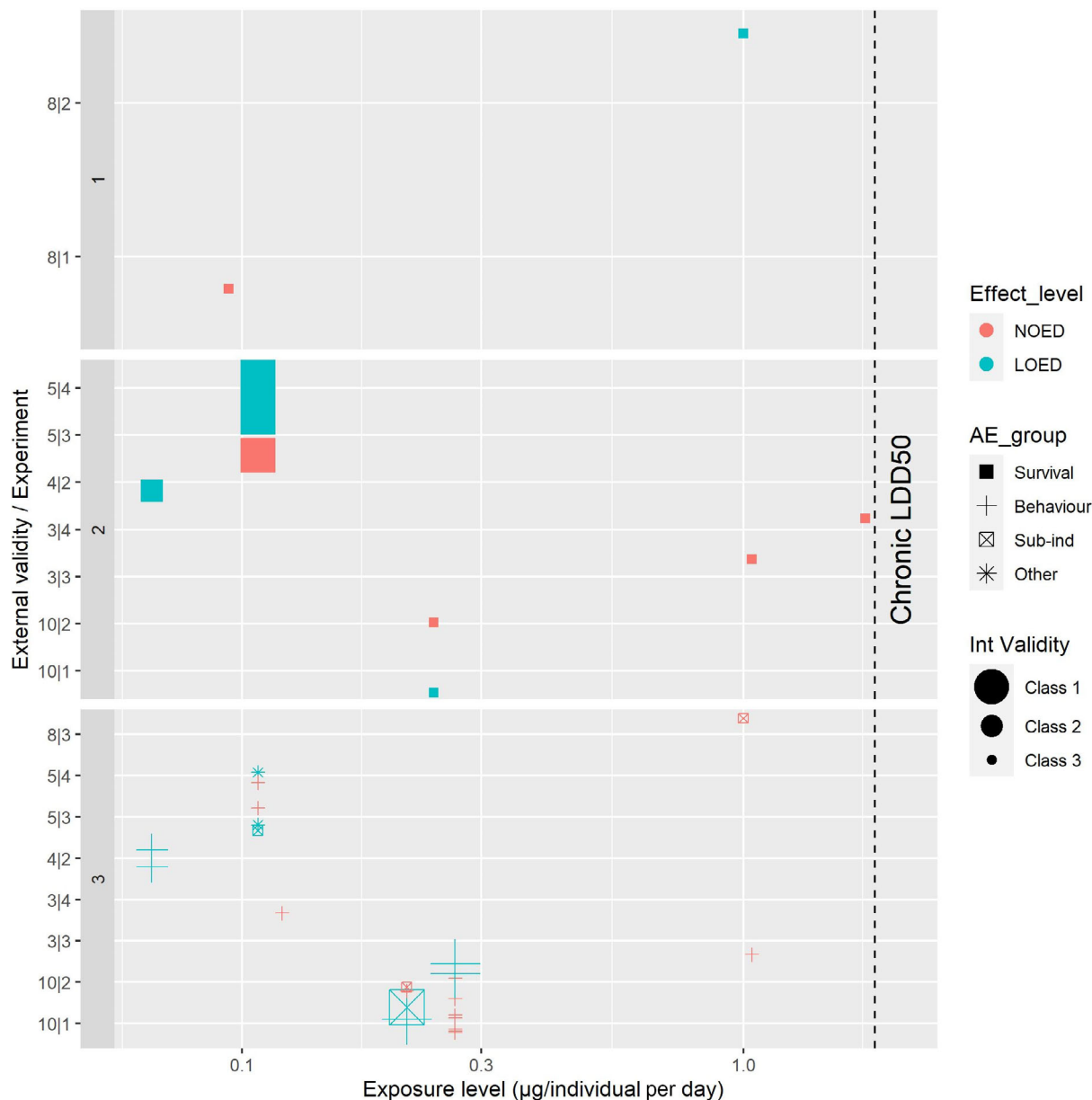


Figure 8: Summary plot of the prolonged oral exposure of honey bees to flupyradifurone. Each line on the y-axis represents an experiment within a reference (e.g. XX|Y indicate experiment Y within reference XX), organised by external validity class (class 1 representing low risk of bias). Colours identify the effect level, shapes the assessment endpoint group ('other' is used for *Nosema* infection intensity, as it could not be grouped with any of the other) and size of the markers identify the internal validity class (class 1 representing low risk of bias). Vertical dashed lines highlight the endpoint available in the EU peer review (EFSA, 2015a). Please note that such line represents an LDD₅₀, while all other points from the new data are LOED and NOED. For a more informative comparisons of effect size, please see Figure 7

Table 6: Weight of evidence and uncertainty analysis of the available prolonged exposure laboratory data for bees. EV = external validity; IV = internal validity; Prec = precision; IC = internal consistency

Assessment endpoint group	RefID exp	Strength of the line of evidence	Potential to indicate a higher hazard compared to EFSA (2015a)	
			Judgment	Rationale
Survival	3 3 3 4 4 2 5 3 5 4 8 1 8 2 10 1 10 2	EV: low to moderate RoB IV: low to high RoB Prec: low IC: moderate	Low with moderate certainty	<p>The data are relevant for the risk assessment of honey bees. Nevertheless, experiments were conducted with very different exposure length (from 1 to 10 days), therefore some risk of bias for external validity must be accounted for.</p> <p>When considering the relationship between either the daily dose or the total dose and the resulting effect level, the available data do not provide a very consistent picture. However, an overall tendency can be identified. It should be noted, that the chronic risk assessment is generally based on LDD₅₀. Hence, 50% effects is the most relevant threshold. Large effects (> 50%) are only recorded at daily doses > 2 µg a.s./bee per day (see Figure 7). None of the available datapoints suggests that a mortality higher than 50% is expected at doses below the LDD₅₀ (1.83 µg a.s./bee per day) available from the dossier considered during the peer review (EFSA, 2015a). The only doubt is represented by the data from Tan et al. (2017; RefID 4), which reported a mortality of 50% at a lower daily dose (0.066 µg a.s./bee per day) administered for only 3 days. Nevertheless, it must be considered that Tan et al. (2017; RefID 4) reports on the only experiment carried out on <i>Apis cerana</i>, which is generally smaller than <i>Apis mellifera</i> and for which thus a lower measured endpoint is to be expected.</p> <p>All experiments also present issues in terms of precision, as generally the number of tested doses was insufficient to describe a dose–response or even to identify an effect threshold. Furthermore, the internal validity was low for most experiments. All in all, the robustness of the data is questionable</p>
Behaviour	3 3 3 4 4 2 5 3 5 4 10 1 10 2	EV: high RoB IV: low to high RoB Prec: low IC: low/NA	Low with low certainty	<p>The data are of limited relevance for the risk assessment, due to the impossibility to link behavioural alteration to effects at the colony level for honey bees.</p> <p>All available endpoints present a low precision, and a diverse level of internal validity.</p> <p>Different experiments generally reported on different assessment endpoints, so that the information is rather scattered. However, two independent experiments reported on flupyradifurone experiments on sugar consumption. One found significant effects after 3 days of exposure at 0.2 µg a.s./bee per day, while another did not record any effect after 6 days of exposure at a threefold higher exposure.</p> <p>Experiments reporting on locomotion activities (including flight) found evidence of impairment mainly in winter bees, for which arguably movement is less important.</p> <p>All in all, even if effects were seen at doses substantially lower than the LDD₅₀, there is limited indication with the potential of changing the hazard</p>

Assessment endpoint group	RefID exp	Strength of the line of evidence	Potential to indicate a higher hazard compared to EFSA (2015a)	
			Judgment	Rationale
				characterisation available from the previous peer review (EFSA, 2015a). Additionally, there is considerable uncertainty in the data
Subindividual alteration	5 3 8 3 10 1 10 2	EV: high RoB IV: low to high RoB Prec: low IC: NA	Low with moderate certainty	The data are not relevant for the risk assessment, as the biological meaning of most of the monitored assessment endpoint is not fully clear even at the individual level, thus their relevance at the colony level is, for the time being, considered very low. All available endpoints present a low precision, and a diverse level of internal validity. Different experiments generally reported on different assessment endpoints, so that the information is rather scattered, and the level of internal consistency cannot be checked properly. Alteration of some subindividual assessment endpoint (gene expression, thermoregulation) was recorded, but it is considered unlikely that the recorded alterations can be used to predict effects at the colony level
Other: <i>Nosema ceranae</i> infection intensity	5 3 5 4	EV: high RoB IV: high RoB Prec: low IC: low	Moderate with low certainty	The indirect effect of exposure to pesticides in increasing the impact of pathogens and diseases is normally not considered in the standard risk assessment. In fact, this kind of interaction is generally not explicitly addressed. Hence, the risk of bias for external validity is considered high. While keeping this as a starting point, there is evidence suggesting that an increase in the infection intensity of <i>Nosema ceranae</i> can lead to effects at the colony level. The infection intensity detected in some bees exposed to flupyradifurone is slightly above a value previously considered to be of alert for the colony health. This was recorded at exposure levels substantially lower than the LDD ₅₀ . Nevertheless, it must be pointed out that the only experiments available presented a high risk of bias for internal validity and a low internal consistency, as this effect was only detected in bees proceeding from one out of the four colonies used

Larvae laboratory experiments

Among the studies submitted by France and the Netherlands, survival endpoints for honey bees exposed during their larval stage are available from two references (see Table 7). One of them measured survival during adulthood, which is not standard practice, reporting no significant effects at the tested doses. Hence, there are no indications of lethal delayed effects of flupyradifurone.

However, both available references reported the effect on survival at emergence, which is considerably more in line with what is normally done in standard risk assessment. On the basis of these experiments, exposure during the larval stage triggered effects on emergence. All assessment endpoint for survival were characterised by low RoB (class 1), but the difference in the exposure duration triggered difference in the classification for external validity. Exposure for 3 days was considered more in line with the standard practice than exposure for 6 days. The measured endpoints were expressed either as total dose during the developmental period (Al Naggar and Baer, 2019; RefID 5) or as daily dose (Tan et al., 2017; RefID 4). In order to make a meaningful comparison, and also to be in line with the standard practices, all measured endpoints were converted into total doses over the developmental period (see Figure 9).

Converted data from Tan et al. (2017; RefID 4) highlight that a total dose of 0.198 µg a.s./larva per dev. period triggered a maximum of 10% effect, while a 10-fold higher dose (1.98 µg a.s./larva per dev. period) triggered 35% effect. This suggests that the dose response is rather shallow. The dose tested in Al Nagggar and Baer (2019; RefID 5) is considerably lower (0.025 µg a.s./larva per dev. period) and triggered an increase in mortality of only about 2–3%, which, while being statistically significant, it is of negligible relevance from both a biological (EFSA Scientific Committee, 2011a) and an ecological perspective.

The NOED available from the previous peer review (1.32 µg a.s./larva per dev. period) was considered only qualitatively in the risk assessment reported in EFSA (2015a). This NOED is close to the higher tested dose in Tan et al. (2017; RefID 4), which as mentioned above, triggered a non-negligible mortality of 35%. However, it should be noted that the two experiments are not fully comparable in terms of exposure duration (6 days in Tan et al., 2017 vs 3 days in the dossier study) and in terms of species tested (*Apis cerana* vs. *Apis mellifera*). For the same reasons, Tan et al. (2017; RefID 4) may represent a worst-case compared to standard testing.

In consideration of all of this, the previously derived NOED might need to be revised on the basis of a new study carried out according to the standard relevant standard guideline (OECD GD 239) which was not yet available at the time of the previous peer review. However, it should also be noted that in the previous peer review many higher tier experiments were available addressing, among others, effects on brood. Hence, potential concerns related to survival of honey bee larvae from laboratory experiments are to some extent already addressed by the previously available higher tier experiments.

Measured endpoints concerning behaviour alteration, subindividual alteration and *Nosema* infection (group 'other' in Figure 9) were also available.

Effects on behavioural assessment endpoints (i.e. learning and memory) were observed at doses comparable to the ones that triggered effects on survival. Hence, no additional concern is raised based on these.

Effects on some subindividual assessment endpoints were recorded at considerably low doses (0.025 µg a.s./larva per dev. period). These endpoints are normally not used to characterise the hazard properties and, more in general, trying to establish a relationship between a certain effect level and its impact at the colony level is extremely challenging.

Nosema ceranae infection rates are also not generally considered in the hazard characterisation. As previously mentioned for prolonged exposure studies, Emsen et al., 2020 showed that infection intensity has quantifiable impacts on honey bee colony health. Results from Al Nagggar and Baer (2019; RefID 5) showed that honey bee exposed as larvae to flupyradifurone and later inoculated with *Nosema* spores can present infection intensities up to 2 million spores per bee, which may potentially create health issues at the colony level. Nevertheless, it must be noted that this only happened in bees from one out of the 3 colonies used in the experiment. In the other two, the infection rate was well below 1 million spores per bee, which Emsen et al. (2020) used as threshold for classifying 'low intensity'.

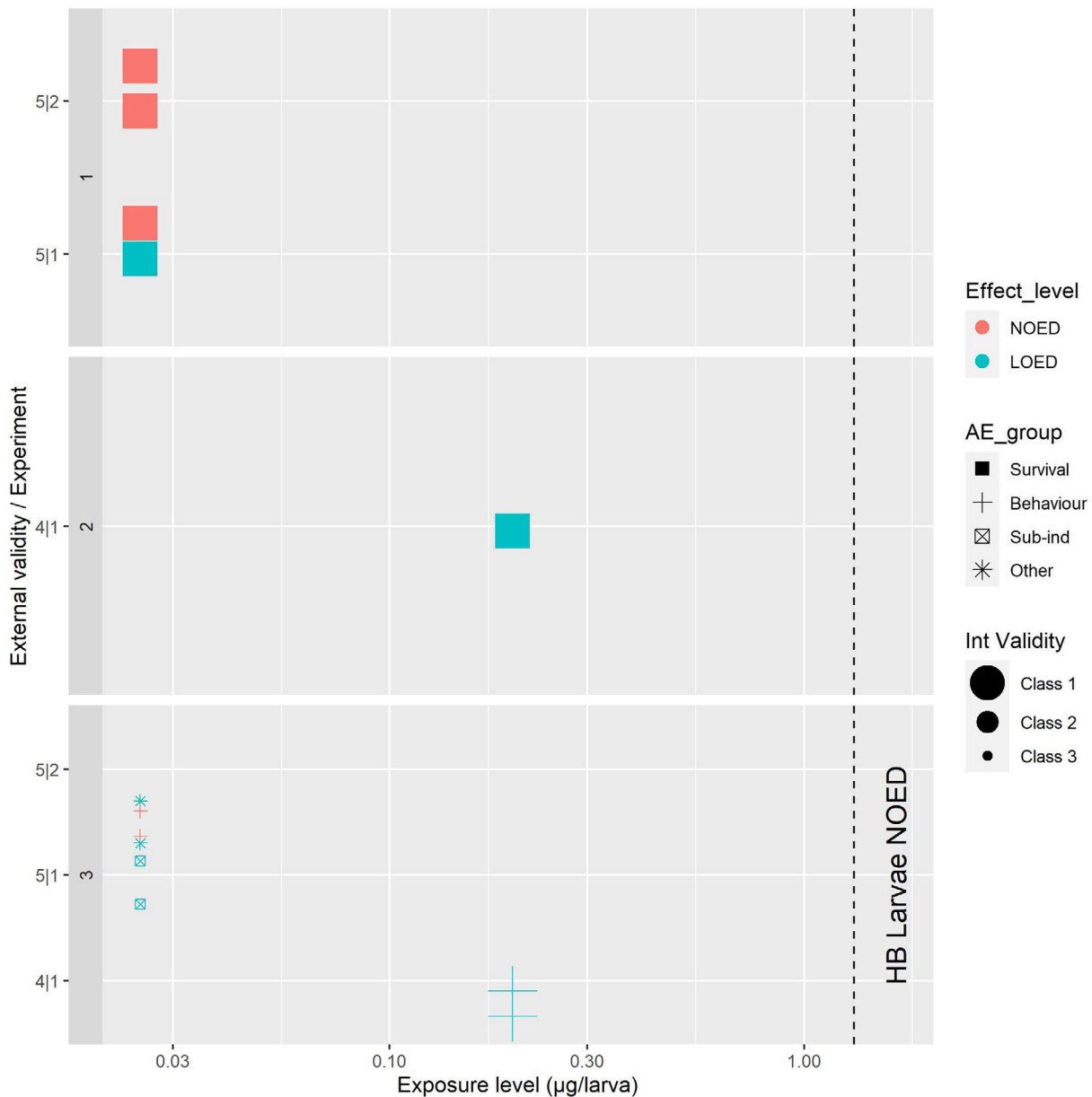


Figure 9: Summary plot of the data from experiments where honey bees were exposed to flupyradifurone as larvae. Each line on the y-axis represents an experiment within a reference (e.g. XX|Y indicate experiment Y within reference XX), organised by external validity class (class 1 representing low risk of bias). Colours identify the effect level. Different shapes identify the assessment endpoint group ('other' is used for *Nosema* infection intensity, as it could not be grouped with any of the other). Marker size identify the internal validity (class 1 representing low risk of bias). Vertical dashed lines represent the endpoints available in the EU peer review (EFSA, 2015a). Note that three of the four survival endpoints in external class 1 are NOED, meaning that no effects were observed at the only tested dose. One point is marked as a LOED, however the corresponding effect size observed was extremely small (2–3%)

Table 7: Weight of evidence and uncertainty analysis of the available laboratory data for bee larvae. EV = external validity; IV = internal validity; Prec = precision; IC = internal consistency

Assessment endpoint group	RefID exp	Strength of the line of evidence	Potential to indicate a higher hazard compared to EFSA (2015a)	
			Judgement	Rationale
Survival	4 1 5 1 5 2	EV: low to moderate RoB IV: low RoB Prec: low IC: moderate	Moderate with low certainty	<p>The data are relevant for the risk assessment of honey bees. Nevertheless, experiments were conducted with different exposure length (up to 6 days); therefore, some risk of bias for external validity must be accounted for.</p> <p>All experiments also present issues in terms of precision, as generally the number of tested doses was insufficient to describe a dose-response or even to identify an effect threshold. However, the risk of bias for internal validity was low for all assessment endpoints experiments. The number of experiments and the doses tested limit the assessment of internal consistency, but what is available does not present any issue of consistency. Hence, overall, the data are reasonably robust.</p> <p>The only tested dose in experiments 5 1 and 5 2 was considerably lower than the NOED available from the previous peer review (1.32 µg a.s./larva per dev. period). Three of the four survival endpoints from these experiments are NOED, meaning that no effects were observed at the tested dose. One point is marked as a LOED, however the corresponding effect size observed was extremely small (2–3%) and considered not to be of biological and ecological relevance.</p> <p>The NOED available from the previous peer review is close to the higher tested dose in Tan et al. (2017; RefID 4), which triggered a non-negligible mortality of 35%. However, it should be noted that the two experiments are not fully comparable in terms of exposure duration (6 days in Tan et al., 2017 vs 3 days in the dossier study) and in terms of species tested (<i>Apis cerana</i> vs. <i>Apis mellifera</i>). For the same reasons, Tan et al. (2017; RefID 4) may represent a worst-case compared to standard testing.</p> <p>In consideration of all of this, the newly available data provide some evidence that the previously available larvae endpoint might not be protective, despite this remains to be determined.</p> <p>In the previous peer review many higher tier experiments were available. Transient effects on brood were noted, but the experiments were used to achieve a conclusion of low risk. Hence, while some data indicate a potential concern, there are indications of low effects that were observed in the (semi)field. However, if an updated lower tier risk assessment would flag issues, then all the available field studies should be re-assessed according to the current state of science and in light of the new data</p>

Assessment endpoint group	RefID exp	Strength of the line of evidence	Potential to indicate a higher hazard compared to EFSA (2015a)	
			Judgement	Rationale
Behaviour	4 1 5 1 5 2	EV: high RoB IV: low to high RoB Prec: low IC: NA	Low with moderate certainty	<p>The data are of limited relevance for the risk assessment, due to the impossibility to link behavioural alteration to effects at the colony level for honey bees.</p> <p>All available endpoints present a low precision, and a diverse level of internal validity.</p> <p>Different experiments generally reported on different assessment endpoints, so that the information is rather scattered.</p> <p>The only assessment endpoint which was significantly impaired concerned learning and memory. These effects were observed at doses comparable to the ones that triggered effects on survival. Hence, no additional concern is raised based on these</p>
Subindividual alteration	5 1 5 2	EV: high RoB IV: high RoB Prec: low IC: low	Low with moderate certainty	<p>The data are not relevant for the risk assessment, as the biological meaning of the monitored assessment endpoint is not fully clear even at the individual level, thus their relevance at the colony level is, for the time being, considered very low.</p> <p>All available endpoints present a low precision, and a high risk of bias for internal validity.</p> <p>Only experiments from a single reference are available, both reporting significant alteration of some gene expression. Nevertheless, such alteration did not follow a consistent pattern for bees exposed to flupyradifurone technical and for bees exposed to a formulation. In addition, a different pattern was observed for bees inoculated with <i>Nosema</i> and bees which were not inoculated. The internal consistency does not appear high. All in all, the strength of the line of evidence is considered low, and in addition it is considered unlikely that the recorded alterations can be used to predict effects at the colony level</p>
Other: <i>Nosema ceranae</i> infection intensity	5 1 5 2	EV: high RoB IV: high RoB Prec: low IC: low	Moderate with low certainty	<p>The indirect effect of exposure to pesticides in increasing the impact of pathogens and diseases is normally not considered in the standard risk assessment. In fact, this kind of interaction is generally not explicitly addressed. Hence, the risk of bias for external validity is considered high.</p> <p>While keeping this as a starting point, there is evidence suggesting that an increased infection intensity of <i>Nosema ceranae</i> can lead to effects at the colony level.</p> <p>The infection intensity detected in some bees exposed during the larval stage to flupyradifurone is slightly above a value previously considered to be of alert for the colony health. This was recorded at exposure levels substantially lower than the NOED for larvae considered in the previous peer review (EFSA, 2015a).</p> <p>Nevertheless, it must be pointed out that the only experiments available presented a high risk of bias for internal validity and a low internal consistency, as this effect was only detected in bees proceeding from one out of three colonies used in the experiments, and only for bees exposed to flupyradifurone technical but not to formulation Sivanto</p>

Effect field experiments

Only two effect field experiments were available in the data package relative to this mandate for flupyradifurone (see Table 8). Both experiments were described in the same paper (Hesselbach et al., 2020; RefID 8). Only behavioural endpoints were described therein. Effect on some assessment endpoints related to foraging behaviour were altered by a prolonged (7 days) artificial exposure to a dose of flupyradifurone equal to 0.102 µg/bee per day. The observation of behavioural effects linked to flight activities at daily doses in the range of 0.1–0.2 µg/bee per day is consistent with the findings of the laboratory prolonged exposure experiments. On the basis of the available evidence, some behavioural effects at these exposure levels is plausible.

In general, when it comes to higher tier studies, it is expected that effects on the main attribute to protect (i.e. colony strength) are addressed as well. This was not the case for Hesselbach et al. (2020; RefID 8), and hence linking individual behavioural effects to colony-level effects is not possible with the available data.

Exposure in the other higher tier experiments available in the peer-reviewed dossier (five tunnel studies, two field studies and one feeder studies) was not quantified in terms of daily doses per bee. However, in the two field studies residues were measured in pollen and nectar collected from foragers and in the combs. Maximum residues in nectar were around 4,000 µg/kg when taken from foragers, and around 1,000 µg/kg when taken from combs. The estimated daily sugar consumption rates for foragers according to EFSA (2013) are between 32 and 128 mg/day. Assuming a sugar concentration of 50% (reasonable worst-case), the estimated intake is between 0.16 and 1.0 µg/bee per day. A further feeding study in the peer-reviewed dossier used concentrations up to 10,000 µg/kg for both pollen and nectar. In all these experiments, the intake from the bees were thus likely above the doses which, according to the results from Hesselbach et al. (2020; RefID 8), would create behavioural issues. These higher tier studies were used to conclude on the risk assessment for flupyradifurone, and the new information submitted within this mandate do not challenge their findings. On the other hand, it is highlighted that none of the higher tier studies available in the flupyradifurone dossier during the last peer review would, when taken individually, measure up to the standards recommended in EFSA (2013) (Table 8).

Table 8: Weight of evidence and uncertainty analysis of the available field effect studies with bees. EV = external validity; IV = internal validity; Prec = precision; IC = internal consistency

Assessment endpoint group	RefID exp	Strength of the line of evidence	Potential to indicate a higher hazard compared to EFSA (2015a)	
			Judgment	Rationale
Behaviour	8 1 8 2	EV: high RoB IV: low to high RoB Prec: low IC: NA	Low with moderate certainty	The data are of limited relevance for the risk assessment, due to the impossibility to link behavioural alteration to effects at the colony level for honey bees. All available endpoints present a low precision, and a diverse level of internal validity. Experiments were carried out in summer and autumn. While results from the two experiments were different (some assessment endpoints impacted in one season, but not in the other), this is not necessarily a sign of inconsistency, but it could just be an indication of variability in the bee behaviour and tolerance to external stressors. Overall, the presence of just these two experiments hampers the evaluation of the consistency of the line of evidence. The observation of behavioural effects linked to flight activities at daily doses in the range of 0.1–0.2 µg/bee per day is consistent with the findings of the laboratory prolonged exposure experiments. On the basis of the available evidence, some behavioural effects at these exposure levels are plausible.

Assessment endpoint group	RefID exp	Strength of the line of evidence	Potential to indicate a higher hazard compared to EFSA (2015a)	
			Judgment	Rationale
				<p>Exposure in other higher tier experiments available in the peer-reviewed dossier was likely higher than in the newly submitted one. Hence, if this kind of behavioural effects are linked to exposure to flupyradifurone, it is reasonable to assume that they also occurred in those cases, even if not explicitly measured.</p> <p>Those higher tier experiments were used to conclude on the risk assessment for flupyradifurone, and the new information submitted within this mandate do not challenge their findings. On the other hand, it is highlighted that, when taken individually, none of the higher tier studies available in the flupyradifurone dossier during the last peer review would measure up to the standards recommended in EFSA (2013)</p>

Mechanistic experiments

Survival

Along with the data discussed under Section 3.3.2.2, this WOE includes endpoints from three references (i.e. Bayer, 2017a,b,c, RefID 1001, 1002, 1003) characterising lethal hazards of flupyradifurone (i.e. formulated as flupyradifurone 200 g/L SL³) to *Megachile rotundata*, *Osmia cornuta* and *Osmia bicornis*. These endpoints were further compared to EFSA (2015a):

- Honey bee 72-h contact LD₅₀ = 15.7 µg/bee
- Bumble bee 48-h contact LD₅₀ > 100 µg/bee

This data set provided important information on the inter-species sensitivity of bees towards nAChR competitive modulators alone or interactively with P450 inhibitors.

- Individual substances:

These data confirm previous evidence that nAChR competitive modulators are not equally toxic to bees (Figure 10), with *N*-cyanoamidine (i.e. thiacloprid) and butenolide (i.e. flupyradifurone) compounds being less lethally toxic than *N*-nitroguanidine (i.e. imidacloprid) to *Apis mellifera*, *Bombus terrestris* and *Osmia spp.* An additional, key information is that such difference was not observed in *Megachile rotundata*, which, instead, appeared similarly and highly sensitive to imidacloprid, thiacloprid and flupyradifurone. Additionally, *Megachile rotundata* appeared more sensitive than *Apis mellifera* to flupyradifurone and thiacloprid by 2 and 3 orders of magnitudes respectively (Figure 11). This finding is of particular relevance, given that the sensitivity of *Megachile rotundata* towards flupyradifurone would not be covered by the standard assessment factor of 10 applied to honey bee endpoints, and considered protective of other bee species (EFSA, 2013). Additionally, although not directly informative of this assessment, the results observed for thiacloprid were consistent with what observed for flupyradifurone, hence, further supporting the mechanistic basis of the high sensitivity of *Megachile rotundata* towards neonicotinoids and butenolide insecticides.

³ BYI 02960 SL 200 G = BYI 02960 SL 200 g/L = Flupyradifurone SL 200 G = Sivanto.

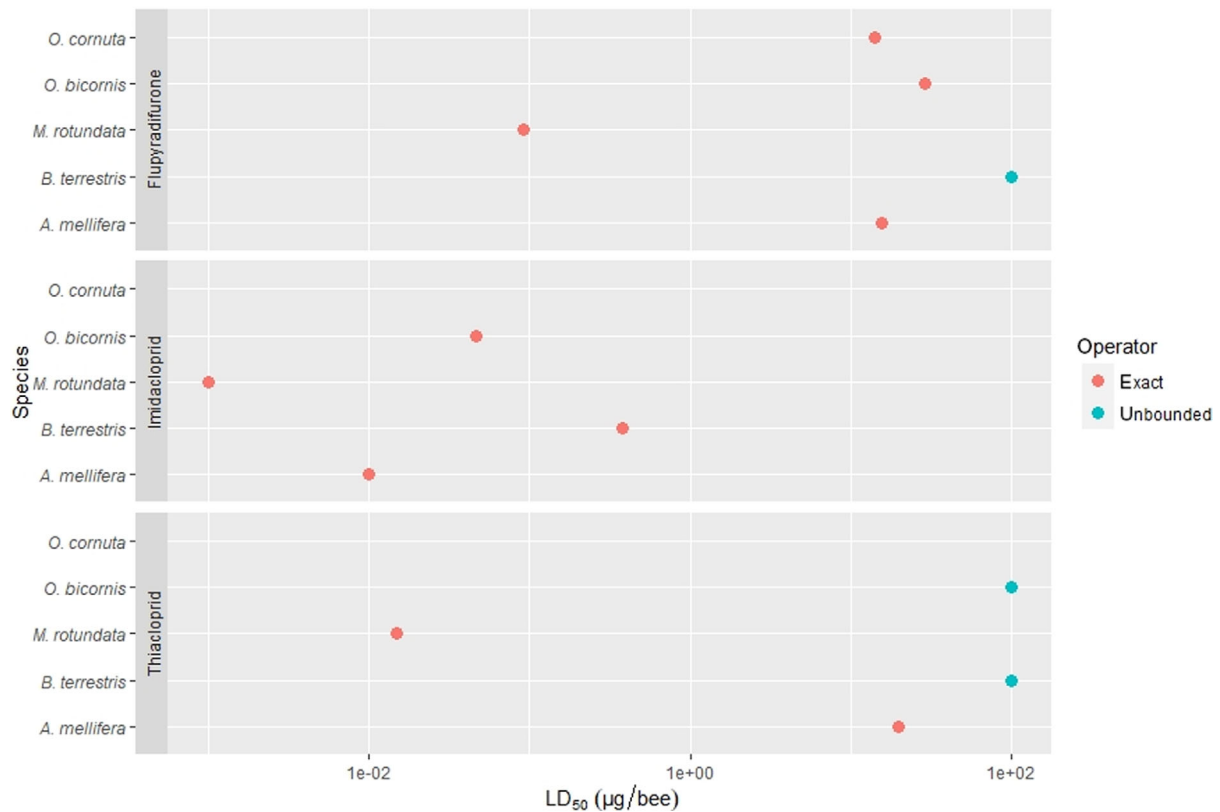


Figure 10: The acute toxicity of flupyradifurone, imidacloprid and thiacloprid (top to bottom) to *Apis mellifera*, *Bombus terrestris*, *Megachile rotundata*, *Osmia bicornis* and *Osmia cornuta*. Bee species were listed on the y axis, while the acute contact LD50 values were plotted as dots against the x axis. Unbounded (i.e. higher than) and exact values were colour coded as specified in the plot legend

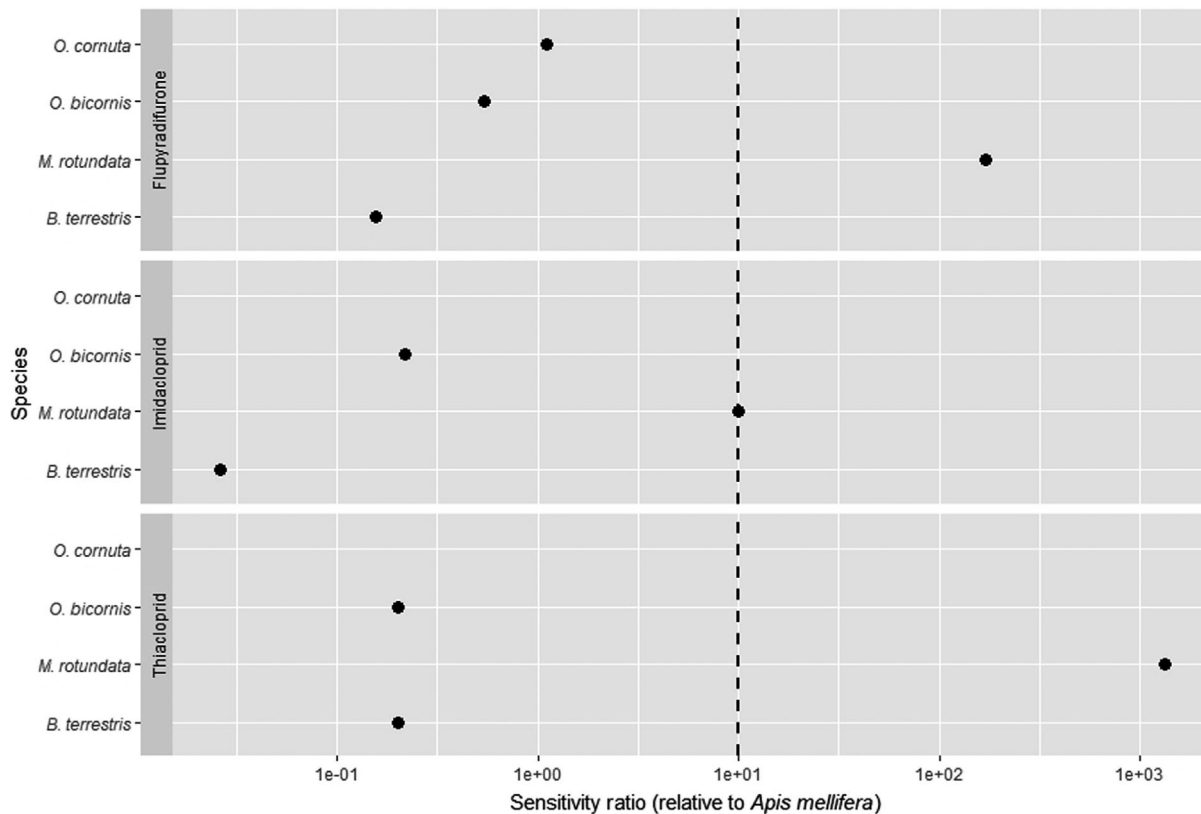


Figure 11: The sensitivity of bees to flupyradifurone, imidacloprid and thiacloprid (top to bottom). Bee species were listed on the y axis, while the sensitivity ratio (i.e. calculated as the honey bee LD_{50} divided by the LD_{50} of other bee species) was reported on the x axis (base-10 log scale). The dashed vertical line represents the sensitivity ratio = 10, used as default safety factor by EFSA (2013). Values on the right of the dashed line indicate higher sensitivity than what covered by previous assessments. The comparison is based on the bee 72 = h contact LD_{50} = 15.7 μg a.s. per bee (EFSA, 2015) from the formulation endpoint; however, the endpoint from the active substance study was higher

A number of considerations relative to assessment of survival endpoints for individual substances were made, including the following:

- i) Risk assessment schemes routinely rely on toxicity data on few bee species. Moreover, a limited proportion of species has been tested in pesticide toxicity bioassays by non-regulatory research. Consequently, the available knowledge of the interspecies sensitivity to pesticide may be considerably biased and potentially incomplete. Therefore, other bee species may show similar patterns of sensitivity to as *M. rotundata*
- ii) Similarly, with the available knowledge of pesticide metabolism by bees being limited, it cannot be excluded that *M. rotundata* might be more sensitive to other pesticides too.
- iii) In the assessment above, survival was only tested upon contact exposure. However, flupyradifurone 200 SL was more toxic via oral than contact exposure in honey bees by a factor ~ 5 .
 - Interactions with P450 inhibitors: In addition to the survival experiments above, a series of toxicity studies were produced to explore the interaction of imidacloprid and thiacloprid with bee P450 enzymes (Figure 12).

Apis mellifera became about 200 times more sensitive to thiacloprid, but only 2.7 times more sensitive to imidacloprid, upon pretreatment with a P450 inhibitor. *Bombus terrestris* became 4.16 times more sensitive to thiacloprid, and 1.19 times more sensitive to imidacloprid, upon pretreatment with a P450 inhibitor. *Osmia bicornis* became > 7.5 times more sensitive to thiacloprid, and 2 times more sensitive to imidacloprid, upon pretreatment with a P450 inhibitor. Overall, this body of evidence suggests that the tolerance of bees to thiacloprid, but not imidacloprid is downregulated by a P450 inhibitor. In other words, the higher tolerance towards thiacloprid may be linked to one or more

members of the cytochrome P450 superfamily. This does not seem to be the case of imidacloprid (at least not at the same extent), which is, indeed, more lethally toxic than thiacloprid.

However, this assessment was not specific to acetamiprid or flupyradifurone, and it is not fully clear if extrapolating such evidence across substances would be fully justified. Additionally, the differences in the sensitivity ratios for the two tested substances was not very consistent across bee species and did not clearly match the n-fold difference in the toxicity between imidacloprid and thiacloprid shown in (Figure 10). Moreover, the route of exposure was not entirely consistent across experiments.

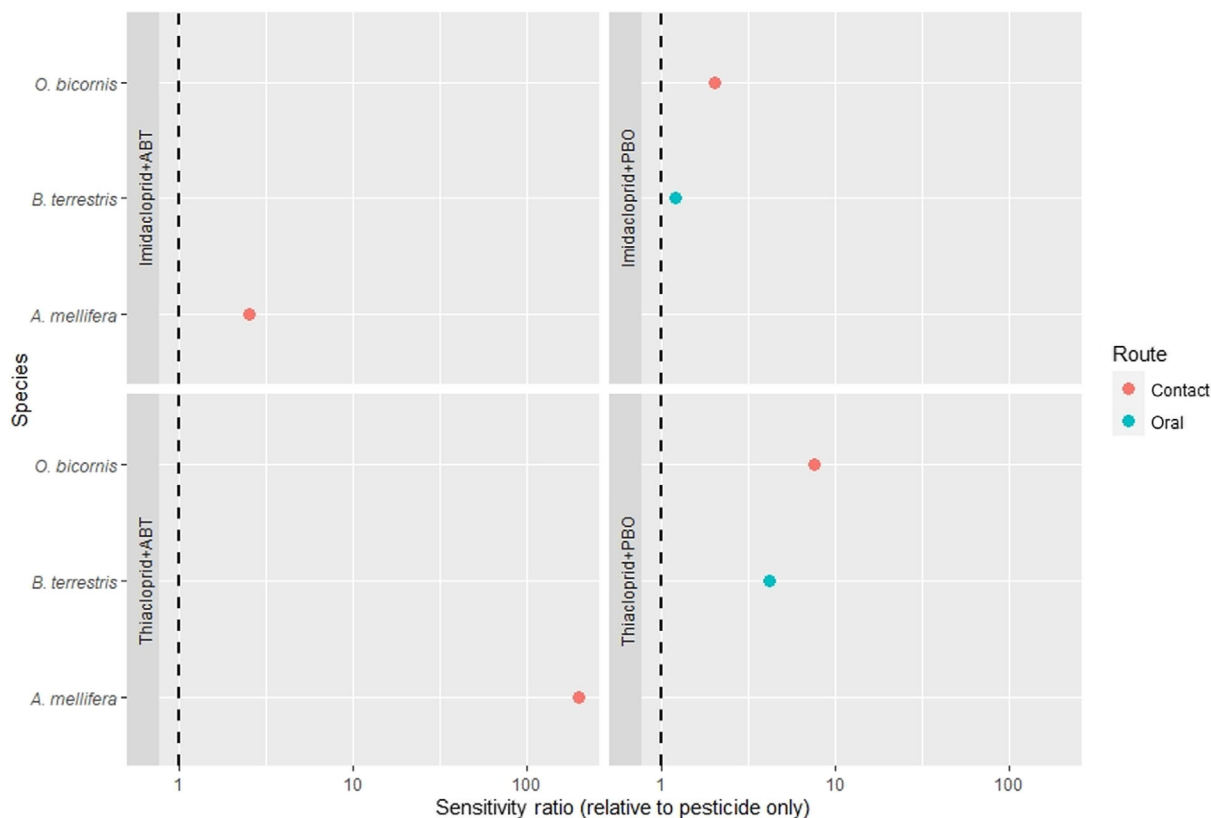


Figure 12: The interactive toxicity of imidacloprid (top) and thiacloprid (bottom) with the P450 inhibitors piperonyl butoxide (PBO – right) and 1-aminobenzotriazole (ABT – left). The bee species were listed on the y axis, while the sensitivity ratio (i.e. the toxicity ratio of the pesticide alone/pesticide + synergist) was reported on the x axis (base-10 log scale). Data points (dots) were colour-coded by route of exposure, as specified in the plot legend. The dashed vertical lines represent the sensitivity ratio = 1, indicating no interactive toxicity. Data on the right side of the dashed line indicate higher sensitivity induced by the P450 inhibitor

Phylogenetic analyses

Building on the results observed *in vivo*, further experiments including phylogeny studies were carried out, to test the hypothesis that the difference in sensitivity across bee species towards nAChR competitive modulators was driven by differences in their ability to produce cytochrome P450s, which are known to be a key route of xenobiotic detoxification in bees, as well as insects in general. Therefore, phylogenetic studies aimed to explore potential differences in the CYP9s (i.e. the genes encoding for P450s).

Therefore, phylogeny of bee P450s was explored across three species in four studies, each including a phylogenetic analysis. These data showed that:

- iv) The CYP9Q subfamily, which has a primary role in neonicotinoid detoxification, is shared by both *Apis mellifera* and *Bombus terrestris*, with the second having six genes (CYP9P1, CYP9P2, CYP9R1, CYP9Q4, CYP9Q5 and CYP6) clustering with honeybee CYP9s.
- v) The genome of *O. bicornis* lacks the CYP9Q subfamily, but, instead, has the CYP9BU subfamily
- vi) However, *M. rotundata* did not have the CYP9Q gene family or closely related genes.

Altogether these data were deemed informative, although uncertainties were identified concerning the methodological approaches (i.e. mainly related to the use of potentially fragmented or non-optimal

genome assemblies by today's standards), which are further discussed in Annex F. These methodological limitations may have influenced the outcome of the phylogenetic analysis, although it is difficult to predict how much weight these choices actually had. In principle, it cannot be excluded that the use of suboptimal, relatively fragmented genome assemblies might have had a negative impact on the detection of specific CYPs. Nonetheless, based on the information presented across mechanistic studies, it appears that differences exist in the distribution and phylogeny of CYP, with *M. rotundata* lacking a family of P450s proven to metabolise neonicotinoids in other species.

Pharmacokinetics

A possible explanation of the differences in sensitivity observed across substances and species might be the speed of cuticular penetration. Therefore, to explore the role of uptake rate on pesticide sensitivity, cuticular penetration was studied using radiolabelled ^{14}C -imidacloprid and ^{14}C -thiacloprid in *Osmia bicornis*. No difference in the absorption of the two compounds was observed, suggesting that cuticular penetration did not explain the higher sensitivity of *O. bicornis* towards imidacloprid.

Similar to other assessments, these endpoints are not specific to acetamiprid or flupyradifurone, and it is not fully clear if extrapolating such evidence across substances and bee species would be fully justified.

Receptor binding

Another factor potentially driving differences in toxicity across species and substances is the interaction at the molecular target site. Specifically, higher binding affinity at the nAChR may exacerbate toxic effects. Therefore, across the mechanistic studies, 10 endpoints provided information on receptor (radioligand) binding affinity (Figure 13) of imidacloprid ($n = 4$), thiacloprid ($n = 4$) and flupyradifurone ($n = 2$) in *Osmia bicornis* ($n = 2$), *Megachile rotundata* ($n = 3$), *Apis mellifera* ($n = 3$) and *Bombus terrestris* ($n = 2$). Results showed that imidacloprid, thiacloprid and flupyradifurone reversibly bound bee nAChRs with nanomolar affinity. The resulting half maximal inhibitory concentrations (IC_{50} s) differed across substances and tested species. However, such differences were within a 10-fold range, which suggest that receptor binding might not be a primary factor in determining either of the inter-species sensitivity or the differential toxicity of different nAChR modulators.

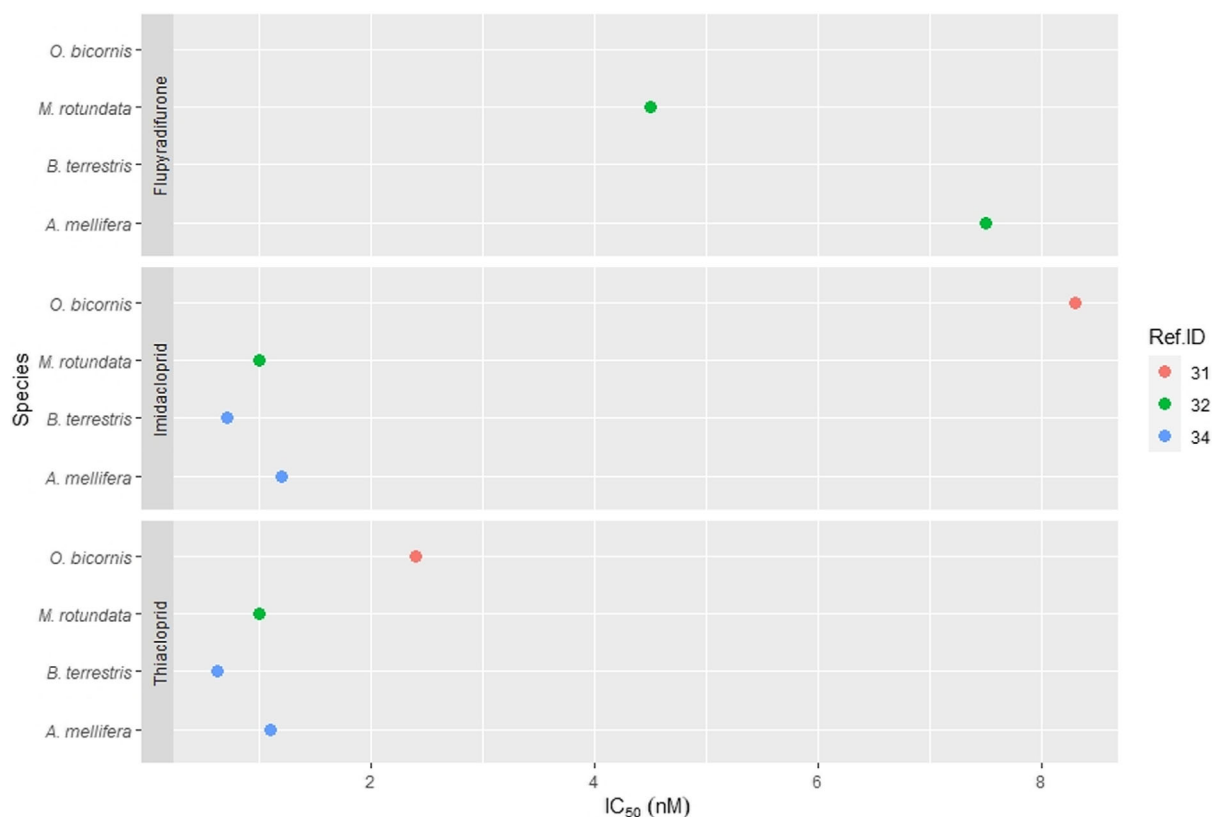


Figure 13: Radioligand binding examined by displacement of tritiated imidacloprid by unlabelled imidacloprid, thiacloprid and flupyradifurone. Dots represent the half maximal inhibitory concentration IC_{50} (nM). Lower IC_{50} values indicate higher binding affinity

Metabolism

Seventeen metabolism endpoints provided information on the ability of microsomal preparation (7) or cell lines (10) expressing P450s from *Osmia bicornis* (n = 3), *Megachile rotundata* (n = 2) and *Apis mellifera* (n = 2) to metabolise thiacloprid (n = 6), imidacloprid (n = 4), flupyradifurone (n = 1), acetamiprid (n = 4) tau fluvalinate (n = 1) and nicotine (n = 1). Altogether, this body of evidence functionally confirmed the primary role of CYP9Q (or closely related) subfamilies in the metabolism of nAChR modulators (i.e. acetamiprid, imidacloprid and thiacloprid).

A brief outline of the main findings across studies is given below,

- vii) CYP9Q1–5 (honey bee, bumble bee) significantly metabolised **acetamiprid**, with CYP9Q2 and CYP9Q3 (honey bee) resulting in the highest level of metabolisation.
- viii) Across 27 honey bee recombinant P450s, CYP9Q3 showed the highest level of **imidacloprid** and **thiacloprid** metabolisation. The overall activity against thiacloprid was higher than that of imidacloprid.
- ix) Among 5 bumble bee candidate P450s, CYP9Q4 and CYP9Q5 metabolised **thiacloprid and imidacloprid**. In general, thiacloprid was metabolised more efficiently than imidacloprid by these CYPs. However, recombinant CYP9Q6 was later shown to metabolise thiacloprid and imidacloprid more efficiently than CYP9Q4 CYP9Q5 with no clear difference across substances.
- x) *O. bicornis* CYP9BU1 and CYP9BU2 showed more efficient metabolic activity against **thiacloprid**, than **imidacloprid**.
- xi) Microsomal preparations from *M. rotundata* did not show metabolic activity for **flupyradifurone, thiacloprid, imidacloprid** or tau fluvalinate, but significantly metabolised the naturally occurring xenobiotic nicotine.

This body of evidence appears as a functional validation of phylogenetic analyses, highlighting that the difference in sensitivity between highly toxic neonicotinoids and less acutely lethal substances might be at least partially explained by different metabolisation efficiency. Indeed, across species, thiacloprid metabolism appeared more efficient than imidacloprid. However, this does not seem to be the case of *M. rotundata*, which was shown to be unable to metabolise flupyradifurone, thiacloprid and imidacloprid. This latter finding seems as a plausible functional validation of the hypothesis that the lack of CYP9Q genes drives higher sensitivity towards thiacloprid and flupyradifurone.

Although representing a robust body of evidence, the pesticide metabolism experiments were not performed comprehensively across substances and bee species. Specifically, data for acetamiprid and flupyradifurone are scarce comparatively to the other tested substances. Particularly the ability of *M. rotundata* to metabolise acetamiprid is unknown. Additionally, most studies used cell lines recombinantly expressing target genes to test the metabolic rate. It is unclear if these data can be considered fully representative of the *in vivo* metabolic response in bees.

Expression profiling

A series of studies focussing on the expression of P450 candidate genes showed that (i) candidate gene expression is not upregulated by the exposure to acetamiprid and thiacloprid; (ii) candidate genes are mainly expressed in the brain, midgut and malpighian tubules.

Survival of transgenic flies

Sixteen survival endpoints investigated if and how the functional, in-vivo expression of key P450 genes (*O. bicornis*, *A. mellifera* and *B. terrestris*) induced increased tolerance to imidacloprid (n = 7), thiacloprid (n = 8) and acetamiprid (n = 1). Overall, although not consistently across transgenes, recombinant expression of candidate bee genes induced slight to moderate tolerance to thiacloprid in *D. melanogaster*. However, this did not seem to be clearly the case of imidacloprid (Figure 14).

Flies expressing bee CYP transgenes CYP9Q2 and CYP9Q3 conferred slightly higher (less than twofold) tolerance to imidacloprid in transgenic flies.

- i) Flies expressing bee CYP transgenes (i.e. CYP9BU1 and CYP9Q2, 3, 4, 6) gained higher tolerance towards thiacloprid. However, CYP9BU2, CYP9Q1 and CYP9Q5 did not confer resistance to transgenic flies.
- ii) Data for acetamiprid were comparatively scarce, with only one transgene (i.e. CYP9Q6) conferring slight (2.3-fold) tolerance to transgenic flies (Figure 14).

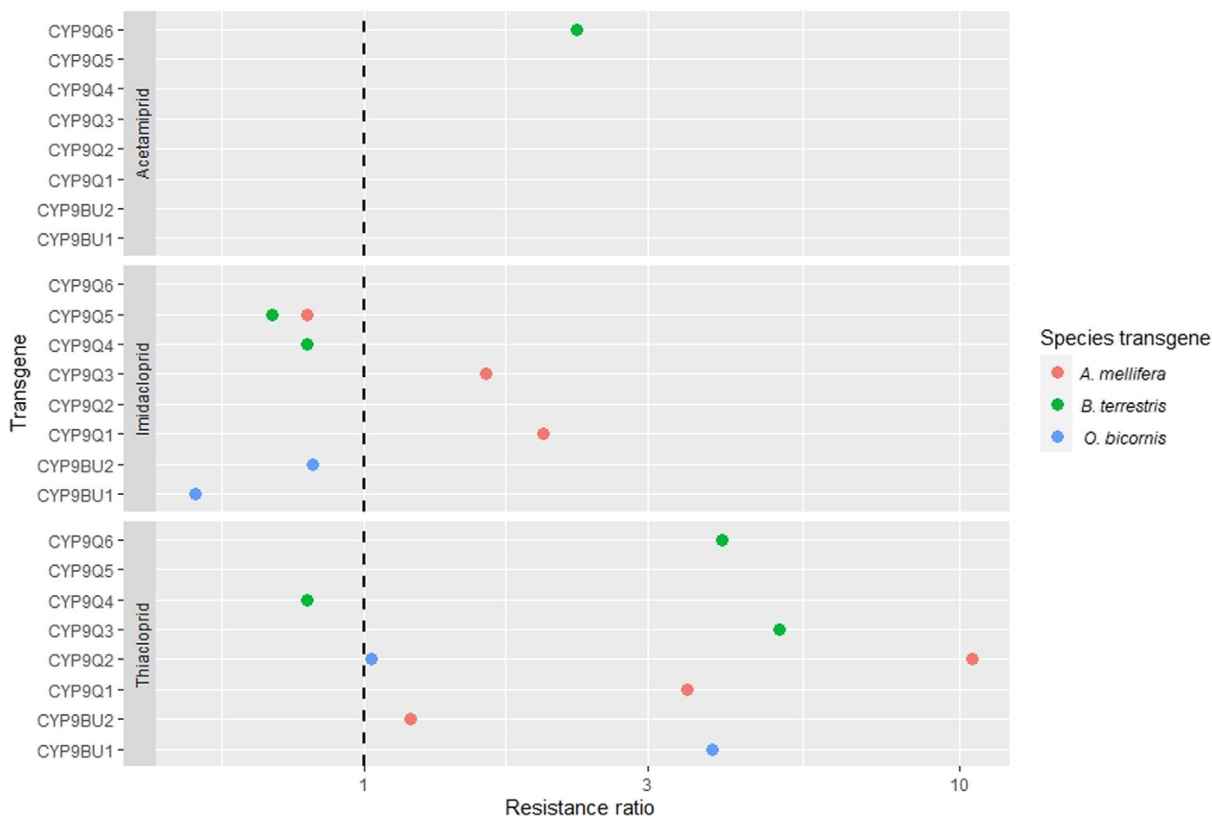


Figure 14: The resistance Ratio (RR) calculated as the ratio of the LC_{50} s of flies expressing the transgene to the LC_{50} s of flies not expressing the transgene ($x = \log$ scale). Values at the right of the dashed line indicate higher pesticide tolerance in transgenic flies

This body of evidence was produced by testing transgenic flies expressing bee P450s. Considering that the model species is a dipteran, there is uncertainty on how accurately it may represent the responses across bee species. Moreover, it is unclear whether the 'basal' response of a non-transgenic fly line represents a true control for the functional characterisation of P450s.

Similarly, it may be argued that the functional characterisation of P450s, as presented across studies, might not fully map to the *in vivo* toxicity in bees. As an example, *Megachile rotundata* was found to be more than 1,000 times sensitive to thiacloprid than *Apis mellifera*. Such difference could be partially explained by the lower bodyweight of the former species. Additionally, it was justified in Hayward et al. (2019; RefID 32) that the higher sensitivity of *Megachile rotundata* than *Apis mellifera* is related to the lack of CYP9Q genes. However, in Manjon et al. (2018 ref. ID 34), CYPQ1-3 - which were found to be the most active in neonicotinoid metabolism - only caused a 1- to 10.5-fold increase in tolerance to thiacloprid.

3.3.2.5. Conclusion for bees

Honey bees hazard

For honey bees, the new information submitted by France and the Netherlands present a low to moderate potential to indicate a higher hazard compared to what was available in the previous peer review (EFSA, 2015a). For acute exposure studies, this can be stated with fairly high certainty. For chronic exposure to adult and repeated exposure to larvae, the studies available in the peer-reviewed dossier were carried out before the publication of the relevant standard OECD guidelines. In a hypothetical new revision of the dossier data, new studies compliant with the OECD guidelines should be submitted. Nevertheless, the new data submitted by France and the Netherlands do not indicate a higher hazard for chronic exposure to adult honey bees. This could be stated with an overall moderate certainty. For honey bee larvae, there is some evidence that the previously available endpoint might not be fully protective, but such conclusion presents low certainty when considering the available evidence. As an additional remark, it is noted that very limited information in terms of quality and

quantity indicates that exposure to flupyradifurone could increase susceptibility of honey bees to *Nosema ceranae*, in some isolated cases up to levels which can impair the colony health. However, this kind of interaction with other stressor is generally not included in the risk assessment. The two available higher tier experiments did not challenge the findings of the previously available higher tier studies (five semi-field, two field and one feeder experiments).

Honey bee risk assessment

New relevant exposure information for honey bees were not submitted.

For acute exposure, as mentioned, the new information does not change the hazard characterisation done in EFSA (2015a). Hence, if the same risk assessment scheme (SANCO, 2002) is used, no changes are expected.

For chronic exposure to adult and repeated exposure to larvae, it is highlighted that the previously available endpoints were only qualitatively considered, as SANCO (2002) does not make use of those. The mandate did not request a newer assessment scheme to be used or specify which flupyradifurone uses should be assessed. Hence, no further predictions on the risk assessment can be made at this stage.

Similarly, while the newly submitted data do not challenge the findings of the previously available higher tier studies, their ability to address the risk assessment cannot be fully evaluated.

Solitary bee hazard

During the previous peer review (EFSA, 2015a), data on solitary bees were not available. On the contrary, the new information submitted by France and the Netherlands include acute contact toxicity studies with two *Osmia* species and *Megachile rotundata*.

For *Osmia* spp. the data do not indicate an increased toxicity compared to honey bees under contact exposure. Nevertheless, since for honey bees the oral LD₅₀ was considerably lower than the contact one, there are indications that the available endpoints for *Osmia* spp. are not protective for other routes of exposure. Overall, it is possible that the acute oral endpoint of honey bees is still protective for *Osmia* spp., but this extrapolation presents some uncertainties.

Megachile rotundata on the contrary is considerably more sensitive than the other tested bee species under acute contact exposure. It is noted that the difference among the LD₅₀s is higher than the standard assessment factor of 10 that EFSA (2013) suggests using to extrapolate from honey bees when solitary bee data are not available. While a lower LD₅₀ for *Megachile rotundata* is to be expected, due to its smaller size compared to honey bees and *Osmia* spp., the observed difference goes beyond that. Indeed, the available mechanistic experiments offer valid alternative explanations for the observed difference, by considering specific detoxification processes.

Other information available from the papers reporting on mechanistic experiments suggests that while such pattern is not unique for flupyradifurone, it is probably limited to specific chemical families. For example, evidence was found for *N*-cyanoamidine neonicotinoids and butenolide insecticides, but not for nitroguanidine neonicotinoids.

Since for honey bees the oral LD₅₀ was considerably lower than the contact one, there is additional uncertainty about whether the available endpoint is also protective for other routes of exposure.

Furthermore, it should be noted that experimental data are only available for a handful of species. However, the mechanistic experiments indicate that the reason behind the difference in sensitivity may be related to the different genomes, and thus may be reflected in the bee taxonomy. In Europe, there are about 80 species belonging to the *Megachile* genus, and other genera may present the same issue in the Megachilidae family, even if *Osmia* spp. does not seem to present a particular sensitivity.

Solitary bee risk assessment

Within the data package submitted, two references reported on information which are informative of the foraging behaviour of *Megachile rotundata* (O'Neill and O'Neill (2011; RefID 35), Sinu and Bronstein 2018; RefID 36). However, neither of the two contained substance-specific information and they were not considered to be fully relevant for EU agricultural areas. Furthermore, since the uses of flupyradifurone were not specified in the mandate, information on foraging preferences is of limited use. In consideration of all of this, these references were excluded at the screening level (see Section 3.3.1).

The previous peer review (EFSA, 2015a) made use of SANCO (2002). In this risk assessment scheme, solitary bees are not considered. Hence, unless a different scheme is used, no definitive consideration can be made concerning the risk assessment. It appears however very unlikely that the

present risk assessment based on either lower tier or higher tier honey bee data is protective of solitary bees as well.

3.4. Conclusions for the environment part

See conclusion for bees in Section 3.3.2.5.

3.5. Recommendation for the environment part

The current assessment was made on selected scientific evidence notified by French and the Netherlands authorities. The PPR Panel recommends that elective selection of evidence should be avoided and that a systematic evidence-based approach should be applied instead, in order to avoid bias.

For honey bees, if the SANCO (2002) risk assessment scheme is to be applied for decision-making, no further action is triggered. If, on the contrary, a more up-to-date and comprehensive risk assessment scheme can be used, as agreed in the first general expert meeting on recurring issues in ecotoxicology (EFSA, 2015b) it is recommended that:

- new laboratory experiments addressing chronic toxicity to adults and repeated exposure to larvae are carried out in accordance with the relevant OECD standards, in order to perform the first-tier risk assessment in accordance with EFSA (2013).
- The available higher tier studies are re-assessed against the principles laid out in EFSA (2013).

For solitary bees, it is recommended that an appropriate specific risk assessment for the intended uses is performed considering the available data.

References

- Al Naggar Y and Baer B, 2019. Consequences of a short time exposure to a sublethal dose of Flupyradifurone (Sivanto) pesticide early in life on survival and immunity in the honeybee (*Apis mellifera*). *Scientific Reports*, 9, 19753. <https://doi.org/10.1038/s41598-019-56224-1>
- Bayer, 2017a. Flupyradifurone SL 200: acute contact toxicity test with the solitary bee *Osmia bicornis* in the laboratory (non-GLP study).
- Bayer, 2017b. Flupyradifurone SL 200: acute contact toxicity test with the solitary bee *Osmia cornuta* in the laboratory (non-GLP study).
- Bayer, 2017c. Acute contact toxicity test with Flupyradifurone SL 200 with the solitary Leafcutter Bee (*Megachile rotundata* L.) in the laboratory (non-GLP study).
- Beadle K, Singh KS, Troczka BJ, Randall E, Zaworra M, Zimmer CT, Hayward A, Reid R, Kor L, Kohler M, Buer B, Nelson DR, Williamson MS, Davies TGE, Field LM, Nauen R and Bass C, 2019. Genomic insights into neonicotinoid sensitivity in the solitary bee *Osmia bicornis*. *PLoS Genetics*, 15, e1007903. <https://doi.org/10.1371/journal.pgen.1007903>
- Chakrabarti P, Carlson EA, Lucas HM, Melathopoulos AP and Sagili RR, 2020. Field rates of Sivanto (flupyradifurone) and Transform(R) (sulfoxaflor) increase oxidative stress and induce apoptosis in honey bees (*Apis mellifera* L.). *PLoS One*, 15, e0233033. <https://doi.org/10.1371/journal.pone.0233033>
- EFSA (European Food Safety Authority), 2013. EFSA Guidance Document on the risk assessment of plant protection products on bees (*Apis mellifera*, *Bombus spp.* and solitary bees). *EFSA Journal* 2013;11(7):3295, 268 pp. <https://doi.org/10.2903/j.efsa.2013.3295>
- EFSA (European Food Safety Authority), 2014. Guidance on Expert knowledge elicitation in food and feed safety risk assessment. *EFSA Journal* 2014;12(6):3734, 278 pp. <https://doi.org/10.2903/j.efsa.2014.3734>
- EFSA (European Food Safety Authority), 2015a. Conclusion on the peer review of the pesticide risk assessment of the active substance flupyradifurone. *EFSA Journal* 2015;13(2):4020, 106 pp. <https://doi.org/10.2903/j.efsa.2015.4020>
- EFSA (European Food Safety Authority), 2015b. Technical report on the outcome of the pesticides peer review meeting on general recurring issues in ecotoxicology. *EFSA Supporting Publication* 2015:EN-924, 62 pp. <https://doi.org/10.2903/sp.efsa.2015.EN-924>
- EFSA (European Food Safety Authority), 2016. Conclusion on the peer review of the pesticide risk assessment of the active substance acetamiprid. *EFSA Journal* 2016;14(11):4610. <https://doi.org/10.2903/j.efsa.2016.4610>
- EFSA (European Food Safety Authority), Ippolito A, del Aguila M, Aiassa E, Guajardo IM, Neri FM, Alvarez F, Mosbach-Schulz O and Szentes C, 2020. Review of the evidence on bee background mortality. *EFSA Supporting Publications* 2020:EN-1880, 76 pp. <https://doi.org/10.2903/sp.efsa.2020.EN-1880>
- EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), Hernández-Jerez AF, Adriaanse P, Berny P, Coja T, Duquesne S, Focks A, Marinovich M, Millet M, Pelkonen OR, Pieper S, Tiktak A, Topping CJ, Widenfalk A, Wilks M, Wolterink G, Rundlöf M, Ippolito A, Linguadoca A, Martino L, Panzarea M, Terron A and Aldrich AP, 2021. Statement on the active substance acetamiprid. *EFSA Journal* 2021;19(12):7030, 60 pp. <https://doi.org/10.2903/j.efsa.2021.7030>

- EFSA Scientific Committee, 2011a. Statistical significance and biological relevance. *EFSA Journal* 2011;9(9):2372, 17 pp. <https://doi.org/10.2903/j.efsa.2011.2372>
- EFSA Scientific Committee, 2011b. Scientific Opinion on genotoxicity testing strategies applicable to food and feed safety assessment. *EFSA Journal* 2011;9(9):2379, 69 pp. <https://doi.org/10.2903/j.efsa.2011.2379>
- EFSA Scientific Committee, Hardy A, Benford D, Halldorsson T, Jeger M, Knutsen HK, More S, Naegeli H, Noteborn H, Ockleford C, Ricci A, Rychen G, Silano V, Solecki R, Turck D, Younes M, Aquilina G, Crebelli R, Gürtler R, Hirsch-Ernst KI, Mosesso P, Nielsen E, van Benthem J, Carfi M, Georgiadis N, Maurici D, Parra Morte J and Schlatter J, 2017. Scientific Opinion on the clarification of some aspects related to genotoxicity assessment. *EFSA Journal* 2017;15(12):5113, 25 pp. <https://doi.org/10.2903/j.efsa.2017.5113>
- EFSA Scientific Committee, More SJ, Bampidis V, Bragard C, Halldorsson TI, Hernández-Jerez AF, Hougaard Bennekou S, Koutsoumanis K, Lambré C, Machera K, Naegeli H, Nielsen SS, Schlatter J, Schrenk D, Turck D, Younes M, Aquilina G, Bignami M, Bolognesi C, Crebelli R, Gürtler R, Marcon F, Nielsen E, Vleminckx C, Carfi M, Martino C, Maurici D, Parra Morte J, Rossi A and Benford D, 2021. Scientific Opinion on the guidance on aneugenicity assessment. *EFSA Journal* 2021;19(8):6770, 27 pp. <https://doi.org/10.2903/j.efsa.2021.6770>
- Emsen B, De la Mora A, Lacey B, Eccles L, Kelly PG, Medina-Flores CA, Petukhova T, Morfin N and Guzman-Novoa E, 2020. Seasonality of *Nosema ceranae* infections and their relationship with honey bee populations, food stores, and survivorship in a North American region. *Veterinary Sciences*, 7, 131. <https://doi.org/10.3390/vetsci7030131>
- Hayward A, Beadle K, Singh KS, Exeler N, Zaworra M, Almanza MT, Nikolakis A, Garside C, Glaubitz J, Bass C and Nauen R, 2019. The leafcutter bee, *Megachile rotundata*, is more sensitive to N-cyanoamidine neonicotinoid and butenolide insecticides than other managed bees. *Nature Ecology & Evolution*, 3, 1521–1524. <https://doi.org/10.1038/s41559-019-1011-2>
- Hesselbach H and Scheiner R, 2019. The novel pesticide flupyradifurone (Sivanto) affects honeybee motor abilities. *Ecotoxicology*, 28, 354–366. <https://doi.org/10.1007/s10646-019-02028-y>
- Hesselbach H, Seeger J, Schilcher F, Ankenbrand M, Scheiner R and Beggs J, 2020. Chronic exposure to the pesticide flupyradifurone can lead to premature onset of foraging in honeybees *Apis mellifera*. *Journal of Applied Ecology*, 57, 609–618. <https://doi.org/10.1111/1365-2664.13555>
- Jeschke P, Nauen R, Gutbrod O, Beck ME, Matthiesen S, Haas M and Velten R, 2015. Flupyradifurone (Sivanto) and its novel butenolide pharmacophore: structural considerations. *Pesticide Biochemistry and Physiology*, 121, 31–38. <https://doi.org/10.1016/j.pestbp.2014.10.011>
- Johnson RM, Harpur BA, Dogantzis KA, Zayed A and Berenbaum MR, 2018. Genomic footprint of evolution of eusociality in bees: floral food use and CYPome “blooms”. *Insectes Sociaux*, 65, 445–454. <https://doi.org/10.1007/s00040-018-0631-x>
- Manjon C, Troczka BJ, Zaworra M, Beadle K, Randall E, Hertlein G, Singh KS, Zimmer CT, Homem RA, Lueke B, Reid R, Kor L, Kohler M, Benting J, Williamson MS, Davies TGE, Field LM, Bass C and Nauen R, 2018. Unravelling the molecular determinants of bee sensitivity to neonicotinoid insecticides. *Current Biology*, 28, 1137–1143.e1135. <https://doi.org/10.1016/j.cub.2018.02.045>
- NTP, 2015. OHAT risk of bias rating tool for human and animal studies. Office of Health Assessment and Translation, National Toxicology Program, 2015.
- NTP, 2016. Monograph on immunotoxicity associated with exposure to perfluorooctanoic acid (PFOA) or perfluorooctane sulfonate (PFOS). National Toxicological Program, Research Triangle Park.
- NTP, 2019. Office of Health Assessment and Translation (OHAT) -NTP (2019): handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration.
- O'Neill RP and O'Neill KM, 2011. Pollen load composition and size in the leafcutting bee *Megachile rotundata* (Hymenoptera: Megachilidae). *Apidologie*, 42, 223–233. <https://doi.org/10.1051/apido/2010057>
- O'Neill KM, Blodgett S and Fultz JE, 2004. Variation in *Megachile rotundata* pollen load composition in relation to flower diversity (Hymenoptera: Megachilidae). *Journal of the Kansas Entomological Society*, 77, 619–625.
- OECD, 2016. Guidance Document on honey bee (*Apis mellifera*) larval toxicity test, repeated exposure. Series on testing and assessment No. 239 series on testing and assessment No. 239. ENV/JM/MONO(2016)34.
- OECD, 2017. Test No. 245: Honey Bee (*Apis Mellifera* L.), chronic oral toxicity test (10-day feeding), OECD guidelines for the testing of chemicals, section 2. OECD Publishing, Paris. <https://doi.org/10.1787/9789264284081-en>
- R Core Team, 2021. R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. Available online: <https://www.R-project.org/>
- SANCO, 2002. SANCO/10329/2002 rev 2 final. Guidance document on terrestrial ecotoxicology under council directive 91/414/EEC.
- Sekeroglu ZA, Aydin A, Yedier SK and Sekeroglu V, 2018. Cytogenetic alterations induced by flupyradifurone, a new butenolide insecticide, in human lymphocytes. *Toxicology and Industrial Health*, 34, 737–743. <https://doi.org/10.1177/0748233718788989>
- Sinu PA and Bronstein JL., 2018. Foraging preferences of leafcutter bees in three contrasting geographical zones. *Diversity and Distributions*, 24, 621–628. <https://doi.org/10.1111/ddi.12709>

- Tan K, Wang C, Dong S, Li X and Nieh JC, 2017. The pesticide flupyradifurone impairs olfactory learning in Asian honey bees (*Apis cerana*) exposed as larvae or as adults. *Scientific Reports*, 7, 17772. <https://doi.org/10.1038/s41598-017-18060-z>
- Tang Q, Ma K, Chi H, Hou Y and Gao X, 2019. Transgenerational hormetic effects of sublethal dose of flupyradifurone on the green peach aphid, *Myzus persicae* (Sulzer) (Hemiptera: Aphididae). *PLoS One*, 14, e0208058. <https://doi.org/10.1371/journal.pone.0208058>
- Tong L, Nieh JC and Tosi S, 2019. Combined nutritional stress and a new systemic pesticide (flupyradifurone, Sivanto(R)) reduce bee survival, food consumption, flight success, and thermoregulation. *Chemosphere*, 237, 124408. <https://doi.org/10.1016/j.chemosphere.2019.124408>
- Tosi S and Nieh JC, 2019. Lethal and sublethal synergistic effects of a new systemic pesticide, flupyradifurone (Sivanto((R))), on honeybees. *Proceedings. Biological Sciences*, 286, 20190433. <https://doi.org/10.1098/rspb.2019.0433>
- Traynor KS, Pettis JS, Tapy DR, Mullin CA, Frazier JL, Frazier M and vanEngelsdorp D, 2016. In-hive pesticide exposome: assessing risks to migratory honey bees from in-hive pesticide contamination in the Eastern United States. *Scientific Reports*, 6, 33207. <https://doi.org/10.1038/srep33207>
- Trocza BJ, Homem RA, Reid R, Beadle K, Kohler M, Zaworra M, Field LM, Williamson MS, Nauen R, Bass C and Davies TGE, 2019. Identification and functional characterisation of a novel N-cyanoamidine neonicotinoid metabolising cytochrome P450, CYP9Q6, from the buff-tailed bumblebee *Bombus terrestris*. *Insect Biochemistry and Molecular Biology*, 111, 103171. <https://doi.org/10.1016/j.ibmb.2019.05.006>
- U.S. EPA, 2003. Generic Ecological Assessment Endpoints (GEAEs) for Ecological Risk Assessment. Washington, DC, USA: Risk Assessment Forum EPA/630/P-02/004B
- Wickham H, 2016. *ggplot2: elegant graphics for data analysis*. Springer-Verlag, New York.

Abbreviations

AIR	Annex I Renewal
CAT	Critical Appraisal Tool
DH RoB	risk of bias definitely high
DL RoB	risk of bias definitely low
EKE	expert knowledge elicitation
LD ₅₀	lethal dose, median
LDD ₅₀	lethal daily dose, median
LOEC	lowest observed effect concentration
LOED	lowest observed effect dose
nAChRs	nicotinic acetylcholine receptors
NOED	no observed effect dose
OHAT/NTP	The Office of Health Assessment and Translation/National Toxicology Programme
PFAS	perfluoroalkyl substances
PFOS	perfluorooctane sulfonate, perfluorooctane sulfonic acid
PH RoB	risk of bias probably high
PL RoB	risk of bias probably low
RAR	Renewal Assessment Report
RoB	risk of bias
US EPA	United States Environmental Protection Agency

Appendix A – Detailed results of the appraisal phase for hazard experiments (environment)

The following figures (Figure A.1, Figure A.2, Figure A.3) are a graphical representation of the appraisal exercise performed on the literature studies considered eligible according to the criteria listed in the protocol (see Annex A).

Results are presented for each assessment question (Table A.1, Table A.2, Table A.3). For every figure, the strings on the left identify the *RefID|ExperimentID|Endpoint*, or in other words, the identifiers for: the reference, the experiment within the reference (where multiple experimental units were identified in the same reference) and the assessment endpoint investigated.

The acronyms for the single criteria are explained by means of tables for each assessment question. The colours used to fill each cell of the matrix represent the risk of bias or precision, in accordance with the following legend.

	Definitely low risk of bias/high precision
	Probably low risk of bias
	Probably high risk of bias
	Definitely high risk of bias/Low precision
	Criterion not applicable

A.1. Bees (laboratory experiments)

Table A.1: Outline of the appraisal questions for bee laboratory experiments

Section	Acronym	Question
External validity	Q1_EV	How confident are we that the assessment endpoint can be used to inform the risk assessment of bees?
	Q2_EV	Are the test organisms exposed to either flupyradifurone or acetamiprid in isolation (without any other active substances)?
	Q3_EV	Are the tested organisms relevant for Europe?
	Q4_EV	Is the duration of the exposure and observation in line with the standard testing?
Internal validity	Q1_IV	Is the origin of the tested organism trustable?
	Q2_IV	Is the age and sex of the tested organisms known and appropriate?
	Q3_IV	Were the test organisms properly acclimatised to the study setup before the exposure started? For acute studies this is generally not a problem
	Q4_IV	Are the test organisms healthy and stress-free at the start of the experiment?
	Q5_IV	Is the methodology used (including the experimental setup) for measuring the assessment endpoint reliable?
	Q6_IV	Is the negative (blank) control performing adequately?
	Q7_IV	Is the system sensitive enough?
	Q8_IV	If a solvent is used, is the effect of the solvent appropriately accounted for?
	Q9_IV	Are the test conditions appropriate?
	Q10_IV	Is the test item clearly identified and characterised?
	Q11_IV	Is exposure characterised by analytical measurements (residues or confirmed dose)?
	Q12_IV	Is exposure underpinned by appropriate measurements/estimation of test item consumption? (only relevant for oral exposure)
	Q13_IV	Was a clear dose-response observed in the study?
	Q14_IV	Is the derivation of the measured endpoint(s) performed with sound statistical methods?

Section	Acronym	Question
Precision	Q1_PR	Are the sample size and replication appropriate?
	Q2_PR	Is the number of tested concentrations/doses appropriate?
	Q3_PR	Is doses selection (including the space between them) appropriate?

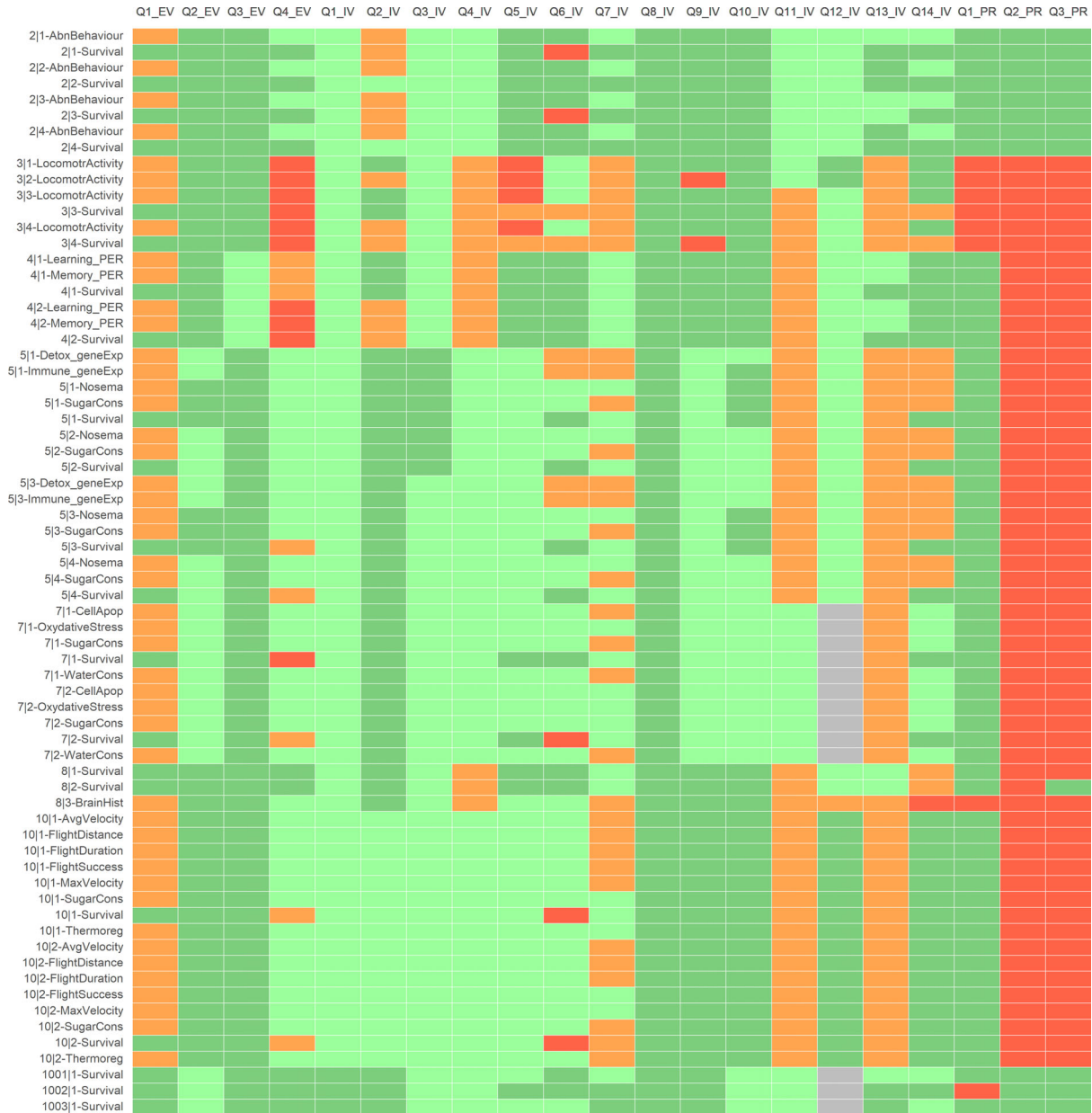


Figure A.1: The heatmap summarising the outcome of the appraisal of bee laboratory experiments

A.2. Bees (field effect experiments)

Table A.2: Outline of the appraisal questions for bee field effect studies

Section	Acronym	Question
External validity	Q1_EV	How confident are we that the assessment endpoint can be used to inform the risk assessment of bees?
	Q2_EV	Are the test organisms exposed to either flupyradifurone or acetamiprid in isolation (without any other active substances)?
	Q3_EV	Are the tested organisms relevant for Europe?
	Q4_EV	Is the study location representative of any EU biogeographical region?
	Q5_EV	Is the study setting representative of an EU agricultural landscape?
	Q6_EV	Do the experimental conditions represent a reasonable worst-case for both exposure and possible triggering of the effects?
Internal validity	Q1_IV	Is the origin of the tested organism trustable?
	Q2_IV	Were the test organisms properly acclimatised to the study setup before the exposure started?
	Q3_IV	Are the test organisms healthy and stress-free at the start of the experiment?
	Q4_IV	Is the methodology used (including the experimental setup) for measuring the assessment endpoint reliable?
	Q5_IV	Is the negative control free from contamination and performing adequately?
	Q6_IV	Are the treatments and exposure levels well characterised?
	Q7_IV	Is the test item clearly identified and characterised?
	Q8_IV	Is the duration of the test appropriate to characterise the assessment endpoint?
	Q9_IV	Is the presence of other stressors checked and accounted for?
	Q10_IV	Is the derivation of the measured endpoint(s) performed with sound statistical methods?
Precision	Q1_PR	Are the sample size and replication appropriate?

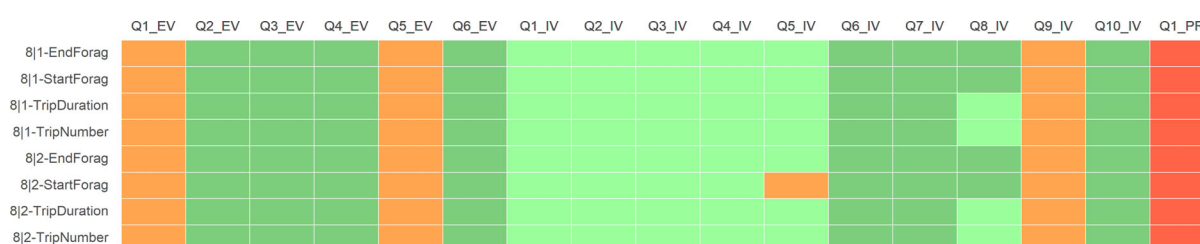


Figure A.2: The heatmap summarising the outcome of the appraisal of bee field effect experiments

Annex A – Protocol

Annex B – Outcome RoB flupyradifurone for Human Health

Annex C – RoB HeatMap flupyradifurone for Human Health

Annex D – Data extraction flupyradifurone for Human Health

Annex E – Uncertainty analysis table flupyradifurone

Annex F – Data extraction, weight of evidence and uncertainty analysis for mechanistic studies

Annex G – Detailed results of the appraisal for laboratory experiments with bees

Annex H – Detailed results of the appraisal for field experiments with bees