

COMMENT

Open Access



# Are micro-/mesocosm studies really not suitable for the risk assessment of plant protection products? A comment on Reiber *et al.* (2022)

Udo Hommen<sup>1\*</sup>, Marie Brown<sup>2</sup>, Eric Bruns<sup>3</sup>, Klaus Peter Ebke<sup>4</sup>, Ivo Roessink<sup>5</sup>, Tido Strauss<sup>6</sup> and Nadine Taylor<sup>2</sup>

## Abstract

**Background** A recently published article, by Reiber *et al.*, on the representativity of macroinvertebrate communities in outdoor micro- or mesocosm studies, used as a higher tier tool in the environmental risk assessment of plant protection products (PPPs) in the EU, concluded that ‘micro-/mesocosm studies do not represent natural macroinvertebrate communities’. Fundamentally, the article based its conclusion on the analysis of data from 26 streams used in a monitoring project in Germany (2018–2019), in comparison to taxa found in seven lentic micro- and mesocosm studies, conducted at four test sites (2013 – 2018), and submitted to the UBA, Germany.

**Results** There are multiple reasons why this conclusion is incorrect, e.g. the number of taxa, for which the Minimum Detectable Differences (MDDs) were low enough to allow a detection of direct effects in the seven lentic mesocosm studies, cannot be compared to the number of taxa just present in at least five of 26 streams. We have further investigated the data from five of the seven studies which were analysed in detail by Reiber *et al.* and determined that the MDDs of 12 to 18 invertebrate taxa per study fulfilled the current recommendation to allow a detection of medium effects (MDD up to 70%). However, which taxa can be considered potentially sensitive depends on the specific test item. While lentic test systems may not be suitable to test effects on typical stream taxa, taxa occurring in lentic systems such as ponds and ditches are not by definition less sensitive, or vulnerable, to pesticides than taxa living in streams, and their relative sensitivity can be checked in laboratory tests, or artificial streams, if needed.

**Conclusions** In our view, well conducted micro- and mesocosm studies do provide reliable and useful data for the environmental risk assessment of plant protection products covering long-term, as well as indirect, effects under semi-natural conditions.

**Keywords** Microcosm, Mesocosm, Macroinvertebrates, Sensitivity, Vulnerability, Plant protection product, Regulatory acceptable concentration, Risk assessment

\*Correspondence:

Udo Hommen

[udo.hommen@ime.fraunhofer.de](mailto:udo.hommen@ime.fraunhofer.de)

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Background

Micro- and mesocosm studies have been used for many years in the risk assessment of chemicals. The first recommendations on how to conduct these studies were published in the early 1990s [1]. Experimental approaches, and the evaluation and use of the data in risk assessment, have been improved over time, resulting in an OECD guidance document on lentic field tests [2] and recommendations on how to use mesocosm studies in the risk assessment for other types of chemicals, e.g. plant protection products (PPPs) [3], biocides [4], industrial chemicals [5], and for setting Environmental Quality Standards under the Water framework directive [6]. Details in experimental design, and in the use of the data generated for the risk assessment, vary depending on the exposure scenario and specific protection goals involved. Risk assessment for PPPs aims to protect aquatic organisms in edge-of-field water bodies, i.e. 'moderate sized ditches, streams and ponds' [7] in agricultural areas against the effects of, often highly dynamic, exposure. Most micro- and mesocosm studies conducted within regulatory frameworks are conducted for PPPs and the most recent guidance is given in the 'Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field water bodies (Aquatic Guidance Document; AGD [3]).

Reiber et al. [8] imply that most micro-/mesocosm studies do not meet the recommendations in the AGD and that, in their view, micro-/mesocosms are, in general, not suitable for environmental risk assessment because they do not represent natural macroinvertebrate communities. Here, we outline where, and why, we disagree with the conclusions of Reiber et al. [8].

## Results

### Invertebrate communities in micro-/mesocosms and streams are inherently different

Reiber et al. [8] used monitoring data from streams in Germany as the reference for comparison to taxa evaluated in seven outdoor microcosm studies, conducted 2013–2018. These test systems were either enclosures, of 1–2 m<sup>3</sup> which are introduced shortly before the first application into a larger pond or experiment ditch, or isolated small artificial ponds (approximately 5 m<sup>3</sup>) which were set up some months before the start of the experiment. For simplicity, we will use only the term 'microcosms' in the following text, as 45 of the 47 studies listed in Reiber et al. were classified as microcosms. The setup is done with sediment from natural water bodies, e.g. pond, ditches or lakes. Water is taken from reservoirs close by or tap water is used. Colonization of the test systems originates from the sediment, the water (if taken from a reservoir), by flying insects and from samples

taken in natural water bodies of the region. Furthermore, these systems can be seeded with specific macroinvertebrates or macrophytes species that are expected to be sensitive. During the establishment period, the microcosms may be interconnected via tubes, or the water is mixed in an alternative manner, to reduce divergent development before the start of the test. From this setup procedure, the invertebrate community which develops cannot be expected to include typical stream taxa, but will be similar to communities found in ponds or ditches. Since the guidance document [3] explicitly considers edge-of-field water bodies, the general statement that microcosm communities do not represent natural macroinvertebrate communities seems to be overstated. Additionally, microcosm experiments are easier to manage than artificial streams in a flow-through design, and can be considered more protective as a realistic, worst-case, static exposure, rather than the faster dissipation exposure scenario in flowing water systems.

As a comparative dataset, Reiber et al. [8] used 26 streams labelled as having 'good' or 'very good' quality according to the standardized SPEAR index, as given in Liess et al. [9]. Of these sites, 10 are characterized having less than 5% agricultural area in their catchment, corresponding to more than 95% surrounding forest. Such sites are indeed likely to contain a different species composition in comparison to edge-of-field water bodies, and therefore the communities of such streams do not provide an adequate reference for edge-of-field water bodies.

### Understanding species sensitivity in microcosm studies

Some macroinvertebrate orders comprised species that prefer stream habitats, e.g. Ephemeroptera, Plecoptera, Trichoptera (the EPT taxa) and Gammaridae (Amphipoda). Thus, a direct comparison of taxa in lentic test systems and streams cannot be considered meaningful, a fact that was also acknowledged by Reiber et al. [8]. The more pertinent question is whether a sufficient number of sensitive and/or vulnerable taxa can be evaluated in the experiments. For the derivation of a Regulatory Acceptable Concentration under the ecological threshold option (ETO-RAC, only negligible effects on populations are accepted [3]), and the intrinsic sensitivity of the species is relevant. There is no reason to assume a priori that stream species are more sensitive than lentic species. Maltby et al. compared the species sensitivity distribution of lentic and lotic macroinvertebrates, for eight insecticides, and found 'no evidence of a significant difference among or within compounds' [10]. However, since communities in the field can be exposed to multiple chemical stressors, the SPEAR index uses a general sensitivity trait. Von der Ohe & Liess [11] obtained acute toxicity data from the US EPA data base AQUIRE

for organic chemicals, here, the  $LC_{50}$  value from a given taxon was divided by the  $LC_{50}$  value for *D. magna* for the same chemical as a reference value, and finally calculated a mean standardized sensitivity of the logarithm of the different quotients per taxon. Using this method of analysis, Plecoptera and Amphipoda were the only macroinvertebrate orders found to be on average more sensitive than *D. magna*. This analysis included data on insecticides (25.6%), fungicides (4.3%), herbicides (9.8%) and others (60.3%). It is unclear whether the consideration of organic substances in general affects the ranking of sensitivity, if the focus is constrained to effects of insecticides and fungicides. However, taxa with a standardized sensitivity below -0.36, the median sensitivity found in [11], are not considered sensitive within the SPEAR framework ([12, 13]). Consequently all macroinvertebrates which do not belong to Arthropoda, e.g. the isopod *Asellus aquaticus*, are considered as 'not sensitive' within the SPEAR classification [14]. However, depending on the specific test item, these taxa can be highly sensitive in micro-/mesocosm studies [14]. Also, planktonic invertebrates which are not considered in SPEAR can be highly sensitive (see data for Cladocera in [11], also Copepoda or Rotifera), thus, can provide valuable data for the ETO-RAC determination. For example, for pyrethroids, the most sensitive invertebrate species in laboratory tests are *Chaoborus* sp. (planktonic phantom midge), *Asellus aquaticus* (isopod), *Hyalella azteca* (amphipod, not native in the EU) and gammarids [15, 16], yet, in the SPEAR context only gammarids are considered as sensitive and none of these other sensitive taxa are classified as 'species of risk'. Thus, applying such a generic classification of taxa to studies with specific, well-known, test items results in an underestimation of the presence of potentially sensitive taxa analysed in microcosm studies.

Currently, in the Central regulatory zone of Europe, only the ETO-RAC is used in the risk assessment process [17], therefore, the presence of vulnerable taxa, in addition to sensitive taxa, is not generally required for the acceptability of a microcosm study. If short-term effects, followed by recovery of a population, are accepted in the risk management (ecological recovery option; ERO, [3], then vulnerable taxa must also be present, in sufficient numbers in a study, to estimate an ERO-RAC. In addition to intrinsic sensitivity, vulnerability includes the potential for exposure (likelihood of exposure due to habitat preference and organism life-cycle) and the potential for recovery (dependent on species generation time and colonization potential) [18]). The Specific Protection Goals proposed in the AGD [3] for ERO allow 'small effects for a few months, medium effects for weeks, and large effects for days, on the abundance and/or biomass of vulnerable populations of invertebrates, as long as

their reduction does not result in more persistent indirect effects'. As a pragmatic criterion, effects in a micro- or mesocosm study must be restricted to less than 8 weeks to be considered acceptable under ERO [3]. For quickly dissipating substances recovery can usually demonstrated only for planktonic species, including *Chaoborus*, and other insects due to recolonization. Non-flying taxa may have high recolonization potential in the field, but less so in microcosm studies, where the test systems are isolated (e.g. Isopoda, Amphipoda). Thus, these taxa can be representative of other vulnerable species. Recovery of populations with short lifecycles, as is the case for many planktonic species, should not be used to define an ERO-RAC, unless it can be demonstrated the more vulnerable taxa are unlikely to be affected at this concentration.

#### Comparison of the number of macroinvertebrate families present in microcosms and in streams

From Fig. 4, Reiber et al. conclude that 'for all insects and crustaceans, communities at field sites are more family-rich with a 3.6 times higher number of families in the field than families<sub>MDD% low</sub> in M/M studies'. However, two distinctly different types of data were compared. Reiber et al. compared the numbers of families present (independent of their abundance) in at least five of 26 streams, against data from seven microcosm studies, where the number of families were restricted only to those that were considered sufficiently abundant for statistical analysis. Our alternative approach, looking at the mean number of families present at either type of site, determined that 14 crustacean and insect families were found in the streams, which is lower than the mean number of 24 Arthropoda families found the macroinvertebrate samples in the seven microcosm studies (Table 1).

An advantage of microcosm studies, compared to field monitoring, is the reduced variability and the focus on effects at the different, and defined, levels of stressor(s). If the five sites used by Liess et al. [9] as reference streams to normalize the SPEAR index are assumed to be control sites in a hypothetical experiment; in a typical microcosm study design (with five controls and five test concentrations with 3 replicates each), no family would reach the MDD criterion of 70%, while seven families would provide MDD's between 80 and 100% (Gammaridae, Chironomidae, Pediciidae, Simuliidae, Baetidae, Heptagenidae, Limnephilidae). The last three are considered vulnerable according to SPEAR, which means only (medium to) large effects [3] would be statistically detectable based on the communities in the SPEAR reference streams.

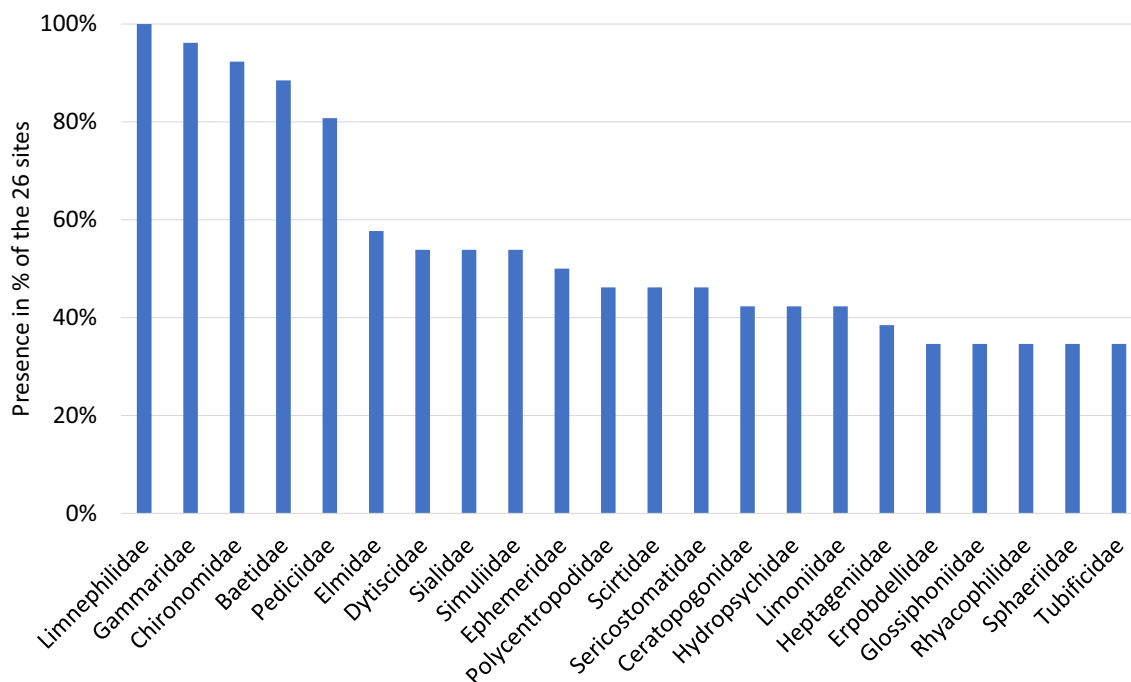
Most EPT taxa prefer streams and cannot be expected to be found in the lentic microcosms. Often the only EPT species that is consistently abundant enough for

**Table 1** Basic characteristics of the seven micro-/mesocosm studies compared to stream monitoring data in [8]

	Study-20	Study-21	Study-22	Study-23	Study-34	Study-35	Study-43
Year of conduct	2014	2013	2014	2018	2013	2017	2013
Location of study site	DE	DE	DE	DE	NL	NL	UK
Water volume [m <sup>3</sup> ]	1.8	1.8	1.8	5.0	1.3	1.5	1.0
Establishment period [month], approx	12	7	18	5	9	>6	3–6
Substance group	Neonicotinoide	Pyrethroid	Dithiocarbamate	Pyrethroid	Botanical	Dithiocarbamate	Juv. hormone mimics
Pesticide type	I	I	F	I	I	F	IGR
n controls	6	5	5	5	5	5	4
n exposure levels	4	5	5	5	5	5	4
n replicates treatments	4	2	2	2	3	3	4
Test units in total	22	15	15	15	18	20	20
MASS/cosm	1	1	1	2	2	2	
other samplers/cosm	1	1	1	3	2	2	2
Netting	Yes	Yes	Yes	No	No	No	No
Emergence traps/cosm	1	1	1	2	1	1	1
<i>Gammarus</i> bioassay	Yes	Yes	No	No	No	No	No
Macroinv. families	28	25	33	29	50	45	37
Arthropoda families	19	16	23	17	35	31	28

statistical evaluation is the mayfly *Cloeon* sp. (Baetidae). Other mayflies or caddisflies may be present as well, but usually not in sufficient numbers for effect assessment. However, with the exception of caddisflies of the family Limnephilidae and the mayflies of Baetidae, other EPT

taxa were also relatively rare in the streams. Additionally, only Ephemeroidea are among the 10 families found in at least every second site and four other caddisfly families and one mayfly family were found in at least every third site (Fig. 1). Stoneflies are restricted to lotic waters and



**Fig. 1** Presence of macroinvertebrate families in the 26 streams considered ‘good’ or ‘very good’ according to SPEAR [9, 19]. Only the families found in at least every third stream are shown

thus, not found in lentic microcosms, However, stonefly families were found only in less than a third of the 26 reference streams. Reiber et al. [8] report 9 times more sensitive and 5 times more vulnerable families, according to the SPEAR framework, in the streams than in the microcosms. On the other hand, studies in lentic systems allow the evaluation of effects on zooplankton, which are not typically present in flowing water, but can include highly sensitive species which would add conservatism for extrapolation to streams. Taylor & Blake [18] compared macroinvertebrates communities in lentic mesocosms of four test sites, and thirty six water bodies, in the UK (14 streams, 10 ditches, 12 ponds) and found that ‘mesocosms are still protective of edge-of-field water bodies, as they are at least as sensitive, and typically more sensitive, than edge-of-field water bodies.’

We believe that the application of the general SPEAR sensitivity, to microcosm studies with specific test items, is not meaningful, and propose an alternative approach where taxa in the mesocosm studies enable an evaluation of direct effects in the next section.

#### Additional evaluations of the selected microcosm studies

Since all seven microcosm studies analysed in more detail by Reiber et al. [8] for Figs. 2, 3, 4 were conducted by the authors of this reply paper, the study reports and data were available to provide additional information on these studies. The basic characteristics of the seven studies are provided in Table 1. Note that three studies (21, 34 and 43) were conducted in 2013 and, consequently, were unlikely to have fully considered the AGD requirements [3] within the study design.

We further investigated five of the seven studies, where we had approval of the owners of the studies, in more detail to identify those invertebrate taxa for which an evaluation of direct effects was possible within the studies. Crucially, our evaluation differs in several aspects from the evaluation of Reiber et al. [8]:

1. We considered all invertebrate species, not just macroinvertebrates. This was done to provide an overview of which taxa could be typically expected to allow an evaluation of direct effects. For the derivation of an ETO-RAC, zooplankton taxa are relevant, and, depending on the test item, not just crustaceans and insects should be considered as highly sensitive. For the derivation of an ERO-RAC, vulnerable species should be assessed. Since many zooplankton species have shorter generation times than macroinvertebrates, and insensitive life-stages (such as resting eggs), they are less vulnerable. However, an ERO-RAC can still be derived, e.g. based on demonstrated recovery of rotifers, as long this ERO-RAC is not

higher than an ETO-RAC derived for vulnerable species. Thus, the taxa to be considered for effect assessment in a mesocosm study should not be restricted to the SPEAR taxa used for monitoring macroinvertebrate communities in streams.

2. Reiber et al. [8] used only the number and variances of the controls and the number of replicates per treatment level to calculate MDDs using the Student's t-test (a 2-sample statistical test). We do not feel this is an appropriate method for testing data from multiple test concentrations against a control. Instead, we have used the MDDs that were calculated in the study reports, following Brock et al. [20], whereby the MDDs for the NOEC are determined by the Williams multiple t-test and utilizes the data from all test units [21].
3. We are aware that low MDDs appearing late in the study, when exposure can be very low, do not indicate that direct effects could have been detected. The relevant period for detecting indirect effects depends on different factors: How fast is the dissipation in the water? Does the substance partition into the sediment and, if so, what is the bioavailability of this compound? What is the life cycle of the species and what endpoint is measured? For example, emergence of insects, even weeks after application of a quickly dissipating substance, can still indicate direct effects since the larvae must have been exposed weeks before. However, for the generic analysis here, only data within 30 days after the (first) application were considered, in an attempt to be comparable to the analysis by Reiber et al. [8].
4. Reiber et al. [8] used the mean MDD, over this 30-day period, in their evaluation. For the analysis here, we have used the minimum MDD, since low MDDs, within 4 weeks after application, should be sufficient to detect a direct effect, especially for macroinvertebrates.
5. According to the AGD [3], MDDs up to 70% are preferable since they allow the detection of ‘medium effects’, although data with higher MDDs can also be useful to detect and evaluate clear direct effects. Nevertheless, for the present analysis, taxa with a minimum MDD < 70%, during the first 30 days after application, were utilized.

Using the criteria outlined above, direct medium effects could be assessed for 8 to 18 invertebrate, and for 6 to 13 arthropod, taxa per study (Table 2). Based on our practical experience, and from studies conducted more recently, we consider that at least 14 invertebrate taxa, including Crustacea (Cladocera, Copepoda, Isopoda and Amphipoda), Insecta (Baetidae, Chaoboridae,

**Table 2** Invertebrate taxa with minimum MDDs < 70% within the first 30 days after (first) application in five microcosm studies

Taxon	Family	Group	MDD <sub>Min</sub> %					Mean	Likely evaluable
			Study-20	Study 21	Study 22	Study-23	Study-43		
Crustacea									
<i>Daphnia</i>	Daphniidae	Cladocera	66		60	50	64	60	X
<i>Chydorus</i>	Chydoridae	Cladocera	69	65				67	
<i>Simocephalus</i>	Daphniidae	Cladocera	65				68	67	
Cyclopodiae	Cyclopodiae	Copepoda	40	37	35	61	29	40	X
Diaptomidae	Diaptomidae	Copepoda		64			11	38	
<i>Asellus sp.</i>	Asellidae	Isopoda	38	46	32	51	21	38	X
<i>Crangonyx</i>	Crangonyctidae	Amphipoda				47		47	X <sup>1)</sup>
<i>Gammarus</i> (assay)	Gammaridae	Amphipoda	12	12				12	
Insecta									
<i>Chaoborus</i>	Chaoboridae	Diptera	42	45	39	42	29	39	X
Chironominae	Chironimidae	Diptera	47	51		54		51	X
Tanytopodinae	Chironimidae	Diptera		69		46		58	X
Orthoclaadiinae	Chironimidae	Diptera			68			68	
Culicidae	Culicidae	Diptera				58		58	
<i>Cloeon</i>	Baetidae	Ephemeroptera	56	58	61	64		60	X
Corixidae	Corixidae	Heteroptera		62				62	
<i>Notonecta</i>	Notonectidae	Heteroptera				53		53	
Anisoptera	Anisoptera n.d	Odonata				50		50	
Zygoptera	Zygoptera n.d	Odonata		49		46		48	X
Coenagrionidae	Coenagrionidae	Odonata	57		45	66		56	
Polycentropodidae	Polycentropodidae	Trichoptera			65			65	
Rotifera									
<i>Polyarthra</i>	Synchaetidae	Rotifera				57		57	
<i>Keratella</i>	Brachionidae	Rotifera	41	54	63			53	X
<i>Trichotria</i>	Trichotriidae	Rotifera		64				64	
<i>Lepadella</i>	Lepadellidae	Rotifera					41	41	
<i>Lecane</i>	Lecanidae	Rotifera					67	67	
Clitellata									
Naididae / Tubificidae	Naididae / Tubificidae	Oligochaeta	66					66	X <sup>2)</sup>
<i>Lumbriculus</i>	Lumbriculidae	Oligochaeta			68			68	
<i>Helobdella</i>	Erpobdellidae	Hirudinida		61	54	67		61	X
Turbellaria						62		62	X
<i>Dugesia</i>	Glossiphoniidae	Turbellaria	59		40			50	
Mollusca									
<i>Lymnaea</i>	Lymnaeidae	Gastropoda	47	48		39		45	X
Sphaeriidae	Sphaeriidae	Bivalvia				55		55	
n_Invertebrates (MDD <sub>Min</sub> < 70%)			14	15	12	18	8	54	
n_Arthropods (MDD <sub>Min</sub> < 70%)			10	11	8	13	6	52	
Proposed potentially sensitive taxa in the specific study			8	11	12	13	6		

1) Gammarids can be tested in in situ bioassays or the amphipod *Crangonyx* can be introduced in cases where gammarids cannot be established in the microcosms. 2) The studies did not include sediment samplings which would provide data for Tubificidae and other sediment organisms with lower MDDs

Chironomidae, Coenagrionidae), Rotifera, Clitellata, Turbellaria, and Mollusca, can be evaluated in lentic test systems. Amphipods are usually not included in most studies, but can be assessed by in situ bioassays with

*Gammarus*, or by introducing *Crangonyx pseudagracilis*, which do not depend on lotic conditions. Mayflies are usually represented by *Cloeon sp.* of the Baetidae family, the most abundant mayfly family also found in the

streams. Trichoptera were sometimes found but could only be statistically evaluated in one study.

The AGD [3] recommends that an evaluation of direct effects for '8 populations of the sensitive taxonomic group' should be possible. In our understanding, this does not mean that for at least 8 taxa effects could be demonstrated but that, according to the mode of action of the test item, at least eight potentially sensitive populations, with sufficiently low MDDs, should be present in a study. For example, in Study-20 where a neonicotinoid was tested, Cladocera should probably not be considered potentially sensitive since, in contrast to other insecticides, *Daphnia* have been found to be significantly less sensitive than insects [23, 24]. On the other hand, oligochaetes of the family of Naididae were found to be potentially sensitive in this study. For fungicides, which often have a broad mode of action, all invertebrates can be considered as potentially sensitive. In Study-21, rotifers were found to be the most sensitive taxon. With the exception of Study-43, at least eight taxa can be considered potentially sensitive to the specific test items in the evaluated studies.

In Study-20, no ERO-RAC could be derived as treatment effects at the concentration higher than the derived ETO-RAC were detected until the end of the study. In Study-22, *Chaoborus* was the clearly most sensitive taxon. At the treatment concentration where *Chaoborus* recovered within eight weeks, other taxa with longer generation times, e.g. *Cloeon* sp., Zygoptera, *Gammarus* and *Asellus*, were not, or only slightly, affected. Thus, the derivation of an ERO-RAC, according to the AGD, is possible. In the fungicide study (Study-22), rotifers were the most sensitive group. At the concentrations where recovery within 8 weeks could be demonstrated, other taxa of different taxonomic groups including crustacea, insects, molluscs, and oligochaetes showed no effects, except for slight temporary effects on copepods. Thus, the derivation of an ERO-RAC is also possible in this study. In the second study with a pyrethroid, *Chaoborus* was again the most sensitive taxon, here, the ERO-RAC was proposed based on the data for the amphipod, *Crangonyx*, which could not recover when affected. In the field, recovery by recolonization may be possible from uncontaminated sections upstream but the missing recolonization potential in the microcosms provides a more conservative, and protective, estimation of an ERO-RAC. Thus, the derivation of an ERO-RAC according to the AGD [3] has to be applied in a study-specific manner.

#### The use of micro- and mesocosm studies in chemical risk assessment

Reiber et al.[8] discuss the use of micro- and mesocosms with respect to statistical aspects, the derivation

of regulatory acceptable concentrations, and their use in a broader regulatory context in the tiered risk assessment of PPPs. They identify the statistical demonstration of treatment-related effects as one of the great challenges. The AGD [3] provides a first proposal on the use of the MDDs, while Brock et al.[20] suggested criteria for acceptable MDDs. Duquesne et al. [22] suggested to consider using a higher power in the MDD calculation, but the proposed method only works for the 2-sample T-test, and not for its multiple variants used in the evaluation of micro- and mesocosm studies, as critical t-values for a beta of 0.2 are not available, to our knowledge. Mair et al. [23] suggest replacing the MDD with the confidence intervals for the effects found at the NOEC, which would require the establishment of new criteria to decide for which taxa direct effects can be assessed. We agree that this is a question that needs further research and related guidance.

Reiber et al. [8], focus on crustaceans and insects 'as they are defined as potentially sensitive taxonomic group for pesticidal substances according to the AGD'. This may apply specifically to insecticides, but not to pesticides in general, since fungicides may affect also other invertebrates [24]. Reiber et al., [8] write that it would be 'advisable for each M/M study to give a justification if (i) especially sensitive taxa towards the mode-of-action of the substance assessed are represented and (ii) their abundances allow for a statistical detection of treatment-related effects'. This is the intention of the AGD ('... not be a reason to reject the study if for several relevant endpoints/populations (e.g. 8 populations of the sensitive taxonomic group) a statistical evaluation can be performed' [3]), and is exactly what has been done in these micro- and mesocosm studies. Here, all available information on the test item is taken into consideration, in contrast to using a generic, taxon-specific, sensitivity such as the SPEAR sensitivity trait. Table 2 provides a list of taxa which could be evaluated in the five microcosm studies conducted between 2013 to 2018, showing that, usually, a statistical evaluation of at least 8 populations of the sensitive group(s) is possible. Reiber et al. [8] question 'if the high effort of M/M studies can be justified when the information related to effect thresholds could also be obtained with far less costly laboratory test systems that usually have a better statistical power, are more targeted towards the toxicant effect and less influenced by complex interactions'. There is no evidence to suggest that laboratory tests investigating delayed, and long-term, effects on several species are, in actuality, less effort and less costly than a mesocosm study, particularly since it is not possible to perform full life cycle tests with many species in the laboratory. In addition, the influence of 'complex interactions' is usually considered as

an argument in favour of micro- and mesocosm studies, since effects are investigated under more realistic conditions, compared to laboratory tests, and incorporate the community context. Reiber's argument that additional stressors are not standardized and quantified and, thus, it is unclear whether the stress level is representative for natural conditions is unfounded since field tests cannot be standardized and the abiotic and biotic conditions can be monitored and described for each study. Indeed, most micro- and mesocosms studies focus on the effects of a single test item under the background of natural stressors (weather, food availability, predation) rather than exposure to chemical mixtures. However, this approach is in line with the current regulatory framework of pesticide risk assessment, which focusses on the effects of individual chemicals, and mixture toxicity was not within the scope of the published article. Nevertheless, effects of mixtures, or typical spray sequences, can be, and have already been, tested in mesocosms studies [25–27]).

Reiber et al. [8] do not recommend the derivation of an ERO-RAC because Chaoboridae and Baetidae would usually be the only vulnerable taxa which can be evaluated in micro- and mesocosm studies using the SPEAR criteria. As shown in Table 2, there are more taxa, e.g. Amphipoda, Isopoda, Zygoptera, Mollusca, and also planktonic taxa, which can affect the derivation of an ERO-RAC. If taxa with a high recovery potential demonstrate recovery, this can still be acceptable if a sufficient variety of other taxa, with a lower recovery potential, show only negligible effects.

In addition to the conclusion that micro- or mesocosms are not representative for natural macroinvertebrate communities in the field, Reiber et al. [8] see problems with micro- and mesocosm studies representing only a specific environmental and exposure scenario. This is true; however, these studies provide a more realistic situation than single species tests in the laboratory, with intentionally constant exposure over time, and the environmental conditions optimized for the test organism employed. Extrapolation from micro-/mesocosms studies to other environmental conditions may be possible in the future by the use of ecosystem models [28]. Reiber et al. [8] also criticize that micro- and mesocosm studies are less conservative than lower tiers. However, that is the basic principle of the tiered approach, where Tier 1 (single species tests) is designed to have a low risk of overlooking critical cases, but, subsequently a high probability to overestimate risks [3]. Furthermore, Reiber et al. [8] state the assessment factor applied to the effect concentrations in a mesocosm study are relatively small (i.e. 2–3 for the ETO-RAC and 3–4 for the ERO-RAC), especially for extrapolation from lentic test systems to lotic water bodies in the field. However, historically, lentic

test systems were used, not only because of easier handling than flow-through artificial streams, but because they represent the worst-case prolonged exposure, situation in lentic water bodies.

#### Improving micro-/mesocosm studies

Despite their general reservations against the use of micro- and mesocosm studies and the current tiered approach, Reiber et al. [8] make some useful suggestions to improve such studies. One recommendation being that study design and sampling techniques should be further improved to increase the statistical power of the studies [20, 22, 23]. This cannot be disputed, however, variability is an inherent property of communities, though lower in micro- and mesocosm studies than between sites in the field and is affected by many other confounding factors. The recommendation of an establishment period of at least two years is not particularly practicable or useful. The number of taxa in the microcosms investigated here was higher than in the streams, and a main point of critique in Reiber et al. [8, 18] is the absence of stoneflies and caddisflies, whose occurrence is determined by the physical type of the test system used, rather than by the duration of the establishment period. Pooling of taxa according to their sensitivity or vulnerability instead of pooling taxonomic groups could be an option to improve the detectability of effects, but the grouping should be study and test item specific, not by a generic approach like SPEAR.

The idea to use shorter studies, focussing on direct effects, seems counter-productive since the purpose of micro- and mesocosm studies is to also assess delayed and indirect effects under semi-field conditions. Thus, longer study durations are not only useful under the ERO, but also the ETO approach.

Finally, Reiber et al. [8] concluded that 'it remains questionable if no unacceptable effects indicated by current higher tier approaches can ensure that no population-relevant unacceptable effects will occur in the field, i.e. if the aim of a more exact and explanatory risk assessment in the current context with complex higher tier approaches should be pursued.' In the AGD [3], mesocosm studies are considered the surrogate reference tier, which can be used to calibrate / validate lower tier approaches. Thus, indicating a high confidence in the protectiveness of such studies. Of course, it has to be evaluated whether this is justified. Since large scale field experiments in natural aquatic water bodies are not feasible, longer-term monitoring studies could provide the data to investigate the level of protectiveness. Any monitoring studies would require an acceptable level of sampling frequency, and a reference site that only differs (significantly) in the pesticide load, this could pose a near impossible task,



particularly as pesticide load is usually correlated with other factors [29]. If the ERA works, and products are used as intended, then there should be only acceptable effects on communities in the field. However, if monitoring were to indicate frequent exceedances of concentrations considered acceptable in the ERA (as shown in [9, 19]), the level of protectiveness of the effect assessment cannot be assessed. Under these circumstances, the first priority would be to clarify the causes for the RAC exceedance.

## Conclusions

In our view, the analysis of Reiber et al. [8] is flawed, and the conclusion that micro-/mesocosm communities do not represent natural macroinvertebrate communities and thus, are not suitable tool for the risk assessment of PPP, cannot be supported by the analysis conducted. Comparing families sufficiently abundant for statistical evaluation in a few lentic microcosms with families present in a larger number of reference streams, independent of their abundance, is biased as it is comparing completely different things. Furthermore, using the general SPEAR sensitivity approach, developed for monitoring streams, is not suitable for defining potentially sensitive taxa in micro-/mesocosm studies with a specific test item, and is not applicable for the detection of previously unknown effects of new substances on organisms without SPEAR classification. Whilst we agree that a critical review of existing studies to improve the ERA of pesticides is beneficial, we believe that micro-/mesocosm, or artificial stream studies, can provide meaningful data for sensitive and vulnerable taxa under the current tiered approach. At present, micro-/mesocosm studies, including artificial stream studies, are the best experimental approach to evaluate the protectiveness of lower aquatic tiers (surrogate reference tier, AGD) and monitoring studies could provide additional useful information. The current tiered approach is also constantly evolving and paradigm shifts are inevitable as new needs arise or new data become available, a good example is given in Topping et al. [30] on comparative lower tier testing and landscape level modelling. However, such new (in silico) ERA approaches also have to be continuously evaluated and validated by field tests (aquatic and terrestrial), on the local scale, and by monitoring data, on the landscape scale.

## Abbreviation

AGD	Aquatic Guidance Document, namely 'Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters' by EFSA [3]
EFSA	European Food Safety Authority
EPT	Macroinvertebrate orders Ephemeroptera, Plecoptera, Trichoptera
ETO-RAC	Regulatory Acceptable Concentration based on the

ERO-RAC	ecological threshold option Regulatory Acceptable Concentration based on the ecological recovery option
IGR	Insect growth regulator
LC50	Lethal Concentration of a substance that will lead to death of 50% of the dosed population during the observation period
MASS	Macroinvertebrate artificial substrate sampler
MDD	Minimum detectable difference
OECD	Organisation for economic co-operation and development
PPR Panel	Panel on plant protection products and their residues
PPP	Plant Protection Products
SPEARpesticides	SPEcies at Risk of pesticides index
SPG	Specific protection goal
UBA	Umweltbundesamt (Environmental Agency)

## Author contributions

UH: writing original draft, evaluation of studies; MB, IR: review and editing, evaluation of studies; NT, EB, KPE, TS, IR: review and editing. All authors read and approved the final manuscript.

## Funding

Open Access Publication Funding by Fraunhofer.

## Availability of data and materials

The microcosm and mesocosm study datasets analysed in the current study are not publicly available. They have been submitted for regulatory risk assessments to the German Environment Agency (UBA). The data access is legally restricted.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

CEA, Fraunhofer IME, gaiac, and WUR conduct micro- and mesocosm studies in the context of chemical regulations on behalf of industry and within public research projects. Bayer conducts mesocosm tests on its own in cooperation with external test sites in the context of aquatic risk assessments for plant protection products. UH, MB, NT, IR, TS and KPE (or their institutions) were involved in the seven recent studies analysed in more detail by Reiber et al. (2022).

### Author details

<sup>1</sup>Fraunhofer IME, Schmallenberg, Germany. <sup>2</sup>Cambridge Environmental Assessments (RSK ADAS Ltd), Boxworth, UK. <sup>3</sup>Bayer AG, Crop Science, Monheim, Germany. <sup>4</sup>Mesocosm GmbH, Institut für Gewässerschutz, Homberg-Ohm, Germany. <sup>5</sup>Wageningen Environmental Research, Wageningen, The Netherlands. <sup>6</sup>Research Institute for Ecosystem Analysis and Assessment (gaiac), Aachen, Germany.

Received: 15 February 2024 Accepted: 19 March 2024

Published online: 29 April 2024

## References

- Hill IR, Heimbach F, Leeuwangh P, Matthiessen P. *Freshwater Field Tests for Hazard Assessment of Chemicals*. Boca Raton: Chapman and Hall/CRC; 1994.
- Organisation for Economic Co-operation and Development (OECD). *Guidance document on simulated freshwater lentic field tests (outdoor microcosms and mesocosms)*. Paris: OECD; 2006.

3. EFSA PPR panel. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFS2 2013. <https://doi.org/10.2903/j.efsa.2013.3290>.
4. European Chemicals Agency (ECHA). Guidance on biocidal products regulation: volume IV environment–assessment and evaluation (Parts B+C); 2017.
5. European Chemicals Agency (ECHA). Guidance on information requirements and chemical safety assessment: Chapter R.10: Characterisation of dose [concentration]-response for environment: ECHA; 2008.
6. European Commission. technical guidance for deriving environmental quality standards: guidance document no. 27 updated version 2018; 2018.
7. FOCUS. FOCUS surface water scenarios in the EU Evaluation process under 91/414/EEC: Report of the FOCUS Working Group on Surface Water Scenarios; 2001.
8. Reiber L, Foit K, Liess M, Karaoglan B, Wogram J, Duquesne S (2022) Close to reality? Micro-/mesocosm communities do not represent natural macroinvertebrate communities. *Environ Sci Eur*. <https://doi.org/10.1186/s12302-022-00643-x>
9. Liess M, Liebmann L, Vormeier P, Weisner O, Altenburger R, Borchardt D et al (2021) Pesticides are the dominant stressors for vulnerable insects in lowland streams. *Water Res* 201:117262. <https://doi.org/10.1016/j.watres.2021.117262>
10. Maltby L, Blake N, Brock TCM, van den Brink PJ (2005) Insecticide species sensitivity distributions: importance of test species selection and relevance to aquatic ecosystems. *Environ Toxicol Chem* 24:379–388. <https://doi.org/10.1897/04-025r.1>
11. von der Ohe PC, Liess M (2004) Relative sensitivity distribution of aquatic invertebrates to organic and metal compounds. *Environ Toxicol Chem* 23:150–156. <https://doi.org/10.1897/02-577>
12. Liess M, von der Ohe PC (2005) Analyzing effects of pesticides on invertebrate communities in streams. *Environ Toxicol Chem* 24:954–965. <https://doi.org/10.1897/03-652.1>
13. Becker J. Effektmonitoring (SPEAR). 2023. <https://www.ufz.de/index.php?de=38122>. Accessed 19 Nov 2023.
14. Roessink I, Crum SJH, Bransen F, van Leeuwen E, van Kerkum F, Koelmans AA, Brock TCM (2006) Impact of triphenyltin acetate in microcosms simulating floodplain lakes. I. Influence of sediment quality. *Ecotoxicology* 15:267–293. <https://doi.org/10.1007/s10646-006-0058-4>
15. Maund SJ, Hamer MJ, Warinton JS, Kedwards TJ (1998) Aquatic ecotoxicology of the pyrethroid insecticide lambda-cyhalothrin: considerations for higher-tier aquatic risk assessment†. *Pestic Sci* 54:408–417. [https://doi.org/10.1002/\(SICI\)1096-9063\(199812\)54:4%3c408::AID-PS843%3e3.0.CO;2-T](https://doi.org/10.1002/(SICI)1096-9063(199812)54:4%3c408::AID-PS843%3e3.0.CO;2-T)
16. Roessink I, Arts GHP, Belgers JDM, Bransen F, Maund SJ, Brock TCM (2005) Effects of lambda-cyhalothrin in two ditch microcosm systems of different trophic status. *Environ Toxicol Chem* 24:1684–1696. <https://doi.org/10.1897/04-130r.1>
17. Zonal Steering Committee Center, Central sekretariat. CZ evaluation manual ecotoxicology—version 2.0 August 2023. 2023. <https://circabc.europa.eu/ui/group/0b40948d-7247-4819-bbf9-ecca3250d893/library/d5076189-cfdc-4d77-b166-187c09e417c9/details>. Accessed 20 Nov 2023.
18. Taylor S, Blake N. PS2339: Assessing the temporal / seasonal changes in the ecology of untreated mesocosms and natural water bodies to inform the uncertainty associated with aquatic risk assessments of plant protection products. 2013. <https://randd.defra.gov.uk/ProjectDetails?ProjectID=15439&FromSearch=Y&Publisher=1&SearchText=PS2339&SortString=ProjectCode&SortOrder=Asc&Paging=10#Description>. Accessed 20 Nov 2023.
19. Liess M, Liebmann L, Vormeier P, Weisner O, Altenburger R, Borchardt D, et al. The lowland stream monitoring dataset (KgM, Kleingewässer-Monitoring) 2018, 2019: PANGAEA; 2021.
20. Brock TCM, Hammers-Wirtz M, Hommen U, Preuss TG, Ratte H-T, Roessink I et al (2015) The minimum detectable difference (MDD) and the interpretation of treatment-related effects of pesticides in experimental ecosystems. *Environ Sci Pollut Res* 22:1160–1174. <https://doi.org/10.1007/s11356-014-3398-2>
21. Williams DA (1972) The comparison of several dose levels with a zero dose control. *Biometrics* 28:519. <https://doi.org/10.2307/2556164>
22. Duquesne S, Alalouni U, Gräff T, Frische T, Pieper S, Egerer S et al (2020) Better define beta-optimizing MDD (minimum detectable difference) when interpreting treatment-related effects of pesticides in semi-field and field studies. *Environ Sci Pollut Res* 27:8814–8821. <https://doi.org/10.1007/s11356-020-07761-0>
23. Mair MM, Kattwinkel M, Jakoby O, Hartig F (2020) The minimum detectable difference (MDD) concept for establishing trust in nonsignificant results: a critical review. *Environ Toxicol Chem* 39:2109–2123. <https://doi.org/10.1002/etc.4847>
24. Maltby L, Brock TCM, van den Brink PJ (2009) Fungicide risk assessment for aquatic ecosystems: importance of interspecific variation, toxic mode of action, and exposure regime. *Environ Sci Technol* 43:7556–7563. <https://doi.org/10.1021/es901461c>
25. van Wijngaarden RPA, Cuppen JGM, Arts GHP, Crum SJH, van den Hoorn MW, van den Brink PJ, Brock TCM (2004) Aquatic risk assessment of a realistic exposure to pesticides used in bulb crops: a microcosm study. *Environ Toxicol Chem* 23:1479–1498. <https://doi.org/10.1897/03-80>
26. Arts GHP, Buijse-Bogdan LL, Belgers JDM, van Rhenen-Kersten CH, van Wijngaarden RPA, Roessink I et al (2006) Ecological impact in ditch mesocosms of simulated spray drift from a crop protection program for potatoes. *Integr Environ Assess Manag* 2:105–125. <https://doi.org/10.1002/ieam.5630020204>
27. Mohr S, Meinecke S, Feibicke M, Duquesne S, Frische T, Sahn R (2023) Effects of a realistic pesticide spraying sequence for apple crop on stream communities in mesocosms: negligible or notable? *Environ Sci Eur* 35:1–20. <https://doi.org/10.1186/s12302-023-00739-y>
28. Loerracher A-K, Schmidt J, Ebke P, Schmolke A, Abi-Akar F, Galic N, Ashauer R (2023) Characterization of patterns and variability in the dynamics of outdoor aquatic mesocosms: exploring the capabilities and challenges in data supporting aquatic system models. *Ecotoxicology* 32:782–801. <https://doi.org/10.1007/s10646-023-02685-0>
29. Jones JJ, Lloyd CEM, Murphy JF, Arnold A, Duerdoth CP, Hawczak A et al (2023) What do macroinvertebrate indices measure? Stressor-specific stream macroinvertebrate indices can be confounded by other stressors. *Freshw Biol* 68:1330–1345. <https://doi.org/10.1111/fwb.14106>
30. Topping CJ, Aldrich A, Berny P (2020) Overhaul environmental risk assessment for pesticides. *Science* 367:360–363. <https://doi.org/10.1126/science.aay1144>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.