Peer Review File

Manuscript Title: Agrochemicals interact synergistically to increase bee mortality.

Editorial Notes: *none*

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1 (Remarks to the Author):

A. The manuscript reports the results of a meta-analysis of 385 combined effects of multiple stressors on bees. The main results are: i) synergistic effects occurred when bees were exposed to multiple agrochemicals at field realistic levels; ii) antagonistic interactions occurred when bees were exposed to multiple pathogens, or when nutritionally stressed bees were inoculated with pathogens; iii) almost all interactive effects on proxies of fitness, behaviour, parasite load and immune response were additive.
B. This work the first meta-analysis on the subject. It is a huge undertaking, particularly in the selection process, so I congratulate the authors for their great work.
C. The statistical analysis follows the methodology of similar work on multiple stressors carried out on aquatic ecosystems. The dataset is robust, the analysis was properly conducted and the results are reliable and valid.
D. My only concern is the large hetereogeneity observed between the datasets, which the authors acknowledge but do not comment in detail. As I mention below, this is not of great concern because meta-analysis deal usually with a mixed bag of experimental data, often showing contrary results, and therefore some heterogeneity is to be expected. However, this shouldn't be an excuse for accepting all findings without reservations, so the authors could be more cautious in interpreting the results of their analysis.
E. The conclusions are well spelled out in the abstract, but not so much in the main text. The paper ends with some recommendations but does not stress the main point, i.e. that current regulation of agrochemicals fails to protect bees and the biodiversity of the ecosystem. Further references to back up this main conclusion are suggested – see below.
F. Specific comments:
• Line 41: the initial statement is incorrect; Foley et al. 2005 estimated that roughly 40% of the surface of the planet is dedicated to 'cropland and pastures', but intensive agriculture is only a fraction of that figure, most likely 10-12% (Lambin & Meyfroidt 2011, PNAS 108: 3465; and FAOSTAT Online statistical service).
• Lines 95, 154 and 162: heterogeneity is very high in all cases, although this is partly an inherent feature of meta-analysis done on a diverse array of experimental designs, multiple agrochemicals and pathogens. I noted the authors comment on this point later in the manuscript (lines 180-5), but they should indicate more caution in accepting these results.
• Lines 163-4: the statement "…previous research that has suggested that agrochemical exposure can suppress the immune response of bees and leave them more vulnerable to other stressors 3,17,28,…" is inaccurate. These three citations refer specifically to the effect of neonicotinoids on the bees immune system based on the mechanism of immune suppression described by Di Prisco et al. (ref 17), which has been confirmed by other authors – i.e., see Brandt et al. 2016, Journal of Insect Physiology 86, 40-47, and other work by this lab. However, the paper by Di Prisco and colleagues also showed that the insecticide chlorpyrifos did not produce inhibition of the immune system in bees, and to my knowledge such an inhibitory effect has not been found with any other agrochemical. Therefore, it appears that only neonicotinoids are responsible for the immune

suppression effect on bees and other animals, not other agrochemicals, and this should be made clear in this sentence.
• Lines 162-6: the statement "In contrast… we found an antagonistic effect on bee immune response when exposed to multiple agrochemicals (Immune response, agrochemical*agrochemical, d = -0.27, CI = -0.52 to -0.01),…" indicates a contradiction between the immune suppression effect of neonicotinoids and the findings of this meta-analysis, which is not true and therefore misleading. Bearing in mind the small d value and the huge heterogeneity of the dataset, your finding suggests that mixtures of two or more agrochemicals do not have additive or synergistic effects on the bees immune response, which is entirely different from neonicotinoids inhibiting the immune system and allowing viral infections to become deadly, i.e. synergistic effect of agrochemical and pathogen on mortality. Please rephrase the sentence to describe accurately your findings and not to mislead the readers.
• Lines 166-7: agree with the explanation given; it is likely that exposure to one agrochemical elicits the same or similar immune response as exposure to two or more agrochemicals, so the combined effects on the immune system is less than additive and hence the 'apparent' antagonism – there is no antagonism sensu stricto in this case.
• Line 177: the reference cited (30) is not appropriate for the changes in phenology of plants, which may not be due to climate change alone but also other factors. Suggestions for replacement: Ovaskainen et al. 2013, PNAS 110: 13434-13439, and Ulrich et al. 2020, Frontiers in Plant Science 11: 542125.
• Line 190: agree, and authors should cite here some of the recent publications that advocate a more realistic evaluation of agrochemicals, as the current regulatory system is not protective of the environment, i.e. Topping et al. 2020, Overhaul environmental risk assessment for pesticides. Science 367: 360.
• Line 201: however, the appropriate dose for treating a crop is dictated by the effectiveness of the chemical formulation, not based on ecological considerations. It is easy for researchers to propose such recommendations, but agrochemical manufacturers and users of the products will never abide by them.
• Line 210: please refer here Brühl & Zaller 2019, Biodiversity decline as a consequence of an inappropriate environmental risk assessment of pesticides. Frontiers in Environmental Science 7: 177.
• Line 434: shouldn't you add Table S2 in addition to Figure S2?
• Line 455: replace Figure S3 with Table S3
• Line 436: replace Figure S4 with Table S4
G. Referencing is appropriate in most cases; I pointed above one replacement and a couple of additional references that should be added for completeness.
H. The manuscript is well written. The abstract is clear, concise and to the point. The results are well described and explained with reference to previous work. The figures are of good quality and appropriate.

Referee #2 (Remarks to the Author):

Siviter et al
Overall impression of the MS is that it is very clearly written and presented and the statistical analysis is sound. The findings are significant especially the different levels of synergy (or lack of) among stressor combinations or classes. This will interest not only pollinator biologists - where the interplay between drivers or pressures affecting pollinators is recognised as a potential threat but also a major knowledge gap (see IPBES thematic assessment of pollinators and pollination 2016) - but also to ecologists and epidemiologists studying the interactive effects of multiple stressors on ecological or pathogen systems generally. I enjoyed reading the MS and I can see how it will attract a lot of citation. I only have minor points of clarification.

Minor specific comments:

Summary – make clear that you are working on social and solitary bees somewhere, L27 – just workers or individuals in the case of non-social bees ?

L41 Conventional intensive agriculture – to distinguish from other forms that may be intensive in management but perhaps less harmful to nature e.g. agroecological, organic

L48 maybe a good place to insert "social and solitary" before "bees"

L70 maybe add citation 12 here also ?

L73 is the description of a crossed design in () really needed ?

L75 use of « category » here is slightly confusing with the subsequent use of categories (mortality etc) in text and figures. Suggest for the incidence here where you are refering to different stressors you use « type » or « class » - just a point of consistency/clarity.

Fitness proxies – overall looking at the supp. table I am convinced by most metrics, but I need some justification for the use of lipid ratios and protein content as proxies of fitness ? To be honest I dont expect their omission makes much of a difference to the results (few records), but it just needs to be stated how they correlate with and serve as a proxy of individual (colony) fitness. Is protein/fat content an indicator of fecundity (which is itself a fitness proxy)?

L95 Extended data displays – check the order of appearance and how they are cited, EDF4 appears here first (cf L109), in some cases they are cited as EDF/T (L128) in others written in full – consistency.

L107-117 I wonder if EDF 2 should actually be included as a panel in Fig1 given the importance of it to the main text?

L121 use of "and" and "&"

L142 Osmia synergistic – is it really? EDF1 does not appear to show that, additive at best?

L164 is Alaux et al 2010 worthy of citation here?
https://royalsocietypublishing.org/doi/10.1098/rsbl.2009.0986

L177 what about Baude et al 2016 cited here alongside 29?
https://www.nature.com/articles/nature16532

L183-184 not sure how to read this statement from EDF4? How does it relate to interactions BETWEEN the stressor classes (parasites x agrochemical x nutrition?). What does each row in a figure panel correspond to? Also y axis in EDF4 lacks a label and here X_axis is Observed outcome whereas elsewhere it is Effect size

L210 cite Potts et al here at end of sentence?

Other cites that may be relevant or missing ?:

Collison, E., Hird, H., Cresswell, J. & Tyler, C. (2016) Interactive effects of pesticide exposure and pathogen infection on bee health - a critical analysis. Biological Reviews, 91, 1006-1019.

Retschnig, G., Williams, G.R., Odemer, R., Boltin, J., Di Poto, C., Mehmann, M.M., Retschnig, P., Winiger, P., Rosenkranz, P. & Neumann, P. (2015) Effects, but no interactions, of ubiquitous pesticide and parasite stressors on honey bee (Apis mellifera) lifespan and behaviour in a colony environment. Environmental Microbiology, 17, 4322-4331.

I think an additional online extended data table giving the full citation of the 296 papers examined or the final 100 papers that provided data for this metaanalysis is advisable, so that the reader can identify if there were any notable omissions.

I wonder if the authors would consider a sentence calling for study of other multi-stressor interactions in pollinators or pollination (e.g. including effects of climate variability/extremes or other effects on populations of habitat loss). A major challenge lies in looking beyond these parasite-nutrition-chemical interactions to others happening in real landscapes. Of course that is a major knowledge gap and data are simply not there yet for quantitative analysis, but identifying such a need could be useful.

Referee #3 (Remarks to the Author):

In this paper the authors meta-analyse the evidence for synergic, additive, and antagonistic effects of anthropogenic stressors on bees. I find this paper well written, the literature search comprehensive, and the topic and findings interesting and important from the conservation and agricultural management perspective. I have mostly concentrated on the analysis part of this work, and provide only some minor comments on other aspects.

Meta-analysis

Overall I think that the authors did a good job in collecting relevant studies, reporting on search and extraction procedures, and conducting the analysis itself. All the main components of a meta-analysis have been applied (e.g. calculating heterogeneity, publication bias test). My comments are:

1) search terms
It would be easier for a reader to follow the reasoning behind the search terms if there was a sentence that explains what different parts of the search string stand for. E.g. the first part of the search string describes the target taxa, the second different stressors, and the third different responses.

Second, there are several instances where unnecessary terms appear. For example, if there is 'fitness' there is no need for 'colony fitness'; or if there is 'output' there is no need for 'reproductive output' too. From my experience, the more detailed terms (like reproductive output) are only used when the more general term ('output') provides too many results, and then the reviewer tries to eliminate the irrelevant terms by being more specific in defining the exact type of the 'output' (e.g. 'reproductive output').

Third, why is there no term 'lethal' (while there are different forms of 'sublethal')

Finally, it would be easier to follow the search terms if those relating to a same category were closer together (e.g. 'foraging' and 'navigat*)

2) Screening process and data extraction
What do you consider to be 'consistent' in line 340. From my experience it is rarely 100% agreement.

One article was not possible to translate, out of how many in a non-English language?

In 49 cases authors provided needed information, out of how many?

3) The decision on the selection of the ES when more variables in the same category were reported in the same study was to randomly select one value. First, how many of these cases were there?

Second, the results (especially if the number of these cases was relative high) would be more robust if this process was iterated a number of times (so, random ES selected, MA conducted; new random ES selected, MA repeated). This is easy to do with the loops in R (and one can only check the frequency distribution of the overall effect sizes obtained in this way)

4) While the effect sizes used in the analysis are provided, I would also suggest providing R codes. This is first, because of computational reproducibility, but also as in this way others can benefit from having access to it (I have myself benefited from modifying the existing codes several times), and if the code is published, it can also be cited upon later use.

Other comments:
Line 23 (and other places in the text): I was very confused what field-realistic exposure is. As a 'general' reader I would prefer to see a more self-explanatory term. Maybe 'agrochemical levels that are realistic in agricultural fields'. Also, this is later (line 424) called 'filed realism'.

Line 24/25 Not sure that 'in contrast' is the correct wording here.

Line 25 Could you just provide a couple of examples of proxies of fitness (otherwise it is very difficult to imagine what these could be here). Same for the line 76.

Line 27 Why would these results on additive effects on other variables (behaviour etc) suggest that death of workers is the main concern. My understanding (after reading the rest of the MA) was that these were tested as possible mechanisms that lead to the detected main effects on mortality.

Lines 65-67 I would suggest moving this sentence before the sentence starting with 'Here, we present'… (line 60)
Line 76 here called 'parasite intensity' and 'parasite load' in abstract

Line 94/95 'With individual studies demonstrating…' I am unclear about what this means. Does it mean that the heterogeneity was cause by differences in interaction effects?

Line 110/111 I suggest putting these results first, before the results on agrochemicals and nutritional stress etc.

Line 134-136 I find the reasoning here difficult to understand.

Referee #4 (Remarks to the Author):

This paper presents a meta-analysis that is a systematic review of co-exposure at threats in the literature on bees. It provides a quantitative estimate for the effect of combined stress. The analysis focused on the comparison of combinations of different types of stress, biotics or abiotics. But the sample of compared corpus are very heterogeneous (C. Data & methodology), with those on the effects of pesticides on mortality of the honeybee that is largely the biggest: there are 2 times fewer publications on behavior, more than 4 times on fitness, 5 to almost 6 times less on pathogen load and immune. Regarding the type of stress, nutrition-related factors are weakly represented. Sampling error decreases as sample size increases and, therefore, larger studies should provide more precise estimates of the true effect. Thus, the most robust conclusion of this publication concerns the effects of co-exposure with pesticides, which is not an original result (B. Originality and significance). The discussion of the results lacks content because a lack of robustness for results obtained outside the effect of pesticides on the honeybee are weak. Pesticides are known to be synergistically toxic when they occur together, or when combined with certain others pesticides, in particular with fungicides that are potent cytochrome P450 monooxygenase enzyme inhibitors (e.g. Colin and Belzunces, 1992; Pilling and Jepson, 1993; Thompson and Wilkins, 2003; Andersch et al.

2010; Iwasa et al., 2004). The synergistic effects of the pesticides identified in this meta-analysis may be mainly due to this effect of fungicides inhibiting the detoxification mechanisms. Cedergreen (2014) present a review of the scientific literature in animals or plants and three main groups of environmentally relevant chemical toxicants: pesticides, metal ions and antifouling compounds. Synergistic mixtures for pesticides included cholinesterase inhibitors or azole fungicides in 95% of 69 described cases. I will advise the authors to check whether the results obtained are due to the effects of these categories of pesticides. The authors could thus test the sensitivity of the results of their model according to the weight of the work on pesticides.

A meta-analysis can is subject to publication bias because studies with zero results are less likely to be published and, therefore, results from meta-analyses may overstate/understate a treatment effect. We can assume that this bias is highest with studies on pesticides (D. Appropriate use of statistics and treatment of uncertainties). The hypothesis will be that articles showing synergy of pesticides have a higher submission rate than those without synergic effect. Indeed, the impact of a study on the synergy of pesticides is greater in the scientific community. On the contrary, the works on interactions between pesticides and other threats would be less underestimated by researchers, whichever obtained results. One strategy to minimize publication bias will be to assess well-known researchers in the field of apidologie to discover whether they have conducted a negative study that remains unpublished (F. Suggested improvements). Secondly, the absence of bias in the small corpus (e.g. non-apis, pathogen load, immune) can be checked with a funnel plot, which visually reveal the presence of a publication bias (Duval and Tweedie, 2000). Thirdly, in the face of the unbalanced sample and the risk of overrepresentation of synergistic effects, one solution could be to compare the impact of co-exposure to pesticides in different invertebrates, comparing the honeybee with other animals. For example, the results of Cedergreen (2014) showed that synergy occurred in 7% of the 194 binary pesticide mixtures. This fact would suggest that bee react differently to co-exposure than other organisms - the authors could discuss this hypothesis.

The authors suggest that their results may influence public policy interventions to protect bees (line 60). They must specify the framework within which their results might be used (E. Conclusions: robustness, validity, reliability). I think that the framework is the pesticide risk assessment in the regulatory procedures (pre- and post-homologation). But the readers may understand that agri-environmental measures to reduce exposure to pesticides should be preferred to enhanced food resources. However, the criteria for choosing the publications in this article (Treatment A, Treatment B, Treatment AB, Control) are not adapted at landscape-scale. Yet, the ecological intensification in favor of bees advises the combination of reduced pesticide use with improvement of trophic resources (Zhang et al., 2017; Kovács-Hostyánszki et al., 2017; Garibaldi et al., 2019).

The figures representing the retrieval rate are numerous but they are not presented and discussed in the ms (C. Data & methodology) The authors need to better explain this parameter, the current text is not clear.

References

Colin, M. E., & Belzunces, L. P. (1992). Evidence of synergy between prochloraz and deltamethrin in Apis mellifera L.: A convenient biological approach. Pesticide Science, 36(2), 115-119.

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. Biometrics, 56(2), 455-463.

Cedergreen, N. (2014). Quantifying synergy: a systematic review of mixture toxicity studies within environmental toxicology. PloS one, 9(5), e96580.

Garibaldi, L. A., Pérez-Méndez, N., Garratt, M. P., Gemmill-Herren, B., Miguez, F. E., & Dicks, L. V. (2019). Policies for ecological intensification of crop production. Trends in ecology & evolution, 34(4), 282-286.

Kovács-Hostyánszki, A., Espíndola, A., Vanbergen, A. J., Settele, J., Kremen, C., & Dicks, L. V. (2017). Ecological intensification to mitigate impacts of conventional intensive land use on pollinators and pollination. Ecology Letters, 20(5), 673-689.

Pilling, E. D., & Jepson, P. C. (1993). Synergism between EBI fungicides and a pyrethroid insecticide in the honeybee (Apis mellifera). Pesticide Science, 39(4), 293-297.

Thompson, H., & Wilkins, S. (2003). Assessment of the synergy and repellency of pyrethroid/fungicide mixtures. Bulletin of Insectology, 56(1), 131-134.

Zhang, W., Ricketts, T. H., Kremen, C., Carney, K., & Swinton, S. M. (2007). Ecosystem services and dis-services to agriculture. Ecological economics, 64(2), 253-260.

Author Rebuttals to Initial Comments:

Comment to all reviewers

We are grateful to the reviewers for their helpful and thoughtful comments on our manuscript, which we have addressed in point-by-point responses below (reviewer comments in normal text, our responses in italics). In responding to the reviewers' comments, we have made two main changes to the manuscript, which we believe have significantly strengthened the manuscript and clarified the interpretation of the results:

1. In response to the reviewers' comments, we spent time thoroughly editing any aspects of the analysis that had been unclear, and have provided publicly available copies of the code that we used to promote transparency. When proofreading our raw data, we noticed one step in our methods that had been overlooked, and could bias our results against being able to detect synergy if it is present (meaning that our previous analyses were highly conservative) and potentially suggesting antagonism when the actual result is additive or synergistic (see below for further explanation). For a limited set of studies, predicted additive effects exceeded the boundaries of experimental observation (for example, predicting a mortality rate of more than 100%, which could not be observed). We have explained why this is problematic in the text:

"….in such cases the true interaction effect cannot be estimated. For example, if hypothetical Stressors A and B both cause 60% mortality relative to the control group, the predicted mortality of the combined treatment (120%) exceeds the boundary of observable values (100%), rendering synergistic and additive interactions impossible to detect, and apparently antagonistic interactions unreliable"

*To deal with this issue, following Jackson et al. (2016), we removed any data points for which the predicted additive value was biologically impossible. This resulted in the removal of 29 effect sizes (leaving a total of 356) across 10 studies. Re-running the analyses on the resulting dataset did not change the main results of the study but strengthened our conclusion about synergistic effects of agrochemicals on bee mortality as the overall effect size for agrochemical*agrochemical interactions became larger. As a result, the overall effect of combined stressors on mortality became synergistic as well.*

2. Two of the reviewers commented on the high hetereogeneity in the data. We agree that this should be discussed in more detail than had previously been the case and have significantly added to the discussion to address this point (lines 111-119 & 161-163).

Detailed responses to comments by individual reviewers:

Referee #1 (Remarks to the Author):

A. The manuscript reports the results of a meta-analysis of 385 combined effects of multiple stressors on bees. The main results are: i) synergistic effects occurred when bees were exposed to multiple

agrochemicals at field realistic levels; ii) antagonistic interactions occurred when bees were exposed to multiple pathogens, or when nutritionally stressed bees were inoculated with pathogens; iii) almost all interactive effects on proxies of fitness, behaviour, parasite load and immune response were additive.
B. This work the first meta-analysis on the subject. It is a huge undertaking, particularly in the selection process, so I congratulate the authors for their great work.
C. The statistical analysis follows the methodology of similar work on multiple stressors carried out on aquatic ecosystems. The dataset is robust, the analysis was properly conducted and the results are reliable and valid.
D. My only concern is the large hetereogeneity observed between the datasets, which the authors acknowledge but do not comment in detail. As I mention below, this is not of great concern because meta-analysis deal usually with a mixed bag of experimental data, often showing contrary results, and therefore some heterogeneity is to be expected. However, this shouldn't be an excuse for accepting all findings without reservations, so the authors could be more cautious in interpreting the results of their analysis.
E. The conclusions are well spelled out in the abstract, but not so much in the main text. The paper ends with some recommendations but does not stress the main point, i.e. that current regulation of agrochemicals fails to protect bees and the biodiversity of the ecosystem. Further references to back up this main conclusion are suggested – see below.

Points A, B, C) - Thank you for taking the time to review our paper, and for providing such positive comments.

Point D) - We appreciate that there was a large amount of heterogeneity in the different data sets and are grateful that you highlight that this is not uncommon when collecting such a wide variety of data.

*We agree that caution is needed in interpretation of heterogeneous data. To this end, we have added a paragraph of discussion of this issue (lines 111-119). In this paragraph, we caution against generalization, particularly for non-agrochemical stressors, but also point out that within the agrochemical*agrochemical class, it is likely that synergy is more common for some groups than others.*

Point E) - We now state that the regulatory process is failing to protect bees from the unwanted consequences of complex agrochemical exposure (line 211-213) and have added the suggested reference (line 213).

F. Specific comments:
• Line 41: the initial statement is incorrect; Foley et al. 2005 estimated that roughly 40% of the surface of the planet is dedicated to 'cropland and pastures', but intensive agriculture is only a fraction of that figure, most likely 10-12% (Lambin & Meyfroidt 2011, PNAS 108: 3465; and FAOSTAT Online statistical service).

Thank you for pointing this out. We have removed this percentage from the text (line 40). Unfortunately, Lambin and Meyfroidt do not explicitly state the percentage of land that is dedicated to intensive agriculture around the world, and similarly the FAO data do not distinguish between land use for intensive vs extensive agriculture. Consequently, we have not cited either of these (please note that these are the only suggested references that we have not added). We were unable to find another reference that accurately determined this value, so the sentence has been edited so that the percentage is not required (line 40).

• Lines 95, 154 and 162: heterogeneity is very high in all cases, although this is partly an inherent feature of meta-analysis done on a diverse array of experimental designs, multiple agrochemicals and pathogens. I noted the authors comment on this point later in the manuscript (lines 180-5), but they should indicate more caution in accepting these results.

Thank you – we have responded to this above – but to reiterate, we have now edited the manuscript to make it clear that caution is needed in the interpretation of results due to high heterogeneity (lines 111-119 & 161-163).

• Lines 163-4: the statement "…previous research that has suggested that agrochemical exposure can suppress the immune response of bees and leave them more vulnerable to other stressors 3,17,28,…" is inaccurate. These three citations refer specifically to the effect of neonicotinoids on the bees immune system based on the mechanism of immune suppression described by Di Prisco et al. (ref 17), which has been confirmed by other authors – i.e., see Brandt et al. 2016, Journal of Insect Physiology 86, 40-47, and other work by this lab. However, the paper by Di Prisco and colleagues also showed that the insecticide chlorpyrifos did not produce inhibition of the immune system in bees, and to my knowledge such an inhibitory effect has not been found with any other agrochemical. Therefore, it appears that only neonicotinoids are responsible for the immune suppression effect on bees and other animals, not other agrochemicals, and this should be made clear in this sentence.

Thank you for pointing this out. We have now edited this sentence to make it clear that the references mentioned only concern neonicotinoids. We have also added the suggested references (line 157).

• Lines 162-6: the statement "In contrast… we found an antagonistic effect on bee immune response when exposed to multiple agrochemicals (Immune response, agrochemical*agrochemical, d = -0.27, CI = -0.52 to -0.01),…" indicates a contradiction between the immune suppression effect of neonicotinoids and the findings of this meta-analysis, which is not true and therefore misleading. Bearing in mind the small d value and the huge heterogeneity of the dataset, your finding suggests that mixtures of two or more agrochemicals do not have additive or synergistic effects on the bees immune response, which is entirely different from neonicotinoids inhibiting the immune system and allowing viral infections to become deadly, i.e. synergistic effect of agrochemical and pathogen on mortality. Please rephrase the sentence to describe accurately your findings and not to mislead the readers.

Thank you for pointing this out. We agree that the original version was confusing/unintentionally misleading! We have now changed the language to make the results clearer. We have removed the word 'contradiction' from the paragraph and have generally softened the language. We also explicitly highlight the high levels of heterogeneity for these data sets, and, in the subsequent discussion of heterogeneity, point out that these results should be interpreted with caution (lines (lines 111-119 & 161-163).

• Lines 166-7: agree with the explanation given; it is likely that exposure to one agrochemical elicits the same or similar immune response as exposure to two or more agrochemicals, so the combined effects on the immune system is less than additive and hence the 'apparent' antagonism – there is no antagonism sensu stricto in this case.

Thank you. We are glad that you agree with this point. We would note that the use of antagonism here, and throughout the paper, doesn't imply anything specific about mechanisms, as it is a specific descriptor for the directionality of an effect. We have now added an additional reference to further back this point up (line 157). However, with the updated methods which reduce antagonistic bias (see above) this specific interaction is now additive rather than marginally antagonistic (see Figure 2).

• Line 177: the reference cited (30) is not appropriate for the changes in phenology of plants, which may not be due to climate change alone but also other factors. Suggestions for replacement: Ovaskainen et al. 2013, PNAS 110: 13434-13439, and Ulrich et al. 2020, Frontiers in Plant Science 11: 542125.

Thank you for pointing this out. The original reference has been removed and replaced with Ovaskainen et al. 2013 (line 182).

• Line 190: agree, and authors should cite here some of the recent publications that advocate a more realistic evaluation of agrochemicals, as the current regulatory system is not protective of the environment, i.e. Topping et al. 2020, Overhaul environmental risk assessment for pesticides. Science 367: 360.

We have now added the suggested reference and two other relevant references (line 191).

• Line 201: however, the appropriate dose for treating a crop is dictated by the effectiveness of the chemical formulation, not based on ecological considerations. It is easy for researchers to propose such recommendations, but agrochemical manufacturers and users of the products will never abide by them.

Thank you. We agree that, currently, considerations of appropriate dosages are largely driven by effectiveness, as measured by specific criteria. However, how effectiveness is measured could change (e.g., Colin et al 2020) and IPM explicitly works with agrochemicals in a different way to current intensive agricultural models. Consequently, we believe both that it remains incumbent upon researchers to make recommendations like this, and that there is hope for future compliance. For this reason, we would like to keep this recommendation in the text.

• Line 210: please refer here Brühl & Zaller 2019, Biodiversity decline as a consequence of an inappropriate environmental risk assessment of pesticides. Frontiers in Environmental Science 7: 177.

Thank you. This reference has now been added (line 213).

• Line 434: shouldn't you add Table S2 in addition to Figure S2?

Thank you. This was a typo. The correct link is Table S2. This has been added (line 472)

• Line 455: replace Figure S3 with Table S3

Thank you. Changed (line 474)

• Line 436: replace Figure S4 with Table S4

Thank you. Changed (line 475).

G. Referencing is appropriate in most cases; I pointed above one replacement and a couple of additional references that should be added for completeness.

Thank you. The suggested references have been added with the exception that is highlighted above.

H. The manuscript is well written. The abstract is clear, concise and to the point. The results are well described and explained with reference to previous work. The figures are of good quality and appropriate.

Thank you.

Referee #2 (Remarks to the Author):

Siviter et al
Overall impression of the MS is that it is very clearly written and presented and the statistical analysis is sound. The findings are significant especially the different levels of synergy (or lack of) among stressor combinations or classes. This will interest not only pollinator biologists - where the interplay between drivers or pressures affecting pollinators is recognised as a potential threat but also a major knowledge gap (see IPBES thematic assessment of pollinators and pollination 2016) - but also to ecologists and epidemiologists studying the interactive effects of multiple stressors on ecological or pathogen systems generally. I enjoyed reading the MS and I can see how it will attract a lot of citation. I only have minor points of clarification.

Thank you for taking the time to review our manuscript, and for these positive comments.

Minor specific comments:

Summary – make clear that you are working on social and solitary bees somewhere, L27 – just workers or individuals in the case of non-social bees?

Thank you for pointing this out. We have now changed this to social and solitary bees at the start of the main text (line 47).

L41 Conventional intensive agriculture – to distinguish from other forms that may be intensive in management but perhaps less harmful to nature e.g. agroecological, organic

Thank you. This has now been edited (line 40).

L48 maybe a good place to insert "social and solitary" before "bees"

Thank you. This has been added (line 47).

L70 maybe add citation 12 here also ?

Added (line 70).

L73 is the description of a crossed design in () really needed ?

Thank you. While we agree this definition probably is not required for many readers, it may be helpful for some, and thus we would like to retain it for clarity.

L75 use of « category » here is slightly confusing with the subsequent use of categories (mortality etc) in text and figures. Suggest for the incidence here where you are refering to different stressors you use « type » or « class » - just a point of consistency/clarity.

Thank you for pointing this out! We now refer to stressor class here (lines 75) and either type/ class throughout depending on the context and sentence semantics.

Fitness proxies – overall looking at the supp. table I am convinced by most metrics, but I need some justification for the use of lipid ratios and protein content as proxies of fitness ? To be honest I dont expect their omission makes much of a difference to the results (few records), but it just needs to be stated how they correlate with and serve as a proxy of individual (colony) fitness. Is protein/fat content an indicator of fecundity (which is itself a fitness proxy)?

Thank you. Higher protein lipid ratios and protein content are associated with better energy and fat reserves, which are essential for hibernating queens and forager fitness (Arrese & Soulages 2010)*, which will have knock-on effects for colony fitness. In honeybees, the protein content of the head is associated with the protein level in the hypopharyngeal glands and the capability to feed offspring, which in turn impacts on colony fitness* (van Dooremalen *et al.* 2013)*. We have now added both relevant references* (Arrese & Soulages 2010; van Dooremalen *et al.* 2013) *to Table S1 to make this clear.*

L95 Extended data displays – check the order of appearance and how they are cited, EDF4 appears here first (cf L109), in some cases they are cited as EDF/T (L128) in others written in full – consistency.

Thank you. We have made this consistent throughout and have re-ordered the figures.

L107-117 I wonder if EDF 2 should actually be included as a panel in Fig1 given the importance of it to the main text?

Thank you. We agree that the results of EDF 2 are important and so now have included it as a figure in the main text (see Figure 1C).

L121 use of “and” and “&”

Thank you. This has been edited.

L142 Osmia synergistic – is it really? EDF1 does not appear to show that, additive at best?

Thank you for picking up this error. This has now been edited to additive (line 135).

L164 is Alaux et al 2010 worthy of citation here?
https://royalsocietypublishing.org/doi/10.1098/rsbl.2009.0986

We have not cited it here because the sentence is focused on agrochemicals, and the suggested reference is focused on nutrition. However, we do agree it’s a great paper!

L177 what about Baude et al 2016 cited here alongside 29?
https://www.nature.com/articles/nature16532

Thank you. Reference added (line 181).

L183-184 not sure how to read this statement from EDF4? How does it relate to interactions BETWEEN the stressor classes (parasites x agrochemical x nutrition?). What does each row in a figure panel correspond to? Also y axis in EDF4 lacks a label and here X_axis is Observed outcome whereas elsewhere it is Effect size

Thank you. We agree that this figure link was confusing. We have now removed it. Each line in EDF1 (previously EDF4) refers to an individual effect size, this has now been made clear in the figure legend and on the x axis. The y axis on caterpillar plots is typically left blank (e.g. https://www.metafor-project.org/doku.php/plots:caterpillar_plot), but we have now clarified in the figure caption that individual effect sizes are sorted from most negative to most positive (See Extended data figure 1).

L210 cite Potts et al here at end of sentence?

Thank you - this reference has been added (line 213).

Other cites that may be relevant or missing ?:

Collison, E., Hird, H., Cresswell, J. & Tyler, C. (2016) Interactive effects of pesticide exposure and pathogen infection on bee health - a critical analysis. Biological Reviews, 91, 1006-1019.

Retschnig, G., Williams, G.R., Odemer, R., Boltin, J., Di Poto, C., Mehmann, M.M., Retschnig, P., Winiger, P., Rosenkranz, P. & Neumann, P. (2015) Effects, but no interactions, of ubiquitous pesticide and parasite stressors on honey bee (Apis mellifera) lifespan and behaviour in a colony environment. Environmental Microbiology, 17, 4322-4331.

Both references have been added.

I think an additional online extended data table giving the full citation of the 296 papers examined or the final 100 papers that provided data for this metaanalysis is advisable, so that the reader can identify if there were any notable omissions.

Thank you. We absolutely agree that this is a really important point, not only because we want readers to be able to quickly find the full reference list, but because we also want to ensure that authors are cited for their research. Nature recommends between 30-40 references in an article, so we cannot include a table in the main text, or the extended data that references each paper. We have included a list of all 296 papers here (https://osf.io/8xnua/). This includes all the papers included in the analysis. If this manuscript is accepted by Nature, we will ask them how best to proceed so that all the papers used will be cited. If all else fails, we will refer to this weblink in the published paper. For now, though please find a list of all the papers included here (https://osf.io/8xnua/).

I wonder if the authors would consider a sentence calling for study of other multi-stressor interactions in pollinators or pollination (e.g. including effects of climate variability/extremes or other effects on populations of habitat loss). A major challenge lies in looking beyond these parasite-nutrition-chemical interactions to others happening in real landscapes. Of course that is a major knowledge gap and data are simply not there yet for quantitative analysis, but identifying such a need could be useful.

Thank you. We agree and have added this suggestion (line 186-189).

Referee #3 (Remarks to the Author):

In this paper the authors meta-analyse the evidence for synergic, additive, and antagonistic effects of anthropogenic stressors on bees. I find this paper well written, the literature search comprehensive, and the topic and findings interesting and important from the conservation and agricultural management perspective. I have mostly concentrated on the analysis part of this work, and provide only some minor comments on other aspects.

Thank you for reviewing our paper and for providing such a positive review.

Meta-analysis

Overall I think that the authors did a good job in collecting relevant studies, reporting on search and extraction procedures, and conducting the analysis itself. All the main components of a meta-analysis have been applied (e.g. calculating heterogeneity, publication bias test).

Thank you!

My comments are:

1) search terms
It would be easier for a reader to follow the reasoning behind the search terms if there was a sentence

that explains what different parts of the search string stand for. E.g. the first part of the search string describes the target taxa, the second different stressors, and the third different responses.

Thank you. We agree and now pre-empt our full search terms with the framework we used (line 334-335). This was population, stressors, response variable.

Second, there are several instances where unnecessary terms appear. For example, if there is 'fitness' there is no need for 'colony fitness'; or if there is 'output' there is no need for 'reproductive output' too. From my experience, the more detailed terms (like reproductive output) are only used when the more general term ('output') provides too many results, and then the reviewer tries to eliminate the irrelevant terms by being more specific in defining the exact type of the 'output' (e.g. 'reproductive output').

Thank you. We agree that some of the terms used may not produce many additional papers, but we wanted to ensure that we collected as many studies as possible. Perhaps we were only successful in making more work for ourselves, but we wanted to be extra careful!

Third, why is there no term 'lethal' (while there are different forms of 'sublethal')

The term 'sublethal' is widely used in the ecotoxicological literature, as well as in the broader literature of effects of pathogens and nutritional stress. We used the similarly widely used terms 'mortality' and 'survival' (e.g. Gill, Ramos-Rodriguez & Raine 2012; Retschnig, Neumann & Williams 2014; Kairo *et al.* 2017)*, rather than the term 'lethal', which is little used in the literature.*

Finally, it would be easier to follow the search terms if those relating to a same category were closer together (e.g. 'foraging' and 'navigat*)

Thank you. This has now been edited (lines 335 - 355)

2) Screening process and data extraction
What do you consider to be 'consistent' in line 340. From my experience it is rarely 100% agreement.

Thank you. Consistency was 90%. This has been added to the text line 363.

One article was not possible to translate, out of how many in a non-English language?

Thank you. Most of the papers had English abstracts and were rejected at the abstract stage. We did not record the language of papers when collecting the data so unfortunately we do not have this information available.

In 49 cases authors provided needed information, out of how many?

We contacted approximately 81 people but as we had multiple people collecting the data, and at times we were forwarded to co-authors, we are not 100% sure of how many authors were contacted so we would prefer not to quote it in the text.

3) The decision on the selection of the ES when more variables in the same category were reported in the same study was to randomly select one value. First, how many of these cases were there? Second, the results (especially if the number of these cases was relative high) would be more robust if this process was iterated a number of times (so, random ES selected, MA conducted; new random ES selected, MA repeated). This is easy to do with the loops in R (and one can only check the frequency distribution of the overall effect sizes obtained in this way)

Thank you for suggesting this. We agree that the approach you outline here is a valid one. However, we did not extract every single effect size in each paper, which would be required to action this suggestion. Instead, we took a parallel approach and randomly chose which data to extract before recording the relevant information and calculating the effect sizes. We chose this approach because we had to contact authors for a considerable number of the data sets, and in contacting authors we found we were much more likely to get a response if we requested a specific subset of the data, as opposed to the complete dataset (as would be required to take the approach you suggest). However, we appreciate the value in your suggested approach and would apply it at the appropriate stage in future data-gathering exercises. In this case, we cannot retrospectively apply it but in its absence we believe that random selection is an appropriate method to avoid bias.

4) While the effect sizes used in the analysis are provided, I would also suggest providing R codes. This is first, because of computational reproducibility, but also as in this way others can benefit from having access to it (I have myself benefited from modifying the existing codes several times), and if the code is published, it can also be cited upon later use.

Thank you for suggesting this. The code is now shared here (https://osf.io/8xnua/).

Other comments:

Line 23 (and other places in the text): I was very confused what field-realistic exposure is. As a 'general' reader I would prefer to see a more self-explanatory term. Maybe 'agrochemical levels that are realistic in agricultural fields'. Also, this is later (line 424) called 'filed realism'.

Thank you for highlighting the mis-spelling, which we have now corrected. Field realism is a standard term in agro-ecology but we recognise that not all readers will be familiar with what this means. We have now, the first time this term is used in the main text (line 120-121), referred the readers to the Methods, where we give a clear definition of this term (lines 463-467).

Line 24/25 Not sure that 'in contrast' is the correct wording here.

Thank you. We agree and have re-worded.

Line 25 Could you just provide a couple of examples of proxies of fitness (otherwise it is very difficult to imagine what these could be here). Same for the line 76.

Thank you. We haven't added examples in the abstract due to the word limit, but we have included examples in line 76.

Line 27 Why would these results on additive effects on other variables (behaviour etc) suggest that death of workers is the main concern. My understanding (after reading the rest of the MA) was that these were tested as possible mechanisms that lead to the detected main effects on mortality.

Thank you. We have now edited this to make it clearer (line 25-26).

Lines 65-67 I would suggest moving this sentence before the sentence starting with 'Here, we present'… (line 60)

Thank you. We agree and have made the suggested change (lines 62-67).

Line 76 here called 'parasite intensity' and 'parasite load' in abstract

Thank you. This has been changed to parasite load throughout.

Line 94/95 'With individual studies demonstrating…' I am unclear about what this means. Does it mean that the heterogeneity was cause by differences in interaction effects?

Thank you. Here we are trying to say that heterogeneity was high, perhaps because the interaction effects between different stressors was causing different types of interactions. We have changed this line, and used the word effect size instead of studies (line 97-99).

Line 110/111 I suggest putting these results first, before the results on agrochemicals and nutritional stress etc.

Thank you. We agree and have re-ordered the entire paragraph (lines 120-129)

Line 134-136 I find the reasoning here difficult to understand.

Thank you. We have now removed this sentence.

Referee #4 (Remarks to the Author):

This paper presents a meta-analysis that is a systematic review of co-exposure at threats in the literature on bees. It provides a quantitative estimate for the effect of combined stress. The analysis focused on the comparison of combinations of different types of stress, biotics or abiotics. But the sample of compared corpus are very heterogeneous (C. Data & methodology), with those on the

effects of pesticides on mortality of the honeybee that is largely the biggest: there are 2 times fewer publications on behavior, more than 4 times on fitness, 5 to almost 6 times less on pathogen load and immune.

Thank you. It is true that there are differences between the sizes of datasets for different response variables and stressors, reflecting the different number of studies conducted on a particular topic. We agree that the mortality data is the most robust data set, which is why we discuss it more than the other data categories. Collecting the mortality data, and indeed data from the other categories was, as highlighted by reviewers 1 & 3, a massive undertaking. However, in response to this comment, we have now added a whole new paragraph which highlights that data with high levels of heterogeneity and smaller sample sizes should be treated with caution (lines 111-119). We also specifically discuss the heterogeneity in the data within the discussion (lines 161-163) and make it clear in the text that while meta-analysis is a broad tool used to determine general patterns, the detail of individual interaction types should certainly not be ignored. We believe that these quantitative general results remain both novel and important, as supported by comments from the other reviewers.

Regarding the type of stress, nutrition-related factors are weakly represented.

We agree, and now point out that this is a key knowledge gap at lines (179-189). While, due to small sample size, our analysis of the interaction effects between nutritional stressors and other anthropogenic stressors is not as robust as for agrochemical interactions, it provides the first quantitative estimates of these interactions. This allows us to make more specific suggestions such as, as suggested by reviewer 2, highlighting the need for more landscape scale experiments, that incorporate the interactions between nutritional and other stressor types (line 186-189).

Thus, the most robust conclusion of this publication concerns the effects of co-exposure with pesticides, which is not an original result (B. Originality and significance).

We agree that the interactions between multiple agrochemicals on mortality is our most robust dataset. Individual studies have suggested this may be the case, and systematic review has also put it forward as a possibility, but meta-analysis finally provides quantification of the evidence. Thus, we believe that our findings are novel and important to a broad array of scientists for the following reasons. By quantifying agrochemical interactions and comparing them with other anthropogenic interaction effects (e.g. parasites, nutritional stressors), we have (i) provided the first quantitative synthesis of the current state of the field that, as pointed out by reviewers 1,2 and 3, is novel, important, and useful in identifying knowledge gaps; (ii) quantitatively demonstrated that the interactions between agrochemicals are much more likely to induce a synergistic interaction than other stressors; and (iii) demonstrated that these synergistic agrochemical interactions broadly occur at ***field realistic levels****. This information is directly beneficial for policy makers, conservationists, ecologists, pollination biologists and epidemiologists (as pointed out by the other reviewers, see above).*

The discussion of the results lacks content because a lack of robustness for results obtained outside the effect of pesticides on the honeybee are weak.

We have outlined above why we consider our results to be robust and, in addition, identify gaps in the literature. As with any study, some results are stronger than others (here, as with any meta-analysis, this is determined by the number of previous studies and the heterogeneity within their results), but we

discuss all of our results within the context of previous work in the field. One of the aims of any meta-analysis is to identify knowledge gaps in the existing literature. We believe we have achieved this aim and highlighted the areas where more research is needed; this could not have been done without the systematic review of the literature which we have conducted. Consequently, and with respect, we disagree that the discussion lacks content.

Pesticides are known to be synergistically toxic when they occur together, or when combined with certain others pesticides, in particular with fungicides that are potent cytochrome P450 monooxygenase enzyme inhibitors (e.g. Colin and Belzunces, 1992; Pilling and Jepson, 1993; Thompson and Wilkins, 2003; Andersch et al. 2010; Iwasa et al., 2004). The synergistic effects of the pesticides identified in this meta-analysis may be mainly due to this effect of fungicides inhibiting the detoxification mechanisms. Cedergreen (2014) present a review of the scientific literature in animals or plants and three main groups of environmentally relevant chemical toxicants: pesticides, metal ions and antifouling compounds. Synergistic mixtures for pesticides included cholinesterase inhibitors or azole fungicides in 95% of 69 described cases. I will advise the authors to check whether the results obtained are due to the effects of these categories of pesticides. The authors could thus test the sensitivity of the results of their model according to the weight of the work on pesticides.

Thank you for these suggestions. We were aware that synergistic interactions had been observed in individual experiments, as you note – one of the aims of our meta-analysis was to determine what the overall effect of such multiple exposure was. We found it to be synergistic for mortality, as noted above and in the manuscript. The excellent systematic review by Cedergreen applies a definition of synergy based on differences in means, whereas our definition is statistical and based around both mean and variance. Nevertheless, as noted by the reviewer, Cedergreen identified two categories of pesticides as being represented in most of the cases of synergy observed in their study (cholinesterase inhibitors and azole fungicides). However, Cedergreen also showed that 5 groups of pesticides - the organophosphate and carbamate insecticides (Cholinesterase inhibitors), azole fungicides (Ergosterol biosynthesis inhibitors), triazine herbicides (Photosystem II inhibitors) and pyrethroid insecticides (interferes with sodium channels in nerve cells) – were over-represented in synergistic mixtures more broadly. Consequently, whether cholinesterase inhibitors or azole fungicides are more important than the other 3 categories in driving synergism, or whether this is simply a result of study bias (with more studies containing these former categories), remains unclear from this analysis.

To investigate this with our dataset, in Table S5 we identified the number of effect sizes contributed by each stressor. As can be seen from this, carbamates and organophosphates (the cholinesterase inhibitors) contribute 1 effect size with respect to mortality, so they are unlikely to have caused the synergism identified in this study. The same thing is true for triazine herbicides. However, azole fungicides contribute to 45 effect sizes while pyrethroid insecticides contribute to 15 effect sizes for mortality.

Therefore, in response to the above comments, we have now included a new Extended data figure that specifically quantifies the interactions between azole fungicides and pyrethroids (EDF4). We now explicitly state in the main text that the synergistic interactions between fungicides and other agrochemicals could potentially drive the observed overall synergistic effects on bee mortality (lines 174-178). We have also cited the suggested reference throughout and highlight the importance of understanding specific agrochemical interactions, as commercial formulas regularly contain multiple agrochemicals, and fungicides, a point that we now explicitly state (lines 196-199). We thank the

reviewer for bringing up their point, as we believe these additions have strengthened the message and impact of our manuscript.

A meta-analysis can is subject to publication bias because studies with zero results are less likely to be published and, therefore, results from meta-analyses may overstate/understate a treatment effect. We can assume that this bias is highest with studies on pesticides (D. Appropriate use of statistics and treatment of uncertainties). The hypothesis will be that articles showing synergy of pesticides have a higher submission rate than those without synergic effect. Indeed, the impact of a study on the synergy of pesticides is greater in the scientific community. On the contrary, the works on interactions between pesticides and other threats would be less underestimated by researchers, whichever obtained results. One strategy to minimize publication bias will be to assess well-known researchers in the field of apidologie to discover whether they have conducted a negative study that remains unpublished (F. Suggested improvements). Secondly, the absence of bias in the small corpus (e.g. non-apis, pathogen load, immune) can be checked with a funnel plot, which visually reveal the presence of a publication bias (Duval and Tweedie, 2000).

Thank you for these suggestions. Duval and Tweedie (2000) proposed trim and fill analysis, which is based on funnel plots and which we have already conducted (lines 446-453). This approach assesses the asymmetry of funnel plots and adjusts the mean effect estimates for missing studies due to publication bias. We have now also provided funnel plots as a visual aid (see Figure EDF6). As our previous analyses showed, there is no evidence of a publication bias against close-to-zero (additive) or negative (antagonistic) effects in any of the data sets (see Figure EDF6). The only bias revealed by trim and fill was ***towards*** *antagonistic effects for behaviour and fitness traits*, *suggesting observed results on behaviour and fitness may underestimate the interaction effects between stressors (lines 446-453*). *Therefore, we found no evidence that our dataset is biased towards finding synergistic effects.*

Thirdly, in the face of the unbalanced sample and the risk of overrepresentation of synergistic effects, one solution could be to compare the impact of co-exposure to pesticides in different invertebrates, comparing the honeybee with other animals. For example, the results of Cedergreen (2014) showed that synergy occurred in 7% of the 194 binary pesticide mixtures. This fact would suggest that bee react differently to co-exposure than other organisms - the authors could discuss this hypothesis.

Thank you. As highlighted above we found no evidence of overrepresentation of synergistic effects (EDF6 & lines 446-453), so this is not a concern. We note that our study is not solely about honeybees – it is about social and solitary bees. Consequently, we already compare the honeybee (which has unique biology) to other animals, as you suggest. In addition, the study by Cedergreen did not assess presence of synergy based on the effect size approach and hence cannot be quantitatively compared to our meta-analysis approach used here. As ours is the first meta-analysis on interactive effects of pesticides on invertebrates, it is unfortunately impossible at present to compare the impact of co-exposure to pesticides in different invertebrates, but it could be done in future studies which use the same meta-analytic approach with our study providing a benchmark for such comparisons. We do now, however, cite the study by Cedergreen (2014) within the manuscript (line 175).

The authors suggest that their results may influence public policy interventions to protect bees (line 60). They must specify the framework within which their results might be used (E. Conclusions: robustness, validity, reliability). I think that the framework is the pesticide risk assessment in the regulatory procedures (pre- and post-homologation). But the readers may understand that agri-

environmental measures to reduce exposure to pesticides should be preferred to enhanced food resources. However, the criteria for choosing the publications in this article (Treatment A, Treatment B, Treatment AB, Control) are not adapted at landscape-scale. Yet, the ecological intensification in favor of bees advises the combination of reduced pesticide use with improvement of trophic resources (Zhang et al., 2017; Kovács-Hostyánszki et al., 2017; Garibaldi et al., 2019).

Thank you for this comment. We refer in our manuscript to 'management interventions', which cuts across a swathe of possibilities, including risk assessment and subsequent reduced use of agrochemicals, agri-environment schemes to enhance nutrition, or management of the pathosphere. As our results emerge in the manuscript, we now specifically comment upon the frameworks that are relevant, e.g., the agrochemical regulatory process (lines 207-211, 194-197, 190-191), and agri-environment schemes (lines 185-188). We agree about the importance of improved trophic resources, and hope that this is now clear in the manuscript.

The figures representing the retrieval rate are numerous but they are not presented and discussed in the ms (C. Data & methodology) The authors need to better explain this parameter, the current text is not clear.

No specific line numbers are provided, but we assume this comment refers to the process of the data collection and literature search. We now refer to the PRISMA Figure (EDF5) which illustrate the process of data retrieval throughout to make this clearer (see methods section).

Reviewer Reports on the First Revision:

Referee #1 (Remarks to the Author):

Authors have addressed my queries correctly.

Only a small correction in line 21-22 of the abstract is needed: choose between revealed or provided, but do not use both terms in that sentence.

Referee #2 (Remarks to the Author):

I continue to find this MS worthy of publication in Nature. It was a pleasure to read. I note the efforts made by the authors to respond positively to the comments of all the reviewers and find the MS improved from an already high base.

(A) The key result for me (Fig. 1) is the overall consistent finding of synergy in chemical-chemical interactions impacting bee mortality but less so (there are still some incidence of synergy) in other stressor class combinations (e.g. parasite-nutrition or nutrition-nutrition). The additive or antagonistic patterns in non-lethal responses (eg fitness proxies, behaviour) means it remains intriguing as what the mechanisms are that underpin the synergistic effects, this, however, does not undermine the current analysis but rather serves to stimulate fresh investigations into this area. Specifics & context are obviously also important and well illustrated (eg EDF 2-3).

(B) I concur with the authors' assertion that this is the first meta-analysis, as opposed to a systematic or narrative review, covering such a breadth of multi-stressor interactions and bee impacts. It represents an advance of our understanding of the nature of multi-stressor interactions and their impact on social and solitary bee species. As I mentioned in my last review the MS will interest and inform diverse disciplines (biologists, ecologists and epidemiologists) studying the

interactive effects of multiple stressors on biota and pollinators in particular. It will also have great impact in terms of informing the policy (IPBES, CBD, FAO) and regulatory domains (national and EU environmental agencies) concerned with the interplay of factors affecting pollinators and their services.
The meta-analytical and statistical approaches taken and adjusted following the last peer review appear highly robust and remaining limitations reflect the available evidence base itself and not the analysis undertaken (C & D). The quality of text and data presentation is excellent (C). I feel the authors have made every effort to recognise and display (EDF1, EDF6) the inherent variability and associated uncertainties (D) with caveats in the text around their data and interpretations (E). I think their conclusions are sound, reasoned and justified and throughout they have recognised the limits to their study and identified knowledge gaps. The MS is very clearly and succinctly presented (E & H). References are appropriate (G).
I have no further suggestions for improvement (F) beyond a handful of very minor points (below) that can be rapidly implemented (or not).
Minor points:
Abstract:
L21_22 revealed AND provided – deletion
L23-24 A very minor point. "but interactions did not exceed additive predictions when" I understand but it seems slightly awkward phrasing for the abstract, perhaps reconsider adjusting it and including reiteration of fact you are discussing mortality at this point: but interactive effects on mortality were not greater than expected from additive combinations of parasite or nutritional stressors.
Last sentence, if you need to save a few words, consider : Environmental risk assessment schemes that assume additive effects of agrochemical exposure risk underestimating the interactive impact of anthropogenic stressors on bee mortality and fail to protect pollinators that provide a key ecosystem service supporting a sustainable agriculture.

L88 (e.g. effects of stressors on number of workers), - wonder if better to add worker 'bees' or replace number of workers with 'mortality' or 'survival'

L91 _94 This as explained in the rebuttal seems an evident improvement made during the revision that increases the analysis robustness.

Referee #3 (Remarks to the Author):

I find that the authors have thoughtfully addressed the comments from the reviewers. Only some minor comments from me
Regarding my previous comment on randomly choosing one ES per study – after the authors explained why they did this I agree with their approach as probably the most feasible under the circumstances. Maybe this explanation could be clarified in the MS too.
Lines 21/22: one of the words needs to be removed (revealed, provided)
Line 45 'increase' instead of 'increases'
Line 82 I think this sentence lacks 'would be' before 'observed'

Referee #5 (Remarks to the Author):

The manuscript "Agrochemicals, but not other stressors, interact synergistically to increase bee mortality" presents a very interesting meta-analysis of experimental studies testing the interaction of stressors on bees. This idea of stressors interaction enhancing bee mortality (and thus bee decline) has been in the air for 10 years or so now. Thus, for a decade, several papers showed that the combination of stressors can lead to synergistic effects on survival and other traits. But as with all experimental work, laboratory conditions, protocols, host genetics are all factors that can also have an impact on the outcome of such experiments. Therefore, to measure the impact of such

interactions, it is extremely pertinent to now look at the numbers, and make a meta-analysis, that can, much more than a systematic review can do, give an unbiased picture of the strength and importance of these interactions.

I think this manuscript very nicely achieved this goal. This work was performed with a lot of rigor and I also appreciate the amount of work authors have done filtering the literature and extracting meaningful data. The manuscript is very clear and concise, but also very careful about the limitations of the study -which, if I understand well, this last point mainly came after a first round of review. I found the meta-analysis well conducted in generally, and I particularly liked a couple of details in the analysis that make this work even stronger such as (i) the restriction to field realistic doses, (ii) the analysis of the most common pesticides, an (iii) the removal of studies that could not technically detect synergistic interactions due to large effects of single treatments. I think this work will attract the interest of many, first in the field of research, as authors nicely identified gaps in the literature, but also to the public and policy makers. We all need healthy environments for our pollinators, and this work shows that some progress has to be made for the conservation of bees in agricultural area.

The response to referees from the first round of reviews were satisfying. Authors addressed all points raised by referees, and I can see, as I mentioned above, how the manuscript was improved with more analyses, new discussions about limitations, and more accurate references. I disagree with reviewer 4 that criticized the lack of novelty about pesticide-pesticide interactions, as this is shown here with a meta-analysis, which itself is novel. Also, I don't think it is necessary to compare the result with this of other invertebrates, or try to collect more unpublished data: this meta-analysis has been conducted properly to avoid publication bias. The issue about heterogeneity of data is inherent to meta-analysis, and if I share the point of most reviewers that the data used here are heterogenous, I agree with authors that this is what a meta-analysis is made for: extract meaningful information for different type of data.

However, I was left with two main questions about data analysis, which for me are essential. First, I am not able to tell, from what I have read in the manuscript and the supplemental material, whether authors considered the level of pseudoreplication many studies on honeybees suffer from. Honeybees are eusocial insects. They live in groups, and for this reason most experiments on honeybees are conducted on groups of bees. Generally, such experiments are done in cages, ranging from 10 to 50 workers per cage. In such experimental set up, individuals within one cage must not be considered as independent samples. Unfortunately, this is too often overlooked and statistical analyses that ignore pseudoreplication violate the assumption of independence of samples. The true level of replication should be the cage, such as in block design experiments. Here, I understand that authors of this meta-analysis either calculated the effect size from the average means and SD found in the papers, or re-calculated the means and SD to compute the effect size. Therefore, I would like to ask if they considered cages and block design in their calculation. I can read, in line 388 of the revised manuscript that cage groups were considered, but what for? In the supporting information, I see that Hedge's d was weighted by sample size n. I think n should be clarified here. I believe this should be explicit in the paper, whether pseudoreplication was considered. Pseudoreplication can greatly inflate effect sizes because of the low within-treatment variation, and therefore it can lead to biased meta-analysis and wrong conclusions. See this paper nicely debating this issue: https://conbio.onlinelibrary.wiley.com/doi/full/10.1111/j.1523-1739.2010.01542.x. To some degree, this point is linked with the heterogeneity of data. But I believe that authors can account for it by considering the cage design in their calculation of d.

The second point that makes me uncomfortable with, is the analysis of mortality -which is central for the conclusion of this work about the effect of pesticide-pesticide interactions. I understand from table S1 that authors used the "Time point where first treatment group were all dead […] or the final time point if this was not available". This method is not the best to study survival. In fact, such an analysis of mortality would never be accepted nowadays in ecological studies. Everyone is now performing survival analyses, in which survival time is defined as the time that an individual has survived over a given period, and where hazard ratio are the statistics of reference when comparing treatments. Survival analysis will consider the dynamics of survival, and differentiate the effects of a

stressor which, for example kill 90% of individuals within 3 days, and the other 10% the remaining 10 days, to the other stressor that kill 10% of individual the first 10 days, and the remaining 90% the last 3 days. Here, one should consider the first stressor that induce a quick drop more hazardous than the second. Currently, the method used by authors do not allow to differentiate the effect of these stressors on survival, and they would basically call the same effect size. I think mortality here should be reanalyzed with survival analyses and hazard ratio. Besides, in modern survival analyses, one can include cage design (or plate design for larvae) as random factors. This is very handy for honeybee research (see this method paper: https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.2041-210X.2011.00099.x).

I think these two points should be considered by authors, and the analysis modified if authors did not account for pseudoreplication and survival times.

Author Rebuttals to First Revision:

Referee #1

Authors have addressed my queries correctly.

Thank you for taking the time to review our paper. We appreciate these comments and those made in the previous review.

Only a small correction in line 21-22 of the abstract is needed: choose between revealed or provided, but do not use both terms in that sentence.

Thank you for pointing this typo out. This is corrected on line 22.

Referee #2

I continue to find this MS worthy of publication in Nature. It was a pleasure to read. I note the efforts made by the authors to respond positively to the comments of all the reviewers and find the MS improved from an already high base.

(A) The key result for me (Fig. 1) is the overall consistent finding of synergy in chemical-chemical interactions impacting bee mortality but less so (there are still some incidence of synergy) in other stressor class combinations (e.g. parasite-nutrition or nutrition-nutrition). The additive or antagonistic patterns in non-lethal responses (eg fitness proxies, behaviour) means it remains intriguing as what the mechanisms are that underpin the synergistic effects, this, however, does not undermine the current analysis but rather serves to stimulate fresh investigations into this area. Specifics & context are obviously also important and well illustrated (eg EDF 2-3).

(B) I concur with the authors' assertion that this is the first meta-analysis, as opposed to a systematic or narrative review, covering such a breadth of multi-stressor interactions and bee impacts. It represents an advance of our understanding of the nature of multi-stressor interactions and their impact on social and solitary bee species. As I mentioned in my last review the MS will interest and inform diverse disciplines (biologists, ecologists and epidemiologists) studying the interactive effects of multiple stressors on biota and pollinators in particular. It will also have great impact in terms of

informing the policy (IPBES, CBD, FAO) and regulatory domains (national and EU environmental agencies) concerned with the interplay of factors affecting pollinators and their services.
The meta-analytical and statistical approaches taken and adjusted following the last peer review appear highly robust and remaining limitations reflect the available evidence base itself and not the analysis undertaken (C & D). The quality of text and data presentation is excellent (C). I feel the authors have made every effort to recognise and display (EDF1, EDF6) the inherent variability and associated uncertainties (D) with caveats in the text around their data and interpretations (E). I think their conclusions are sound, reasoned and justified and throughout they have recognised the limits to their study and identified knowledge gaps. The MS is very clearly and succinctly presented (E & H). References are appropriate (G).

I have no further suggestions for improvement (F) beyond a handful of very minor points (below) that can be rapidly implemented (or not).

Thank you for these positive and helpful comments and those provided in your previous review.

Minor points:
Abstract:
L21_22 revealed AND provided – deletion

Thank you. This has now been edited (line 22)

L23-24 A very minor point. “but interactions did not exceed additive predictions when” I understand but it seems slightly awkward phrasing for the abstract, perhaps reconsider adjusting it and including reiteration of fact you are discussing mortality at this point: but interactive effects on mortality were not greater than expected from additive combinations of parasite or nutritional stressors.

Thank you for your suggestion to increase the clarity of this sentence. We have re-phrased this along similar lines to those you suggest. We have added “of bee mortality” at the start of the sentence so that it is clear that the agrochemical statement also refers to bee mortality (line 21-22). We have changed “but interactions did not exceed additive predictions when” to “but interactions were not greater than additive expectations when bees were exposed to parasites and/or nutritional stressors” (line 23-24).

Last sentence, if you need to save a few words, consider: Environmental risk assessment schemes that assume additive effects of agrochemical exposure risk underestimating the interactive impact of anthropogenic stressors on bee mortality and fail to protect pollinators that provide a key ecosystem service supporting a sustainable agriculture.

Thank you, we have edited lines 27 – 29 as you suggest.

L88 (e.g. effects of stressors on number of workers), - wonder if better to add worker ‘bees’ or replace number of workers with ‘mortality’ or ‘survival’

Thank you, we’ve added the word bees (line 87).

L91 _94 This as explained in the rebuttal seems an evident improvement made during the revision that increases the analysis robustness.

Thank you – we are glad that you appreciate the edits we made.

Referee #3

I find that the authors have thoughtfully addressed the comments from the reviewers. Only some minor comments from me

Thank you for reviewing our paper, and for providing such helpful and positive comments, here and with the first review.

Regarding my previous comment on randomly choosing one ES per study – after the authors explained why they did this I agree with their approach as probably the most feasible under the circumstances. Maybe this explanation could be clarified in the MS too.

We agree this could be useful information and have added a summary to the methods in the main text (Line 402-405).

Lines 21/22: one of the words needs to be removed (revealed, provided)

Thank you. This has been edited (line 22).

Line 45 'increase' instead of 'increases'

Thank you. This typo has now been edited (line 44).

Line 82 I think this sentence lacks 'would be' before 'observed'

Thank you. 'would be' has been added (line 81).

Referee #5 (Remarks to the Author):

The manuscript "Agrochemicals, but not other stressors, interact synergistically to increase bee mortality" presents a very interesting meta-analysis of experimental studies testing the interaction of stressors on bees. This idea of stressors interaction enhancing bee mortality (and thus bee decline) has been in the air for 10 years or so now. Thus, for a decade, several papers showed that the combination of stressors can lead to synergistic effects on survival and other traits. But as with all experimental work, laboratory conditions, protocols, host genetics are all factors that can also have an impact on the outcome of such experiments. Therefore, to measure the impact of such interactions, it is extremely pertinent to now look at the numbers, and make a meta-analysis, that can, much more than a systematic review can do, give an unbiased picture of the strength and importance of these interactions.

I think this manuscript very nicely achieved this goal. This work was performed with a lot of rigor and I also appreciate the amount of work authors have done filtering the literature and extracting meaningful data. The manuscript is very clear and concise, but also very careful about the limitations of the study -which, if I understand well, this last point mainly came after a first round of review. I found the meta-analysis well conducted in generally, and I particularly liked a couple of details in the analysis that make this work even stronger such as (i) the restriction to field realistic doses, (ii) the analysis of the most common pesticides, an (iii) the removal of studies that could not technically detect synergistic interactions due to large effects of single treatments. I think this work will attract the interest of many, first in the field of research, as authors nicely identified gaps in the literature, but also to the public and policy makers. We all need healthy environments for our pollinators, and this work shows that some progress has to be made for the conservation of bees in agricultural area.

The response to referees from the first round of reviews were satisfying. Authors addressed all

points raised by referees, and I can see, as I mentioned above, how the manuscript was improved

with more analyses, new discussions about limitations, and more accurate references. I disagree with reviewer 4 that criticized the lack of novelty about pesticide-pesticide interactions, as this is shown here with a meta-analysis, which itself is novel. Also, I don't think it is necessary to compare the result with this of other invertebrates, or try to collect more unpublished data: this meta-analysis has been conducted properly to avoid publication bias. The issue about heterogeneity of data is inherent to meta-analysis, and if I share the point of most reviewers that the data used here are heterogenous, I agree with authors that this is what a meta-analysis is made for: extract meaningful information for different type of data.

Thank you for these positive comments. We are very glad that you appreciate the importance of our research!

However, I was left with two main questions about data analysis, which for me are essential. First, I am not able to tell, from what I have read in the manuscript and the supplemental material, whether authors considered the level of pseudoreplication many studies on honeybees suffer from. Honeybees are eusocial insects. They live in groups, and for this reason most experiments on honeybees are conducted on groups of bees. Generally, such experiments are done in cages, ranging from 10 to 50 workers per cage. In such experimental set up, individuals within one cage must not be considered as independent samples. Unfortunately, this is too often overlooked and statistical analyses that ignore pseudoreplication violate the assumption of independence of samples. The true level of replication should be the cage, such as in block design experiments. Here, I understand that authors of this meta-analysis either calculated the effect size from the average means and SD found in the papers,
or re-calculated the means and SD to compute the effect size. Therefore, I would like to ask if they considered cages and block design in their calculation. I can read, in line 388 of the revised manuscript that cage groups were considered, but what for? In the supporting information, I see that Hedge's d was weighted by sample size n. I think n should be clarified here. I believe this should be explicit in the paper, whether pseudoreplication was considered. Pseudoreplication can greatly inflate effect sizes because of the low within-treatment variation, and therefore it can lead to biased meta-analysis and wrong conclusions. See this paper nicely debating this issue: https://conbio.onlinelibrary.wiley.com/doi/full/10.1111/j.1523-1739.2010.01542.x. To some degree, this point is linked with the heterogeneity of data. But I believe that authors can account for it by considering the cage design in their calculation of d.

Thank you for bringing to our attention the lack of clarity in the methods regarding studies with caged bees. When extracting data for the meta-analysis we quickly realised that the majority of studies that reported mortality did so simply as the percentage that survived/died and a p value for the difference between treatments (54/64 of the studies did not provide an extractable SD/SE). Out of 33 studies that used groups of bees in cages, 30 did not provide information on the SD/SE at the cage level, or report the raw data that would allow it to be calculated. We thought carefully about the best way to deal with this, and we decided that for the mortality analysis it would be best to extract data for all studies at the individual level. This was because:

(1) As it is not possible to keep A. mellifera *workers or queens as individuals, removing studies with cage level replication that did not report relevant metrics would essentially remove the majority of studies that were carried out on* A. mellifera *adults, creating huge bias in the dataset*

(2) Removing studies with cages would also substantially reduce the sample size of our analysis (removing 74 data points from 29 studies).

(3) As it was not possible to calculate an estimate of standard deviation at the cage level for the majority of studies, we used the number of individuals as the n value because it corresponds to the level of the standard deviation that we calculated from the percentage mortality values.

For other categories of dependent variable, a much higher proportion had standard deviations or measures of error reported in the study. Although it was not always clear what level this was calculated at, it was most likely to have been estimated from nested models that included cage as a random factor. Where the n-value (individual or cage) relating to the standard deviation or measure of error was not obvious for a given study, an individual level n is reported. Of the studies that used cages, the standard deviation was clearly calculated at the cage level and reported as such for the following number of studies: fitness 3/8; behaviour 9/18; parasite 2/11; immune 0/6.

We have now clarified how we treated caged studies in lines 406-414. "For categories other than mortality, the sample size for studies using cages of more than one bee was at the cage level, where relevant data were reported and the n value relating to the SD was clear. The number of studies for which relevant data was clearly reported at the cage level was 3/8 studies using cages (fitness; total studies = 22); 9/18 (behaviour; total studies = 31); 2/11 (parasite; total studies = 22); 0/6 (immune; total studies =11). For mortality data, all studies with caged animals are reported at the individual level (33/64 studies used cages) as only 3 studies had the raw data to calculate the standard deviation at the cage level or reported a cage level standard deviation."

However, we absolutely agree that it is important to understand whether the results of our analysis would change if cage was considered as a replicate, and thank you for raising this point. To address this, we looked into the most appropriate way to assess the impact of inflated sample size on outcomes of meta-analyses. Following that, we have now conducted an additional sensitivity analysis using effective sample sizes, rather than the sample sizes used in our original analysis, following the method of Higgins et al. 2008 (now described in lines 470-476 in the main text and 57 to 110 in the supplement). As mentioned above, the majority of the cage data did not specify cage level variance. However, following Noble et al. 2017, we were able to use data from the three studies that did provide raw, cage level data to calculate an estimate of the intra-cluster correlation coefficient (ICC) for cage, and used the maximum value of two different estimation methods to estimate the effective sample sizes for each study and account for the non-independence of cages. In this new analysis, our results did not change (Lines 470-476) (e.g. (Mortality (ICC = 0.26), d = 0.18, CI = 0.08 to 0.28: Mortality (ICC = 0.76), d = 0.16, CI = 0.06 to 0.26 compared with the original estimates of d = 0.19, CI

= 0.08 to 0.29). In addition, we have put all the sub-set analysis below, and as you can see the results for these also do not change. From carrying out these additional analyses, it is clear that the interaction between multiple agrochemicals remains the only synergistic interaction. We have added this information to lines 470-476 in the main text and 57 to 110 in the supplement (Table S6). We believe that these additional analyses demonstrate the robustness of our results, and show that the sample size issue raised in your comment does not invalidate our conclusions.

Table S6: Results from sensitivity analysis on non-independence within samples. Note that the table above ICCA = an ICC of 0.26 and ICCB = an ICC of 0.76

Original results	**PE**	**CI**	**CI**
Parasite-Parasite	0.04	-0.16	0.24
Parasite-Agrochemical	0.1	-0.06	0.26
Parasite-Nutrition	-0.12	-0.42	0.17
Agrochemical-Agrochemical	0.32	0.13	0.52
Agrochemical-Nutrition	0.25	-0.008	0.51
Nutrition-Nutrition	-0.51	-0.83	-0.18
Overall	0.18	0.08	0.28
Results ICCA	**PE**	**CI**	**CI**
Parasite-Parasite	0.04	-0.16	0.25
Parasite-Agrochemical	0.12	-0.04	0.3
Parasite-Nutrition	-0.03	-0.25	0.17
Agrochemical-Agrochemical	0.31	0.12	0.49
Agrochemical-Nutrition	0.24	-0.003	0.49
Nutrition-Nutrition	-0.51	-0.83	-0.18
Overall	0.18	0.08	0.28
Results ICCB	**PE**	**CI**	**CI**
Parasite-Parasite	0.03	-0.18	0.26
Parasite-Agrochemical	0.07	-0.1	0.26
Parasite-Nutrition	-0.01	-0.21	0.19
Agrochemical-Agrochemical	0.31	0.12	0.5
Agrochemical-Nutrition	0.21	0.02	0.44
Nutrition-Nutrition	-0.51	-0.83	-0.18
Overall	0.16	0.06	0.26

One reason for the lack of reporting at the cage level for mortality data in the literature may be that whilst the OECD guidelines for toxicity assays in honey bees state that replication at the cage level should be used, the details on statistical analysis and reporting of data do not specify that cage should be included in the analysis or that data should be reported with indications of variation at the cage level (OECD, 2017; https://doi.org/10.1787/9789264284081-en). As this OECD methodology is used as a standard protocol to guide the methods of many toxicity studies, it is not surprising that these data were not reported in many survival studies.

*We do agree that if these data were available for the majority of studies it would be pertinent to use them but, as they are not, we believe that this is the best analysis that we can do based on the information available to us. To clarify this caveat to the reader, and hopefully to encourage a change in reporting that would enable future studies to incorporate cage-level variation, we have added to the methods a statement "*Many studies using *A. mellifera* follow the OECD guidelines when designing mortality studies and we suggest that it may be pertinent for the reporting guidelines to be updated to include data on cage level replication in the future*." (lines 413-415)*

The second point that makes me uncomfortable with, is the analysis of mortality -which is central for the conclusion of this work about the effect of pesticide-pesticide interactions. I understand from table S1 that authors used the "Time point where first treatment group were all dead […] or the final

time point if this was not available". This method is not the best to study survival. In fact, such an analysis of mortality would never be accepted nowadays in ecological studies. Everyone is now performing survival analyses, in which survival time is defined as the time that an individual has survived over a given period, and where hazard ratio are the statistics of reference when comparing treatments. Survival analysis will consider the dynamics of survival, and differentiate the effects of a stressor which, for example kill 90% of individuals within 3 days, and the other 10% the remaining 10days, to the other stressor that kill 10% of individual the first 10 days, and the
remaining 90% the last 3 days. Here, one should consider the first stressor that induce a quick drop more hazardous than the second. Currently, the method used by authors do not allow to differentiate the effect of these stressors on survival, and they would basically call the same effect size. I think mortality here should be reanalyzed with survival analyses and hazard ratio. Besides, in modern survival analyses, one can include cage design (or plate design for larvae) as random factors.This is very handy for honeybee research (see this method paper: https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.2041-210X.2011.00099.x).

We agree that the hazard ratio is generally a much more appropriate way to analyse the effect of stressors on survival. However, in the case of our own analysis we believe that it would be inappropriate to use this metric to investigate the interaction types between stressors. Whilst many of the studies that are included in the mortality analysis use a cox-proportional hazard analysis or similar, of the 64 studies included only 7 report the hazard ratio or model estimates and error for the control, individual and crossed stressor needed for a meta-analysis. As the purpose of a meta- analysis is to collate data already reported within the literature rather than to re-analyse data using statistical models to feed into the meta-analysis, an analysis of hazard-ratios is beyond the scope of our study. Furthermore, and with respect, we disagree that our choice to analyse mortality at a fixed timepoint prevents us from analysing for interaction effects. Whilst there may be differences in the rate at which bees die over the duration of the experiment, and thus their temporal hazard profile, what we are interested in is the overall comparison of the single stressors to the combined stressors. Therefore, at any one time point our null hypothesis (additive interaction) is that the proportion deadin the crossed group will be a function of the individual stressors. Therefore, it is possible to analyse the interaction at any timepoint. The reason we chose the timepoint at which the first treatment group all dies rather than the end of the study is because timepoints later than this will have decreasing sensitivity to detect synergistic interactions. As outlined in the revised manuscript, where the predicted mortality for the additive treatment was greater than 100% studies were excluded, but where the studies were included the additive prediction could still be very close to 100% leaving very little room to detect synergism. Time points earlier than this may have not left long enough for the treatments to take effect. Consequently, we believe that our methodology is the most appropriate forthis dataset.

References

Higgins, J.P., Deeks, J.J. & Altman, D.G. (2008). Special topics in statistics. In: *Cochrane Handbook forSystematic Reviews of Interventions*.

Noble, D.W.A., Lagisz, M., O'dea, R.E. & Nakagawa, S. (2017). Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses. *Mol. Ecol.*, 26, 2410–2425.

Reviewer Reports on the Second Revision:

Referee #5:
Remarks to the Author:
I have gone through all the response to reviewers' comments. I think authors provided written detailed and sound responses and I can see their argument about (i) analysing single bees and not groups/cages and (ii) not re-computing hazard ratio as it is indeed not the purpose of a meta-analysis. After explaining why they did analysed mortality at a given time point, I do agree this was probably the best ways to perform this meta-analysis. Afterall, authors had to do with the heterogeneity of data, which includes how papers reports stats. I am happy with their response to the cage design issue. I do understand the difficulty in running such analysis and I do praise authors for considering my comments and adding a sensitivity analysis to validate their sample size. I think this will show the authors' careful approach and supports once more the robustness of the result. In conclusion, I think this paper is extremely solid. I want to express my congratulations for authors for this great piece of work.