Letter to Editor

Effects of neonicotinoids on bees: an invalid experiment

R. A. Bailey[1] & J. J. D. Greenwood[2]

Introduction

The replication of field experiments and monitoring studies on the potential effect of neonicotinoid pesticides on bees is costly and difficult. Unfortunately, the advice of the European and Mediterranean Plant Protection Organization (2010) on the need for replication and the dangers of pseudoreplication is ambiguous and vague. The European Food Safety Agency (2013) is clearer but has been prominently criticised by Bakker (2016) and further doubt seems to have been raised by the paper of Davies and Gray (2015), although that was concerned with *spurious* accusations of pseudoreplication. Our purpose here is to reinforce the advice of OEPP/EPPO (2010): that strong inference is impossible if there is no replication and that analyses based on pseudoreplication are invalid.

As a concrete example, we use a recent study described by Heimbach et al. (2016). For convenience, we refer to this below as MWP, since it was carried out in Mecklenburg-West Pomerania. Details of the procedures and results are given by Heimbach et al. (2016) and by a set of papers in the same issue of Ecotoxicology (Peters et al. 2016; Rolke et al. 2016a, b; Sterk et al. 2016). In outline they were as follows. Two study sites were used, chosen to be similar in terms of ecology and agriculture, so that "differences were limited to an absolute minimum achievable under field conditions" (Heimbach et al.). In the Test site (T), the oilseed rape seeds sown were treated with clothianidin. In the Reference site (R), no clothianidin was used. Farmers were allowed to decide on all other agricultural activities, including use of other insecticides at the Reference site. Eight honey bee hives and 10 bumble bee hives were placed at each of six locations in each of the two sites; at six other locations, three nesting shelters, each with eight nesting blocks, were set up for mason bees. For each bee species, half of the study locations at each site were at the edge of OSR fields and the other half at approximately fixed distances from the nearest OSR field. Various characteristics of bee performance were measured in each colony (hive or nesting block), except in one in each group of ten bumble bee colonies.

Some characteristics were analysed using simple linear mixed models (LMMs); others were analysed using generalized linear mixed models (GLMMs) or generalized additive mixed models (GAMMs). No matter which of these is used, the model needs to include a term for the effect of study site (the large area chosen), a term for the effect of treatment (clothianidin used or not), a term for study location, a term for distance from OSR (near or far), and a term for colony (which gives the lowest level of variability, and which is included in most statistical software by default). Since there were only two study sites, each with a different

1 School of Mathematics and Statistics, University of St Andrews, Fife KY16 8HT, Scotland.
e-mail: rab24@st-andrews.ac.uk
2 Centre for Research into Ecological and Environmental Modelling, University of St Andrews, Fife KY16 9LZ, Scotland

treatment, it is impossible to differentiate between the effect of site and the effect of treatment. Thus, for comparing treatments, there was no true replication: the study locations within the study sites, and the colonies within study locations, were pseudoreplicates.

It is interesting to see that the report of the honey-bee study (Rolke et al. 2016a) gives the effect of "study site", while the reports of the bumble-bee study (Sterk et al. 2016) and the red mason bee study (Peters et al. 2016) give the effect of "treatment", which clearly shows the confounding between these two effects. Indeed, Rolke et al. (2016a) slip over seamlessly from referring to the effects of study site to referring to the effects of clothianidin, as though these were the same thing rather than two different things that the pseudoreplicated study could not separate.

Contrast this with the effect of distance from OSR. The estimated effect of treatment needs to be compared with variability between study sites, which this experiment gives us no way of measuring. On the other hand, the estimated effect of distance from OSR needs to be compared with variability between study locations within study sites, and there are nine degrees of freedom for estimating this. Thus, for finding out about the effect of distance, the study sites and study locations within study sites provided true replication but the hives within study locations provide only pseudoreplication.

(We note that it would have been possible to include a term, in each model, for the interaction between treatment and distance from OSR, to be assessed against the variability between study locations within study sites. The existence of such an interaction would have been evidence for an effect of clothianidin.)

We do not suggest that the effort put into using several study locations per study site and several colonies per study location was wasted. These pseudoreplications certainly decreased the variance of the estimator of the treatment effect; the problem is that there is no way of estimating that variance, and hence no way of performing a valid hypothesis test and obtaining a valid p-value.

The need for replication and the avoidance of pseudoreplication

Fisher (1926) established that in order to conduct a valid test of statistical significance of any difference caused by treatment one needed a valid estimate of the error of the estimated difference, which can be obtained only by replicating the experiment over more locations than a single control unit and a single treated unit. It was a principle reiterated in his own book *Design of Experiments* (first published in 1935 and appearing in a further eight editions up to 1971 and as part of a compendium in 1990). In addition, the insights of Hurlbert (1984) into pseudoreplication made great impact in the ecological community and have since been further discussed in many publications by Hurlbert and others (e.g. Hurlbert 2004, 2009; Oksanen 2004; Hurlbert & Lombardi 2016). The issues were covered two decades ago in Johnstone's (1998) text on agricultural and ecological experiments and in two publications under the aegis of the Society of Environmental Toxicology and Chemistry (Sparks et al. 1997; Wiens 1996); and they have been fully explained in recent books aimed at biologists or at statisticians who advise scientists (Bailey 2008; Crawley 2015; Gibson 2014; Grafen and Hails 2002; Ruxton and Colegrave 2016). The European Food Safety Authority (2013) Guidance Document on the risk assessment of plant protection products on bees does not explicitly argue that replication is essential but, in its repeated references to replication, takes

it for granted. In the context of other regulatory testing in field studies, Candolfi at al. (2000) emphasized the importance of replication.

Arguments to the contrary

Hurlbert's exceptions
Citing the Hubbard Brook watershed study and Schindler's experiments on Canadian lakes, Hurlbert (1984) allowed that some cases where replication is impossible can deliver useful results. But these cases were characterized by there being "before and after" data and by the informal knowledge that the effects were seen only in the experimental locations, not in the broader landscape. Furthermore, the expected results were of such magnitude that it was implausible to suppose that they were the result of anything other than the experimental treatment.

European and Mediterranean Plant Protection Organization
OEPP/EPPO (2010) advice on the conduct of trials for the evaluation of side-effects of plant protection products on honeybees recognized that it may be impracticable to replicate trials but stated that if there are no replicates "statistical analysis may not be feasible". This suggests that EPPO considers that there are some cases where statistical analysis is possible. However, the advice contained no suggestion as to when this might be and continued "It should also be remembered that individual hives are not replicates", suggesting that EPPO does not believe that it is ever valid to use inferential statistics to compare unreplicated treatments.

Monitoring studies
The EU & SETAC conference *The Effects of Pesticides in the Field* (Liess et al. 2005, Section 2.1) distinguished between two sorts of field studies:

Experimental study: an experiment into the impact of a specific product or active substance applied under controlled conditions in the field. Such studies are performed in the natural environment within an agricultural context.

Monitoring study: an investigation into the overall impact of pesticide use on a specific ecosystem through surveying or monitoring that consists of characterisation of exposure (chemical monitoring, exposure modelling) and observations of effects (biological monitoring) occurring in the field or treated area as a consequence of use and/or misuse of pesticides.

This led to MWP being labelled by Heimbach and his colleagues as a monitoring study. Unfortunately, the report by Liess et al. (2005) did not clearly explain the designs of studies that should be called monitoring studies; nor did it clearly show how inferences could be drawn from them. More clarity was provided Bakker (2016), in his paper critical of the EFSA (2013) guidelines (especially those related to the need for replication): "the field study is the final step in a series of experiments designed to demonstrate the potential impact of a test item on honeybee health... It is, therefore, by nature an experimentation and not a monitoring exercise."

In fact it does not matter whether a study such as MWP is considered to be an experiment or monitoring because, just as in an experiment, statistically sound inferences can be drawn from monitoring only if there is some form of replication. Three examples illustrate this, as follows.

1. Systematic incident investigations, such as the UK Wildlife Incident Investigation Scheme (Liess et al. 2005), where apparent effects following particular potentially damaging events (such as use of a pesticide) are investigated. One such incident can do no more than raise an alarm and stimulate further study. However, the accumulation of repeated similar incidents may allow more robust conclusions about the impact of, for example, the pesticide.

2. Liess et al. also quoted the Sussex Study of Grey Partridges (*Perdix perdix*), an intensive ecological investigation undertaken that extended over 40 years on an area of over 6000ha. This large scale allowed both temporal and spatial comparisons – comparisons that were extended further, both with the general landscape of southern England and with a number of smaller intensive studies. It was possible to model the population ecology of the birds and in some areas to undertake interventions based on that model, the results of the interventions providing further and firmer knowledge of the effects of habitat management, especially the impact of herbicides (Potts 2012).

3. Large-scale and long-term monitoring of wildlife numbers and potentially damaging environmental variables. For example, the decline of British Peregrine (*Falco peregrinus*) populations was both spatially and temporally correlated with the use of cyclodiene seed-dressings (Greenwood & Crick 2015; Newton 2015), widespread declines of British farmland birds (Fuller et al. 1995) were temporally correlated with the intensification of agriculture (Chamberlain et al. 2000), and declines of insectivorous birds in the Netherlands were spatially correlated with neonicotinoid concentrations in surface waters (Hallman et al. 2014). Correlation is a less powerful means of establishing causation than experiment but the Peregrine case was confirmed by the recovery of the population following restrictions on cyclodiene use and the farmland bird case by more detailed examinations of the data, by more intensive ecological studies and by experiments.

Bakker on the EFSA guidance

In designing experiments it is crucial to distinguish between the treatments, the experimental units and the observational units (Bailey 2008, Section 1.4). The treatments are what are applied by the experimenter – for the MWP study they are the two conditions of allowing or not allowing neonicotinoid use. The experimental unit is the smallest unit to which a different treatment can be applied – for the MWP study these units were the two study sites (T and R). The observational unit is the smallest unit on which a response is measured – usually the individual colony in the MWP study. In his criticism of the EFSA Guidance, Bakker (2016) confused these clear distinctions, stating "Although the treatment is physically applied to a field [for 'field' substitute 'study site' in the case of MWP], the field is not necessarily the experimental unit in this case. It is the treatment itself. The hive should be considered the independent unit." This is nonsense: in MWP, for example, all the hives near fields sown with treated seed were in one study site, all those near to fields sown with untreated seeds were in the other, so they were not independent.

In case his confused argument is not accepted and one concludes (correctly) that the colonies in a study site are pseudoreplicates, Bakker then advanced the argument that all this means is that the error variance required for a statistical analysis is underestimated, "which may affect the risk of committing Type I errors". In fact, as Fisher (1926) pointed out, the problem is that the error variance required for the test cannot be estimated because it is the variance between experimental units (*sensu* Bailey) that is needed – the variance between observational units is irrelevant. Bakker (2016) continued "we are comfortable with the use

of statistics to analyze data from (pseudo)replicated designs, because we believe it helps to have a formal analysis that quantifies the risk of finding a false positive result". But what pass for formal analyses of pseudoreplicated studies cannot provide a valid estimate of that risk: they are at best useless; worse, they will usually be misleading.

Bakker (2016) argued that the experimental methodology proposed by EFSA (2013) had "logistical consequences, in particular those related to replication and land use … such that field studies are no longer a feasible option for the risk assessment". He suggested that "It may be necessary to explore new lines of thought for the set-up of field studies and to clearly separate experimentation from monitoring." He neither proposed specific new lines of thought nor how monitoring on a large enough scale to allow the spatial or temporal studies that have been so useful in bird monitoring could be mounted any more easily than properly designed experiments.

Can modern statistical methods get round pseudoreplication?

Davies & Gray (2015) urged "Don't let spurious accusations of pseudoreplication limit our ability to learn from natural experiments (and other messy kinds of ecological monitoring)." We support their plea and would even extend it to cases where the accusations are not spurious, for there is no doubt that "natural experiments (and other messy kinds of ecological experiments)" can provide useful information – even if it is only that there something worthy of further investigation (by looking at independent evidence that bears on the matter or by doing an experiment). Unfortunately, they state that "increased computing power means that there are a number of analytical options for dealing with pseudoreplication", which is misleading. The analytical options to which they refer do not "deal with" pseudoreplication; properly used, they involve fitting complex models that provide valid estimates of the effects of variables for which there is true replication and simply bypass variables for which there is no true replication. Indeed, Davies & Gray redeem themselves by stating clearly that "using such approaches will not be possible if there are only single treated and untreated sites".

We note that some of the analyses of the MWP data involved the calculation of Minimum Detectable Differences (Rolke et al. 2016a, b). However, being based on "error variances" that are derived from pseudoreplicated data, they are themselves invalid as measures of the effect of treatment.

Replication adds bias and variability?

"Intuitively more replicates means more precision ... However, for field studies with bees this may not be necessarily true. Increasing the number of test fields will not necessarily reduce or cancel out noise, but may actually introduce bias This is a consequence of the enormous surface area over which the studies must be laid out." (Bakker, 2016). Heimbach et al. produced a similar argument for avoiding replication, saying that it "would have added considerably to the amount of natural variability and, hence, limited the statistical conclusions possible." We disagree with the view that replication, as such, produces bias. It may introduce more variability but this can be dealt with by blocking and randomization, as Fisher (1926) showed and as Heimbach et al. themselves suggest (page 1645, using paired locations): the appropriate data analysis would then use the pairs as blocks and so remove the geographical variability while still giving valid statistical conclusions. In any case, adding to the between-site variation is not an argument for doing no replication at all: if one has only a single pair of study sites (Treated and Reference) one cannot even estimate the underlying between-site variation because it cannot be separated from the differences due to treatment.

The problem of uncontrolled differences between experimental units

Fisher's technical reasoning about the need for replication, the need to be able to estimate the error variance of the estimate of the difference between the experimental units subjected to different treatments, follows from the possibility of there being differences between the units in addition to the differences imposed by the investigator. As Liess et al. (2005) said in reporting the workshop on Effects of Pesticides in the Field, "The importance of confounding parameters can be reduced by careful selection, but they cannot be completely excluded because they are part of the natural system." MWP attempted to reduce the problem in two ways, both based on measuring various environmental and agronomic conditions that might have differed between the two study sites. First, they included these conditions as covariates in their analyses. Second, they examined the magnitudes of differences in conditions. Neither approach, however, takes into account any conditions that were not measured: one can avoid the influence of such unconsidered conditions only by randomization and replication. Furthermore, although Heimbach et al. concluded that "conditions at the reference and test site were largely similar with the exception of the insecticide treatment", their Table 2 shows differences that could be potentially important – for example, arable land comprised 72% of the area of T but only 50% of R, the Core Area of T had 29% more rape than that of R, early varieties comprised 50% of rape drilled in T but only 18% in R, the mean drilling rate of rape in T was 29% higher than in R and the plant density was 25% higher. Clearly, there were other differences between the sites than the clothianidin treatment.

When replication is too difficult or too costly

Heimbach et al. gave two reasons for their failure to replicate. The first was that "The requirements for separation of reference and test conditions limit the possibility for true statistical replication which would be desirable under ideal conditions (Hurlbert 1984) but is hardly feasible for large-scale, resource intensive studies like large honey bee field trials (OEPP/EPPO 2010; Pilling et al. 2013)." What did these references actually say? Hurlbert did not say that true replication was "desirable": he said that it was obligatory except in special circumstances (see above). Equally, while Pilling et al. pointed out that it was difficult to replicate such experiments, they did not say that it was "hardly feasible"; indeed, they themselves replicated their experiments on the effects of neonicotinoids on bees. Furthermore, replication was undertaken in all bar one of the other purely field experiments on bees and neonicotinoids that have been published (Cutler et al. 2014; Henry et al. 2015; Rundlöf et al. 2015; Tsvetkov et al. 2017; Woodcock et al. 2017). (The exception "was not designed as a definitive statistically robust study" (FERA 2013) and was not published in a refereed journal.) However, OEPP/EPPO did, indeed, state that "although very desirable, replication is often not feasible", unfortunately implying that in this case replication could be forsaken.

Discussion

It is clear that studies without replication can tell us little. It may be, however, that the matter under investigation is so important that even an unreplicated study is judged to be better than no study at all. In that case, great care has to be taken in interpreting the results. Most fundamentally, statistical tests that use the pseudoreplicates as replicates are invalid (as we show above) and must not be used. In this situation, as Wiens (1996) remarked "the

alternative of gauging environmental impacts using qualitative, subjective approaches may seem attractive." However, Wiens continued "Human perturbations of the environment are invariably contentious and emotionally charged, however. Our responsibility as scientists is to provide the most objective and rigorous assessment of environmental effects possible, without which decision-makers have little but guesses, emotions and politics to guide them." If one is reduced to making subjective judgements then, to be useful rather than misleading these must be limited and their subjective nature acknowledged. Thus if, on the face of it, there appears to be a large effect of treatment, the hypothesis that there is a large effect is justified, so long as one stresses that this is not an established fact but merely an hypothesis that needs to be tested by a more powerful and better designed experiment or by gathering of independent ancillary data. Without such testing, the alternative hypothesis that the observed difference results from the underlying differences between the two study sites (that is, confounding factors) is equally valid – and this must be made clear in any account of the study.

Similarly, if there is only a small apparent difference between the outcome in the Treated and Reference study sites in an unreplicated study (such as MWP), this cannot be taken to mean that there is little or no effect of treatment. Because an unreplicated experiment produces no estimate of the error variance for the difference, one cannot calculate confidence limits for the apparent effect so there is no way of knowing whether it really is small or whether it might be interesting large. It is important to acknowledge this point when publishing such results because if several such studies are published people are likely (though mistakenly) to conclude that, since none of them has "shown the effect to be statistically significant", there is no effect; one could equally argue that since none of them has shown that the effect is not interestingly large then one may conclude that it really is interestingly large. (Of course, a proper meta-analysis of the results of several such studies can be informative (Davies and Gray 2015), because it can provide confidence limits of the estimate of effect size over all the studies).

It is sometimes suggested that even where an analysis is not formally valid it can help with informal interpretation based on expert judgement. However, "If the data obtained from a study are inadequate for a formal analysis, then they are, as a matter of principle, inadequate for an informal interpretation. Given that the data are inadequate, the conclusions of the experts are determined by subjective criteria and prior assumptions, often unstated" (Schick, et al. 2017).

Acknowledgements

We thank Dr Fred Heimbach and his colleagues for helping us in trying to produce a note that generates light rather than heat.

Compliance with Ethical Standards

Funding: This work was not funded by any grant or company.

Conflict of Interest: Author Bailey declares that she has no conflict of interest. Author Greenwood declares that he has no conflict of interest.

Ethical approval: This article does not contain any studies performed by any of the authors that involved human participants or animals.

References

Bailey RA (2008) Design of comparative experiments. Cambridge University Press, Cambridge.

Bakker F (2016) Design and analysis of field studies with bees: A critical review of the draft EFSA guidance. Integr Environ Assess Manag 12 (3):422–428. doi:10.1002/ieam.1716

Candolfi M, Bigler F, Campbell P, Heimbach U, Schmuck R, Angeli G, Bakker F, Brown K, Carli G, Dinter A, Forti D, Forster R, Gathmann A, Hassan S, Mead-Briggs M, Melandri M, Neumann P, Pasqualini E, Powell W, Reboulet J-N, Romijn K, Sechser B, Thieme Th, Ufer A, Vergnet Ch, Vogt H (2000) Principles for regulatory testing and interpretation of semi-field and field studies with non-target arthropods. J Pest Sci 73:141–147.

Chamberlain DE, Fuller RJ, Bunce RGH, Duckworth JC, Shrubb M (2000) Changes in the abundance of farmland birds in relation to the timing of agricultural intensification in England and Wales. J Appl Ecol 37 (5):771–788. doi:10.1046/j.1365-2664.2000.00548.x

Crawley MJ (2015) Statistics: An introduction using R, 2nd edn. John Wiley & Sons, Chichester.

Cutler GC, Scott-Dupree CD, Sultan M, McFarlane AD, Brewer L (2014) A large-scale field study examining effects of exposure to clothianidin seed-treated canola on honey bee colony health, development, and overwintering success. PeerJ 2:e652; doi:10.7717/peerj.652

Davies GM, Gray A (2015) Don't let spurious accusations of pseudoreplication limit our ability to learn from natural experiments (and other messy kinds of ecological monitoring). Ecology and Evolution 5(22):5295–5304. doi:10.1002/ece3.1782

European and Mediterranean Plant Protection Organization (2010) EPPO Standards – Efficacy evaluation of plant protection products – PP 1/170(4) Side-effects on honeybees. OEPP/EPPO Bulletin 40:313–319.

European Food Safety Authority (2013) Evaluation of the FERA study on bumble bees and consideration of its potential impact on the EFSA conclusions on neonicotinoids. EFSA J 11(6):3242, 20 pp., doi:10.2903/j.efsa.2013.3242

Fisher RA (1926) The arrangement of field experiments. Journal of the Ministry of Agriculture of Great Britain 33:503–513.

Fisher RA (1935) Design of experiments, 1st edn. Edinburgh, Oliver & Boyd.

Fuller RJ, Gregory RD, Gibbons DW, Marchant JH, Wilson JD, Baillie SR, Carter N (1995) Population declines and range contractions among lowland farmland birds in Britain. Conservation Biology 9(6):1425–1441.

Gibson DJ (2014) Methods in comparative plant population ecology, 2nd edn. Oxford University Press, Oxford.

Grafen A, Hails R (2002) Modern statistics for the life sciences. Oxford University Press, Oxford.

Greenwood JJD, Crick HQP (2015) "It seemed like a dream come true" – Derek Ratcliffe and the peregrine surveys. In: Thomson DBA, Birks HH, Birks, JB (eds) Nature's conscience: The life and legacy of Derek Ratcliffe. Langford Press, King's Lynn, pp 255–279.

Hallmann CA, Foppen RPB, van Turnhout CAM, de Kroon H, Jongejans E (2014) Declines in insectivorous birds are associated with high neonicotinoid concentrations. Nature 511 (7509):341–343. doi:10.1038/nature13531

Heimbach F, Russ A, Schimmer M, Born K (2016) Large-scale monitoring of effects of clothianidin dressed oilseed rape seeds on pollinating insects in Northern Germany: implementation of the monitoring project and its representativeness. Ecotoxicology 25:1630–1647. doi:10.1007/s10646-016-1724-9

Henry M, Cerrutti N, Aupinel P, Decourtye A, Gayrard M, Odoux J-F, Pissard A, Rüger C, Bretagnolle V (2015) Reconciling laboratory and field assessments of neonicotinoid toxicity to honeybees. P Roy Soc B-Biol Sci 282: 20152110. doi:10.1098/rspb.2015.2110

Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. Ecol Monogr 54:187–211. doi:10.2307/1942661

Hurlbert SH (2004) On misinterpretations of pseudoreplication and related matters: a reply to Oksanen. Oikos 104:591–597. doi:10.1111/j.0030-1299.2004.12752.x

Hurlbert SH (2009) The ancient black art and transdisciplinary extent of pseudoreplication. J Comp Psychol 123:434–443. doi:10.1037/a0016221

Hurlbert SH, Lombardi CM (2016) Pseudoreplication, one-tailed tests, neofisherianism, multiple comparisons, and pseudofactorialism. Integr Environ Assess Manag 12: 196–7. doi:10.1002/ieam.1730

Johnstone P (1998) Planning and managing agricultural and ecological experiments. Stanley Thornes, Cheltenham (republished by Routledge, 2013).

Liess M, Brown C, Dohmen GP, Duquesne S, Hart A, Heimbach F, Kreuger J, Lagadic L, Maund S, Reinert W, Streloke M, Tarazona JV (2005) Effects of pesticides in the field. EU & SETAC Workshop, SETAC, Le Croisic, France.

Newton I (2015) Pesticides and birds of prey – the breakthrough. In: Thomson DBA, Birks HH, Birks, JB (eds) Nature's conscience: The life and legacy of Derek Ratcliffe. Langford Press, King's Lynn, pp 281–299.

Oksanen L (2004) The devil lies in details: reply to Stuart Hurlbert. Oikos 104:598–605. doi: 10.1111/j.0030-1299.2004.13266.x

Peters B, Gao Z, Zumkier U (2016) Large-scale monitoring of effects of clothianidin-dressed oilseed rape seeds on pollinating insects in Northern Germany: effects on red mason bees (*Osmia bicornis*). Ecotoxicology 25:1679–1690. doi:10.1007/s10646-016-1729-4

Pilling E, Campbell P, Coulson M, Ruddle N, Tornier I (2013) A four-year field program investigating long-term effects of repeated exposure of honey bee colonies to flowering crops treated with thiamethoxam. PLoS ONE 8:e77193. doi:10.1371/journal. pone.0077193

Potts GR (2012) Partridges. Harper Collins, London.

Rolke D, Fuchs S, Grünewald B, Gao Z, Blenau W (2016a) Large-scale monitoring of effects of clothianidin-dressed oilseed rape seeds on pollinating insects in Northern Germany: effects on honey bees (*Apis mellifera*). Ecotoxicology 25:1648–1665. doi:10.1007/s10646-016-1725-8

Rolke D, Persigehl M, Peters B, Sterk G, Blenau W (2016b) Large-scale monitoring of effects of clothianidin-dressed oilseed rape seeds on pollinating insects in northern Germany: residues of clothianidin in pollen, nectar and honey. Ecotoxicology 25:1691–1701. doi:10.1007/s10646-016-1723-x

Rundlöf M, Andersson GKS, Bommarco R, Fries I, Hederström V, Herbertsson L, Jonsson O, Klatt BK, Pedersen TR, Yourstone J, Smith HG (2015) Seed coating with a neonicotinoid insecticide negatively affects wild bees. Nature 521:77–80. doi:10.1038/nature14420

Ruxton GD, ColegraveN (2016) Experimental design for the life sciences, 4th edn. Oxford University Press, Oxford.

Schick RS, Greenwood JJD, Buckland ST (2017) An experiment on the impact of a neonicotinoid pesticide on honeybees: the value of a formal analysis of the data. Environ Sci Eur 29:4. doi:10.1186/s12302-016-0103-8.

Sparks TH, Bailey RA, Elston DA (1997) Pseudoreplication: Common (mal)practice. SETAC News 17 (3):12–13.

Sterk G, Peters B, Gao Z, Zumkier U (2016) Large-scale monitoring of effects of clothianidin-dressed OSR seeds on pollinating insects in Northern Germany: effects on large earth bumble bees (*Bombus terrestris*). Ecotoxicology 25:1666–1678. doi:10.1007/s10646-016-1730-y

Thompson H, Harrington P, Wilkins S, Pietravalle S, Sweet D, Jones A (2013) Effects of neonicotinoid seed treatments on bumble bee colonies under field conditions. Food and Environment Research Agency, Sand Hutton, York, UK.

Tsvetkov N, Samson-Robert O, Sood K, Patel HS, Malena DA, Gajiwala PH, Maciukiewicz P, Fournier V, Zayed A (2017) Chronic exposure to neonicotinoids reduces honey bee health near corn crops. Science 356:1395–1397. doi:10.1126/science.aam7470

Wiens JA (1996) Coping with variability in environmental impact assessment. In: Baird DJ, Maltby L, Greig-Smith PW, Douben PET (eds) ECOtoxicology: Ecological Dimensions. pp 55–70. Chapman and Hall, London. doi:10.1007/978-94-009-1541-1_6

Woodcock BA, Bullock JM, Shore RF, Heard MS, Pereira MG, Redhead J, Ridding L, Dean H, Sleep D, Henrys P, Peyton J, Hulmes S, Hulmes L, Sárospataki M, Saure C, Edwards M, Genersch E, Knäbe S, Pywell RF (2017) Country-specific effects of neonicotinoid pesticides on honey bees and wild bees. Science 356:1393–1395. doi:10.1126/science.aaa1190