An ecosystem services approach to pesticide risk assessment and risk management of non-target terrestrial plants: recommendations from a 2nd SETAC Europe workshop

Authors (in alphabetical order)

Arts, Gertie H.P. †, Eva Kohlschmid§, Lorraine Maltby ||, Christoph Mayer ‡, Giovanna Meregalli#, Véronique Poulsen ††, Franz Streissl !!

† Wageningen Environmental Research (Alterra)
P.O. Box 47
6700 AA Wageningen
The Netherlands

§ Agroscope
Schloss 1, P.O. Box 185,
CH-8820 Wädenswil
Switzerland

|| The University of Sheffield
Department of Animal and Plant Sciences
S10 2TN
United Kingdom

‡ BASF SE
67117 Limburgerhof
Germany

Dow AgroSciences
Environmental Regulatory Sciences, Ecotoxicology
Milano, Italy

†† ANSES - French Agency for Food
Environmental and Occupational Health & Safety
27-31 avenue du Général Leclerc
94701 Maisons-Alfort Cedex
France

‼ European Food Safety Authority (EFSA)
Via Carlo Magno 1A
I - 43100 Parma
Italy

Society of Environmental Toxicology and Chemistry (SETAC), November 2017

This workshop report presents the results of the second workshop about non-target terrestrial plants that was held in Wageningen, The Netherlands, 21 - 22 September 2015. The workshop was held under the auspices of the Society of Environmental Toxicology and Chemistry (SETAC) Europe, with sponsorship from the European Crop Protection Association, the Dutch Ministry of Economic Affairs and participating stakeholder representatives.

Preface

In 2014 a non-target terrestrial plant (NTTP) stakeholder workshop was held under the auspices of the Society of Environmental Toxicology and Chemistry (SETAC) Europe, with sponsorship from the European Crop Protection Association and participating stakeholder representatives. The workshop took place in the Netherlands between the 1st and 3rd of April 2014 and included invited experts from academia, regulatory bodies and business. The aim of the first NTTP workshop was to consolidate current knowledge and expertise to aid the further development of testing and assessment procedures for non-target terrestrial plant. The agreed recommendations of the first NTTP workshop have been published (Arts et al., 2015a) and a workshop report has been finalized (Arts et al., 2015b). Meanwhile, EFSA published a scientific opinion addressing the state of the science on risk assessment of plant protection products for non-target terrestrial plants (EFSA, 2014).

In order to build on the outcome of the first NTTP workshop and the EFSA scientific opinion of NTTP, and to discuss the approach for non-target terrestrial plant risk assessment in greater depth, a second NTTP stakeholder workshop was organized from 21-22 September 2015 in Wageningen, The Netherlands. This second NTTP workshop was organized by SETAC Europe in collaboration with EFSA and participants represented different stakeholder groups and different fields of expertise. The tripartite structure of SETAC was reflected in the number of participants from each stakeholder group: including 7 representatives from Academia, 10 representatives from Regulatory Authorities and 11 representatives from Business (Appendix 1).

This workshop report is been based on discussions held during the 2nd workshop and the underlying reports of Christl (2017a,b). The latter reports were adapted after the 2nd workshop based on actions defined in the recommendations from the workshop agreed by all participants (see page 6-9), i.e. the requirement of a definitive analysis of a full and extended dataset as well as an additional analysis of the EFSA dataset by the same statistical methods as applied to the full dataset. The analysis of the EFSA dataset was considered necessary for comparison of results. The amended reports of Christl (2017a,b) confirmed the results of Christl (2015a,b).

Table of Contents

Conclusions and recommendations agreed by all workshop participants

The following conclusions and recommendations were agreed during the 2nd Non-Target Terrestrial Plant Workshop, held in Wageningen 21 - 22 September 2015. The exact wording was revised based on feedback from workshop participants and a final version of the recommendations was circulated to all participants and approved on March 30, 2016.

1. Ecosystem services

1.1. The workshop participants agreed that "food web support" is an important ecological function which should be protected. The workshop participants did not reach consensus on how to deal with "food web support" in the classification system of the ecosystem services approach.

2. Effects on reproduction

2.1. Preliminary comparative analysis of reproductive and vegetative endpoints indicates that the reproductive endpoints (mainly seed number and seed biomass as measures of seed production) are on average less than a factor of 2 more sensitive than the vegetative endpoints (vegetative vigour, biomass) when comparing the same point estimate (i.e. ER_{10}, ER_{25} or ER_{50} each). This conclusion was independent of whether the analysis was based only on data collated by EFSA (EFSA, 2014) or on an extended dataset containing published and unpublished information (Christl, 2015a)[1].

ACTION: Definitive analysis of full dataset. EFSA will consider this action as part of the preparation of the guidance for NTTP.

2.2. The preliminary analysis described above suggests that reproductive endpoints may be covered by applying an appropriate extrapolation factor to the vegetative vigour endpoints (e.g. ER_{50}, ER_{25} or ER_{10}). However, in cases where reproductive endpoints are expected to be much more sensitive than vegetative endpoints, reproduction studies may be necessary.

---

[1] Based on the recently completed re-analysis on a larger dataset (Christl, 2017b), the factors to extrapolate from vegetative endpoints for young plants to reproductive endpoints, maintaining the same effect level (i.e. ER_{50}, ER_{25}, ER_{10}), were calculated to range from 0.74 to 1.43. Additionally the extrapolation factors from an ER_{50} based on vegetative endpoints in juvenile plants to an ER_{10} or ER_{25} for reproduction ranged from 6.25 to 8.68 and from 2.32 to 3.69, respectively.

ACTION: Further analysis of datasets under investigation, whether it is possible to predict which PPP have a much larger impact on reproduction than on vegetative growth.

2.3. A tiered approach is proposed to cover potential effects on reproduction Tier 1 – vegetative vigour plus an extrapolation factor (unless MoA analysis indicates that reproduction is particularly sensitive). Higher tier – seed production/germination of annuals and bi-annuals.

ACTION: Modelling studies should be conducted to translate reductions in seed production/germination to effects on population size to link the test endpoints to the specific protection goals.

2.4. Surrogate assessment endpoints of the long term impact of PPP on vascular plant reproduction are flowering, seed production and seed germination. Vegetative reproduction is addressed via the biomass tests.

3. Lower and higher tier testing

3.1. Based on an initial analysis of sensitivities of wild versus crop species (Christl, 2015b), testing with standard crop species appears to be protective of wild species.

ACTION: Definitive analysis of dataset to account for further extension of dataset and re-assessment of the statistics behind.

3.2. If a refined effects assessment is required, potential higher tier approaches include:

- single species tests with refined exposure;
- testing additional species (to allow SSD approach) and/or growth stages;
- single species, multispecies or plot experiments (greenhouse, semi-field or field-testing);
- population / community modelling.

Experimental studies include studies in the greenhouse (e.g. potted plants) and outdoor studies at semi-field or field level, e.g. plot experiments

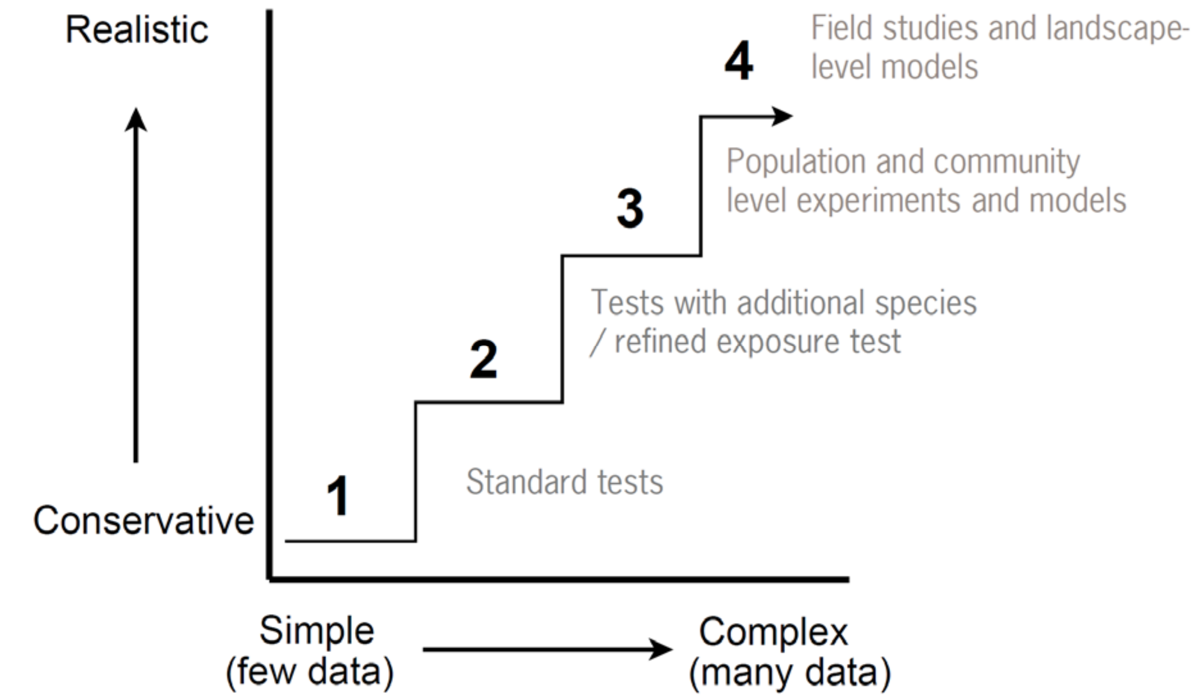
3.3. There is a need for criteria to evaluate and interpret field studies in the context of SPGs.

3.4. There is a need for defining a (surrogate) reference tier in order to calibrate the tiered approach.

3.5. For defining a reference community for an EU level risk assessment, a trait-based approach seems promising in the view of some participants. Traits could be linked to the ecosystem services provided by NTTP.

3.6. The following general tiered approach is appropriate for NTTP effects assesement.

Tiered approach in NTTP effect assessment



4. Mitigation

4.1. Mitigation measures described in MaGPIE are appropriate for non-target terrestrial plants(see the toolbox from MaGPIE). Mitigation measures are considered in the context of the surrounding landscape.

4.2. Risk assessment indicates what proportion of risk reduction is required, but how to achieve this (i.e. implementation) is up to individual MS.

4.3. Vegetated strips need to be managed and the question is what to manage them for e.g. for reducing run-off, to protect annual / perennial NTTP or to provide habitat for other species. Management strategy of vegetated strips should be related to the specific protection goals but need to be regulated in a broader context.

4.4. Management of vegetated strips can be part of a landscape management level strategy for which a broader context of the landscape level is needed.

5. Compensation

5.1. Compensation for in-crop effects should be defined by risk managers in the light of the specific protection goals.

5.2. Several pieces of legislation may be relevant in concert with the pesticide regulation (EC) 1107/2009 when considering compensation (e.g. sustainable use directive; CAP; habitat directive).

SUMMARY

The registration of Plant Protection Products (PPPs) in the EU falls under Regulation 1107/2009, which recommends a tiered approach to assessing the risk to Non-Target Terrestrial Plants (NTTP). However, little information is provided on how to perform and implement higher tier studies or how to use them to refine the risk assessments. Therefore, two SETAC Europe workshops were organized (April 2014, Sept 2015). The participants of the workshops adopted the European Food Safety Authority (EFSA) approach of using an ecosystem services framework for identifying specific protection goals for NTTP. The main scientific questions discussed during the workshops include:

- Is the sensitivity of crop plants protective of wild plants?
- Are vegetative growth endpoints protective of effects on sexual reproduction?
- What options are available for higher tier NTTP testing?
- How to mitigate risks for NTTP?

The first workshop resulted in a list of recommendations agreed to by all participants (Arts et al., 2015). These recommendations related to specific protection goals, risk assessment scheme and mitigation. First, delivery and protection of ecosystem services were discussed for in-crop, in-field *and* off-crop, and off-field areas. Second, lower and higher tier risk assessment methods, including modelling approaches, were evaluated and the benefits from these options were addressed. Third, options for risk mitigation of spray drift and run-off were discussed and evaluated. The workshop participants agreed that the type and relative importance of ecosystem services provided by NTTP differ between areas both in-field and off-field. A number of concerns were raised during the workshop and a need for further data collection and data analyses was expressed by participants in order to reduce uncertainty. This data analysis focussed on the protectiveness of standard test species for wild species and the protectiveness of regulatory endpoints for reproductive endpoints. The second NTTP workshop built on the draft results of data analyses identified by the first NTTP workshop, the recommendations of the first workshop (Arts et al., 2015) and the EFSA opinion on risk assessment of PPPs for NTTP (EFSA, 2014). A preliminary comparative analysis of wild species versus crop species revealed that testing crop species was protective of wild species and additional analyses, performed in response to discussions at the second workshop, have confirmed these initial results (Christl, 2015b). Wild and crop species were found to be similarly sensitive, and there was no trend for either to be more sensitive. The main charge questions identified for the second workshop were:

- How to address reproductive effects in the risk assessment for NTTP?
- How to mitigate risks for NTTP?

- How to conduct higher tier tests (field studies) for NTTP and what options are available?

A preliminary comparative analysis of reproductive and vegetative endpoints was discussed ath the second NTTP workshop. Based on these discussions it was agreed that the data analysis should be extended to incorporate an analysis of the EFSA dataset and to perform further statistical analyses. These additional analyses have confirmed the initial results discussed in the workshop, namely that reproductive endpoints (mainly seed number and seed biomass as measures of seed production) are on average less than a factor of 2 more sensitive than the vegetative endpoints (vegetative vigour, biomass) when comparing the same point estimate (i.e. ER_{10}, ER_{25} or ER_{50} each). Analysis of the EFSA dataset supports this finding. Differences between ER_{50} and ER_{10} are more profound than differences between vegetative and reproductive endpoints. The preliminary analysis available at the workshop indicated that a change from ER_{50} to ER_{10} and from vegetative to reproductive endpoints involved a factor of approximately 10 (Christl, 2015a). Based on the recently completed re-analysis on a larger dataset (Christl, 2017b), the factors to extrapolate from vegetative ER_{50} for young plants to reproductive ER_{10} were calculated to range from 6.25 to 8.68.

A framework for a higher tier risk assessment for NTTP was proposed by workshop participants. Mitigation measures described in MaGPIE were considered appropriate for NTTP. Compensation for in-crop effects should be defined by risk managers in the light of the specific protection goals. Several pieces of legislation may be relevant in concert with the Pesticide Regulation 1107/2009. Further recommendations will be presented.

Chapter 1: Introduction to the workshop

The registration of Plant Protection Products (PPPs) in the EU falls under Regulation 1107/2009, which recommends a tiered approach to assessing the risk to non-target terrestrial plants (NTTP). The tiered approach states that extended laboratory, semi-field and field studies may be conducted. However, little information is provided on how to perform and implement higher tier studies or how to use them to refine the risk assessments.

According to Sanco guidance document 10329/2002, NTTP are non-crop plants located outside the treatment area. However, EFSA 2014 states: "*Furthermore, it is relevant to point out that the new version of the data requirements Commission Regulations (EU) No 283/2013 and No 284/2013 of 1 March 2013, which came into force after the publication of the SPG opinion (EFSA PPR Panel, 2010), no longer provides the following definition: "non-target plants are plants outside the cropped area". Therefore, there is a legal requirement to explicitly also consider non-target plants in the field as well as in the off-field area.*" An official regulatory agreed definition of NTTP is therefore missing. Information provided in Commission Regulations (EU) No 283/2013 and No 284/2013 on exposure conditions and higher tier study designs (e.g. exposure to dust drift, "more realistic exposure conditions", distance of the field), give the distinct impression that the authors of these data requirements regarded NTTP as plants exposed outside of the treated area.

EFSA are in the process of developing guidance to support Regulation 1107/2009 and have produced a scientific opinion on the state of the science on risk assessment of PPPs to NTTP (EFSA, 2014). One of the main objectives of the two NTTP workshops was to consolidate scientific, technical and regulatory expertise as input for the further development of robust, reliable and usable NTTP testing and assessment procedures. In addition, it addressed the following needs:

- Clear guidance on how results from a tiered assessment procedure for non-target terrestrial plants may be interpreted or linked to the relevant ecosystem services.
- Promotion of better understanding among different stakeholders of the state-of the art scientific knowledge relevant to terrestrial non-target plants risk assessment and the sustainable use of PPPs.
- Application of cross-stakeholder understanding to agree on suitable risk management strategies.

The recommendations of the first NTTP workshop were reported in Arts et al 2015. Here we report on the outcome of the second NTTP workshop, which was held under the auspices of the Society of Environmental Toxicology and Chemistry (SETAC) Europe, with sponsorship from the European Crop Protection Association, the Dutch Ministry of Economic Affairs and participating stakeholder representatives. The second NTTP workshop built on the draft results of data analyses identified by the first NTTP workshop, the recommendations of the first workshop (Arts et al., 2015) and the EFSA opinion on risk assessment of PPPs for NTTP (EFSA, 2014). EFSA participated in the Steering Committee and in the workshop. The collaboration between SETAC and EFSA and the involvement of experts from academia, regulatory bodies and business provide an excellent opportunity to capture knowledge and build consensus. The proceedings/recommendations of this workshop will be availablefor possible inclusion in the revision of the EFSANTTP Guidance that is planned after the protection goals have been agreed upon with the European Commission.

The second NTTP workshop was organized around three charge questions:

1. How to address reproductive effects in the risk assessment for NTTP?

2. How to mitigate risks for NTTP?

3. How to conduct higher tier tests (field studies) for NTTP and what options are available?

There were 28 workshop participants from academia (7), business (11) and regulatory communities (10). All participants had a working knowledge of either NTTP research and testing or the regulation of plant protection products in Europe. Background information papers were distributed to all participants prior to the workshop[2] and all Power Point Presentations of keynotes were made available to participants after the workshop.

---

[2] Christl, 2015b (draft report of Christl, 2017a); a presentation of the report Christl (2015a, now available as Christl, 2017b); EFSA, 2014; Arts et al., 2015a; Arts et al., 2015b: draft workshop report of the 1st NTTP workshop; MAgPIE executive summary; Van de Zande, 2015; Krueger, 2015, unpublished report that served as input for the workshop; Abstracts of keynotes, of which the abstract of Poulsen and Van de Zande was added later.

Chapter 2: How to address reproductive effects in the risk assessment for non-target terrestrial plants?

Véronique Poulsen and Giovanna Meregalli

2.1 Introduction

Possible effects on plant reproduction are currently not specifically evaluated in the European terrestrial plant risk assessment framework, which is based on vegetative effects (e.g. fresh weight, shoot length, emergence) assessed in pre- or post-emergence studies. However, it was agreed during the first workshop (Arts et al., 2015) that sublethal effects on non-target terrestrial plant (NTTP) growth and reproduction need to be addressed in order to adequately protect populations of NTTP. Since discussions on reproduction effects play a role with respect to NTTPs in off-field areas and in field margins (Hahn et al., 2014; Schmitz et al., 2013; 2014a,b), it is important to evaluate if the current risk assessment scheme for NTTP is protective of sublethal reproductive effects on terrestrial plants or if any adaptation is necessary, for instance by including reproductive endpoints.

This topic represented one of the main charge questions discussed by the participants in break-out group sessions. An introduction to the topic was provided with the aid of two keynote presentations.

2.2 Abstracts of keynote presentations

2.2.1 Extrapolation from ER_{50}s for vegetative endpoints to ER_{10}s for reproductive endpoints

Robert Luttik (Retired), Beate Strandberg (University of Århus).

Introduction

In 2009 the European Food and Safety Authority asked the Working Group on Non-target terrestrial plant Risk Assessment consisting of Céline Boutin, Michael Klein, Robert Luttik, Mark Rees, Carmen Schweikert, Walter Steurbaut and Beate Strandberg, and EFSA staff members Stephanie Bopp, Franz Streissl, Laura Villamar Bouza and Alessandra Caffi to develop and update the guidance on non-target terrestrial higher plants. In particular it was asked to include effects on biodiversity which were not explicitly addressed in the existing guidance documents at that moment.

The standard endpoint from the plant toxicity tests is an ER_{50} value for vegetative endpoints. Because the aim of the new guidance was to protect plant populations it was advised to use reproductive endpoints. Therefore, in the Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target terrestrial plants (EFSA, 2014) a number of endpoints were proposed for use in risk assessment for non-target terrestrial plants. For some protection goals the endpoint is a 5th percentile of the ER_{10} (effective concentration resulting in a 10% decrease compared with the controls) for reproduction, used as a surrogate for no observable effect rates (NOERs), for others it is the 5th percentile of the ER_{10} for biomass, the 5th percentile of the ER_{50} values for biomass or the 5th percentile of the ER_{50} values for visual endpoints. The data available do not always enable all endpoints to be derived; in such cases, extrapolation between the tested endpoints and the required ones are necessary.

Introducing an ERx and, in this case, an ER_{10} is not completely new because according to the Commission Regulation (EU) No 283/2013, for active substances that exhibit herbicidal or plant growth regulator activity, vegetative vigour and seedling emergence concentration/response tests shall be provided. It is further stated that dose–response tests on a selection of 6 to 10 monocotyledon and dicotyledon plant species representing as many taxonomic groups as possible shall be provided. Finally it is also stated in the Regulation that the ER_{10}, ER_{20} and ER_{50} shall be reported together with the NOER (section 8, Introduction, point 6). The 5th percentile of the species sensitivity distribution for a particular compound is also not new because it was already implemented in the Guidance Document on Terrestrial Ecotoxicology of 2002 (EC, 2002) and is one of the best ways to take the differences in sensitivity between species into account in the risk assessment.

Methods

Materials

Because we were interested in the relation between vegetative endpoints and reproductive endpoints we collected a number of studies where the test time was prolonged to obtain a reproductive endpoint. In total 55 tests were available for nine herbicides (2,4-D, chlorimuron ethyl, glufosinate ammonium, glyphosate, mecoprop, metsulfuron methyl, primisulfuron, sulfometuron and tribenuron) and 34 different plant species were involved (Carpenter et al., 2013; Carpenter and Boutin, 2010; Strandberg et al., 2012; Mathiassen unpublished; Rotchés-Ribalta et al., 2012 and Olszyk et al., 2009). See for more information on the test designs EFSA (2014).

Calculations

In experiments conducted by Carpenter and Boutin (2010) Carpenter et al. (2013) and Rotchés-Ribalta et al. (2012), the ER_{10} and ER_{50} were calculated using non-linear regressions when the data met the assumptions of normality and homogeneity of variance, or else the non-parametric ICPIN method was used (Norberg-King, 1993). Vegetative and reproductive parameters (seed production or measurable equivalent) were used separately in each calculation. In Olsyk et al. (2009), ER_{10} and ER_{50} were recalculated with the raw data provided by the authors using the methodology for calculation as above. Similarly, in

Strandberg et al. (2012) and Mathiassen (unpublished), the ERs were analysed with non-linear regressions using log-logistic dose–response models (Seefeldt et al., 1995). For each herbicide, dose–response curves were estimated for each plant species and growth stage. The fitness of the model was verified using an *F*-test for lack of fit, comparing the residual sum of squares.

Results

When comparing ER_{50} values for a vegetative endpoint with an ER_{50} for a reproductive endpoint 41% of the combinations show a lower vegetative endpoint than that for reproduction. When comparing ER_{10} values for a vegetative endpoint with an ER_{10} for a reproductive endpoint 40% provide a lower vegetative endpoint than that for reproduction and comparing an ER_{50} for a vegetative endpoint with an ER_{10} for a reproductive endpoint, the latter is always lower except for 2 out of 42 combinations (4.8%).

The table below lists factors (EF) to extrapolate from ER_{50} to ER_{10} for vegetative endpoints and from vegetative endpoints (ER_{50} and ER_{10}) to reproductive endpoints (ER_{10}) and the percentages of plant species (combinations) for which the extrapolation factors are not sufficient to prevent effects to happen (i.e. the exposure is still above the rate equivalent to the ER_{10}).

Table 2.1: Extrapolation factors from ER_{50} to ER_{10} for vegetative endpoints and from vegetative endpoints (ER_{50}/ER_{10}) to reproductive endpoints (ER_{10}). The percentages of plant species (combinations) for which the extrapolation factors are not sufficient to prevent effects to happen are also shown.

ER_{50} vegetative endpoint to ER_{10} vegetative endpoint					
EF	10	20	30	34	350
% of plant species not covered	21%	10%	6%	5%	0%
ER_{50} vegetative endpoint to ER_{10} reproductive endpoint					
EF	10	20	30	35	70
% of plant species not covered	19%	10%	8%	5%	0%
ER_{10} vegetative endpoint to ER_{10} reproductive endpoint					
EF	2	2-3	3	5	
% of plant species not covered	6.3%	5%	2.1%	0%	

Remarks

The EFs defined here are associated with some inherent uncertainties owing to the nature of the data used in this exercise. Only a few studies were available and some are carried out by the same authors. It is advisable to redo these calculations when more data are available in future and to pay more attention to the representativeness of the test species and the potential grouping of species (e.g. annual versus perennial species).

Note: a more extensive description of the calculations and information available is provided in EFSA 2014 and in particular in Appendix A.

The second presentation focused on the outcome of the analyses agreed to be conducted as follow up from the first SETAC NTTP workshop comparing the sensitivity of reproductive and vegetative endpoints. The abstract below was made available to the workshop participants before the meeting. Following the second SETAC NTTP workshop, the analysis was completed using a larger dataset and accounting for the recommendations received at the meeting. Results of these final analyses are presented in Christl 2017b. The abstract of this more recent and complete analysis can be found in Appendix 3.

Besides the comparison of the sensitivity of reproductive and vegetative endpoints, an evaluation of the sensitivity of wild and crop plant species has been performed as a follow up from the first SETAC NTTP workshop (Christl, 2015b). This analysis was then revised based on the recommendations received at the second SETAC NTTP workshop and the final results are presented in Christl 2017a. An abstract of this analysis and the results obtained can be found in Appendix 2.

2.2.2 Sensitivity of reproductive and vegetative endpoints

Heino Christl (Tier 3 solutions)

Introduction

Following recommendations of the SETAC workshop on “Non-target terrestrial plants” two literature reviews were performed to compare the sensitivity of endpoint groups i.e. the sensitivity of crop species compared to wild species, and vegetative vigour endpoints and reproductive endpoints of terrestrial plant species. The former tested the hypothesis if wild plant species are more sensitive to plant protection products (PPPs) than crop species (the standard test species), the latter if reproductive endpoints are the more sensitive endpoint than the standard vegetative vigour endpoint available from tier1-tier2 tests. Published literature and unpublished data generated for the registration of PPPs were searched for these reviews, and endpoints from crop species and wild species were compared. An EFSA expert group recently worked on a Scientific Opinion also touching these points (EFSA 2014) that was also considered. In this abstract we focus on the question regarding reproductive vs vegetative vigour endpoints. For the question if there is an intrinsic difference in sensitivity between wild plants and crop species see separate abstract in Appendix 2.

Material and methods

Available Data (focusing on reproductive vs vegetative vigour endpoints)

Formal literature searches were performed, papers with LOER or ERx-endpoints of non-target plant species were assessed, focusing on the availability of reproductive endpoints, and the endpoints included in a data base. Following an initial evaluation of published data in which the test-substance - species combinations were identified for which reproductive endpoints were available, the data sets were complemented utilizing datasets from the species sensitivity evaluations, and EU data (DAR or Review endpoints) were searched for vegetative vigour endpoints of matching test-substance - species combinations. For further details regarding data sources, please refer to Christl 2015b.

Evaluation

Most of the papers generating reproductive endpoints reported also matching vegetative vigour endpoints. In a first step these data pairs were assessed visually (scatter plot) and ratios calculated (dividing the vegetative vigour endpoint by the matching reproductive endpoint), specifying if in a given substance-species- /test combination effects on vegetative vigour were more pronounced (quotients <1) or those on the matching reproductive endpoint (quotients >1). At this stage there were often multiple observations for a given test-substance - species combination. In a second step data were consolidated by generating average (geometric mean) or worst-case (minimum) descriptors for the sensitivity of the two parameters, so that just one quotient per endpoint type (ER_{10}, ER_{25}, ER_{50}) and test-substance - species combination was obtained (each either based on minimum or on geometric mean of all endpoints).

Based on results of the previous assessment (Christl, 2015b)– in which no consistent difference in sensitivity between greenhouse and field data was observed – field and greenhouse data were no longer kept separate. E.g. based on minima the lowest reproductive endpoint was compared with the lowest vegetative vigour endpoint, irrespective of whether it had been generated in a field study or in a lab/greenhouse study. Also the huge variety of different measured variables, from number of inflorescences or flowers via number of pods, number or weight of seeds, up to germination success of the F1-generation, were considered as equivalent[3], and were used both for a geometric mean and for a worst case (minimum) overall measure of toxicity.

[3] We are aware that some of these endpoints are more and others are less relevant, but the decision which endpoints to consider and which not would inevitably include a judgemental aspect and lead to discussions in the course of any potential re-evaluation, so for now we treated them as equals. However, we did excluded some fully anthropocentric endpoints that are not relevant for the population of a non-target plant population, such as ‘colour of cherries’ or ‘marketability of tomato’. More relevant endpoints such as biomass/harvest of fruits were reported in the same studies.

Results

Analysis of the data is ongoing. To date [August 2015] 134 papeüs were assessed. 84 of the ordered papers did not contain reproductive endpoints, 51 did, but from 13 of these no ERx could be derived (e.g. just one level tested). With the initial paired approach, also considering various endpoints of the same test, 148, 1092 and 709 quotients could be calculated based on ER_{10}-, ER_{20}- and ER_{50}-endpoints respectively. With the consolidated approach (minimum endpoints each, or the mean of all reported endpoints of a given substance-species – combination, a total of 46 active-substance–species combinations with reproductive ER_{10}-endpoints, comprising 7 active substances, 108 substance – species combinations with repro ER_{25} encompassing 23 active substances and 150 substance–species combinations with reproductive ER_{50}-endpoints, comprising 27 active substances are available as yet, with numbers expected to further increase.

The resulting ratios – depending on actual parameter settings – are presented in Table 2.2.

Table 2.2: Quotients obtained by dividing minima or mean vegetative vigour ERx endpoint by minima or mean reproductive ERx endpoint per substance/species combination. Censored values were either excluded, or the values (i.e. tested rates) included with a factor of 2 (in line with recommendations of the German UBA).

Effect level	Substance-species - combinations		Ratio VV/Rep	
	Censored values	n	minimum	mean
ER_{10}	*excluded*	23	0.97	1.88
	included with factor of 2	26	0.85	1.56
ER_{25}	*excluded*	56	1.43	1.95
	included with factor of 2	98	1.14	1.44
ER_{50}	*excluded*	71	1.22	1.87
	included with factor of 2	132	1.09	1.48

Based on this initial evaluation and including all data available to date [August 2015], reproductive endpoints were generally somewhat lower than vegetative endpoints, i.e. the former were more sensitive by a factor of 0.9 to 1.9[4]. However there were exceptions, in a few cases vegetative and reproductive endpoints were reported to differ by several orders of magnitude, possible causes of which are still being investigated.

[4] Based on the recently completed re-analysis on a larger dataset (Christl, 2017b), the factors to extrapolate from vegetative endpoints for young plants to reproductive endpoints, maintaining the same effect level (i.e. ER_{50}, ER_{25}, ER_{10}), were calculated to range from 0.74 to 1.43.

An aspect that proved to be important was the age/growth stage of the plant at application, and the subsequent duration of the observation until evaluation. In the initial assessment summarized here generally matching endpoints were compared (as reported in the studies) as effects on vegetative vigour of mature plants were generally assessed together with reproductive effects on mature plants. This approach thus answers the scientifically valid question if the complex physiological processes leading to reproduction are more susceptible to exposure to herbicides than the less complex processes resulting in vegetative growth. This is also the question briefly addressed in EFSA 2014 (Appendix B). However, the regulatory more relevant question is a different one: How does the sensitivity of reproductive endpoints of *mature* plants relate to vegetative vigour endpoints of *young* plants. The latter are the endpoints always available for the risk assessment of substances active on plants, and the current discussion aims to answer the question if a fundamental protection goal is missed when solely vegetative vigour endpoints of young plants are used in the risk assessment, not also reproductive endpoints; these are inevitably measured on mature plants, which normally are much less sensitive than the young plants used in vegetative vigour tests.

To address this question, vegetative vigour endpoints of standard (tier-2) studies will be related to reproductive endpoints, and also to vegetative vigour endpoints of mature plants (data-gathering ongoing in August 2015).

Interim conclusion [August 2015]

Analysis is ongoing, hence this is just an interim conclusion. Based on a comparison of reproductive ER_{10}, ER_{25} and ER_{50} endpoints with corresponding vegetative vigour endpoints of plants of the same age/growth stage/duration of exposure generally reproductive endpoints appeared to be somewhat lower than the corresponding vegetative vigour endpoints, with a few large deviations in both directions. On the other hand standard tier1/tier2 vegetative vigour endpoints are also distinctly lower than the vegetative vigour endpoints generally used for the comparison above. The evaluation is ongoing, further (and hopefully conclusive results) will be presented during the workshop.

2.3 Reproductive effects: Report on workshop discussions (key points from discussion of keynote presentations, breakout groups and plenary)

The questions addressed to participants during the workshop were linked to subjects and concerns raised during the keynote presentations. All discussion points, expressed opinions and questions raised by the participants have therefore been summarized in the following section in order to avoid repetition.

The workshop participants were asked to discuss two charge questions:

- Which reproductive endpoints need to be addressed in NTTP testing and risk assessment in the light of the protection goals for off-crop and in-field, or off-field?
- Can reproductive effects on population level be covered by assessing vegetative effects? If so, what would be an appropriate extrapolation factor?

The participants acknowledged that the preliminary analysis presented by Heino Christl indicates that when vegetative endpoints (e.g. ER_{50} for biomass) derived for young plants (i.e. the endpoints used in the current EU risk assessment scheme) are compared to corresponding reproductive endpoints (i.e. ER_{50} for seed number and seed biomass as measures of seed production) derived for mature plants, only a marginal difference (i.e. a factor ranging between 1.09 to 1.22, Christl, 2015, see section 2.2.2 above[5]) can be found. It should however be noted that such a factor could be within the variability of species sensitivity. Considering this, the workshop participants agreed that studies specifically assessing effects on reproduction of NTTP may not always be necessary to ensure protection of whole plant life-cycle and should not be part of the standard data requirements for all plant protection products. Indeed the basis for the risk assessment can be the pre- and post-emergence studies currently conducted on young plants. Additionally the majority of the participants agreed that if the same endpoint is used in the risk assessment, i.e. ER_{50}, no additional extrapolation factor (EF) was necessary. Other participants felt that an EF may still be required. The available analyses of data-sets can be used to quantify this EF. They show that vegetative and reproductive endpoints differ by a factor less than 2 on average (Christl, 2015a)[6]. An EF may be required if ER_{10} values are extrapolated starting from ER_{50} values. The preliminary analysis presented during Heino Christl's presentation at the workshop indicates that such factors should be approximately 6[7], if focusing on the same parameter (e.g. vegetative ER_{50} to vegetative ER_{10}) and approximately 10 when extrapolating from vegetative ER_{50} to reproductive ER_{10}[8].

[5] Based on the recently completed re-analysis on a larger dataset (Christl, 2017b), the factors to extrapolate from vegetative ER_{50} for young plants to reproductive ER_{50} were calculated to range from 0.97 to 1.38.

[6] This conclusion was confirmed in the recently completed re-analysis on a larger dataset (Christl, 2017b), see Appendix 3.

[7] Based on the recently completed re-analysis on a larger dataset (Christl, 2017b), the factors to extrapolate within the same parameter (i.e. vegetative, reproductive) and different effect levels (e.g. ER_{50} to ER_{25} or ER_{10}) were calculated to range from 4.20 to 5.51.

[8] Based on the recently completed re-analysis on a larger dataset (Christl, 2017b), the factors to extrapolate from vegetative ER_{50} for young plants to reproductive ER_{10} were calculated to range from 6.25 to 8.68.

Several participants had concerns on the appropriateness of an ER_{10} for use in the risk assessment due to the high variability often observed in NTTP studies and to the ecological relevance of such endpoint. Indeed, it is expected that the sensitivities of individuals from the same population may range beyond 10% and so the ecological relevance of an ER_{10} is questionable. It was suggested that if it is determined that ER_{10} endpoints are appropriate for use in the risk assessment and if they are to be derived in the seedling emergence and vegetative vigour studies, the study design may need to be adapted to try to control the variability and increase the statistical power of the test. An alternative approach proposed at the workshop is the use of the ER_{25} endpoint as basis for the terrestrial plant risk assessment[9]. The analyses by Christl, 2017b indicate that an EF ranging from 2.1 to 2.6 may be required to extrapolate from an ER_{50} based on vegetative endpoints in juvenile plants to an ER_{25} for reproduction. However, it is extremely important that the endpoint (e.g. ER_{10}, ER_{25} or ER_{50}) used as basis for the risk assessment is ecologically relevant, so that an adequate level of protection can be ensured.

Studies specifically focusing on reproduction may be required for particular modes of action. Additional analyses are needed to determine the circumstances when the standard studies on young plants are not predictive of possible side effects on reproduction. Information that may be helpful in this analysis may be available in the efficacy data package required to register plant protection products (PPP). However, this will mainly be qualitative information since the methodologies and the level of standardisation used in the efficacy studies is different from the ecotoxicological studies on non-target terrestrial plants.

In cases where reproductive studies could be required, the workshop participants agreed that specific standardised test guidelines would need to be developed and validated to ensure that results are sufficiently robust and that ecologically relevant endpoints are employed in a regulatory context. The parameters relevant to the risk assessment would be an evaluation of flowering or the number of seeds produced. Potentially the subsequent germination of the seed produced could be also evaluated to assess the impact on a second generation and to assess the capacity of the population to produce viable plants. A specific evaluation of vegetative reproduction is not necessary as this was considered a particular growth pattern which is addressed in the standard vegetative vigour studies. Moreover, it was highlighted that measuring the number of seeds for wild species might be difficult. Indeed, the assessment of long-term impacts would imply the need to conduct greenhouse studies, which are more difficult to perform with wild species. Similarly, a long-term reproductive study in the greenhouse may not be practical for bi-annual species. Validated models could aid risk assessors in evaluating the extent of any long-term impacts on populations following possible

[9] The analyses by Christl, 2017b indicate that an EF ranging from 2.32 to 3.69 may be required to extrapolate from an ER_{50} based on vegetative endpoints in juvenile plants to an ER_{25} for reproduction.

effects on reproduction and in determining which effects may be acceptable without impairment of the relevant ecosystem services.

It was also proposed, and the majority of the participants agreed, to include the assessment and testing into a tiered system, starting with vegetative vigour to identify species of concern.

2.4 Conclusions

In conclusion, the following recommendations on the assessment of reproductive effects were agreed by the workshop participants:

- If specific data on the long-term impact (reproductive) of PPP on vascular plant are necessary, surrogate assessment endpoints that should be used are flowering, seed production and seed germination. Vegetative reproduction is addressed via the biomass tests;
- Comparative analysis of reproductive and vegetative endpoints indicates that the reproductive endpoints (mainly seed number and seed biomass as measures of seed production) are on average less than a factor of 2 more sensitive than the vegetative endpoints (vegetative vigour, biomass) when comparing the same point estimate (i.e. ER_{10}, ER_{25} or ER_{50} each)[10]. This conclusion was independent of whether the analysis was based only on data collated by EFSA (EFSA, 2014) or on an extended dataset containing published and unpublished information (Christl, 2015a).
- Based on an analysis of sensitivities of wild versus crop species (Christl, 2015b), testing with standard crop species appears to be protective of wild species, although this result needed to be confirmed in the definitive analyses of the data requested by some of the workshop participants. Such final analysis is presented in Christl, 2017a and indeed confirmed the results of the preliminary analysis. The extended abstract from this work is provided in Appendix 2 of this report.

The following action points were also identified at the workshop:

---

[10] Based on the recently completed re-analysis on a larger dataset (Christl, 2017b), the factors to extrapolate from vegetative endpoints for young plants to reproductive endpoints, maintaining the same effect level (i.e. ER_{50}, ER_{25}, ER_{10}), were calculated to range from 0.74 to 1.43. Additionally the extrapolation factors from an ER_{50} based on vegetative endpoints in juvenile plants to an ER_{10} or ER_{25} for reproduction ranged from 6.25 to 8.68 and from 2.32 to 3.69, respectively.

- Perform a definitive analysis of the complete dataset comparing reproductive and vegetative endpoints, including all the relevant publications with vegetative or reproductive endpoints on NTTP that can be identified.

The preliminary analysis described (Christl, 2015a) above suggests that reproductive endpoints may be covered by applying an appropriate extrapolation factor to the vegetative vigour endpoints (e.g. ER_{50}, ER_{25} or ER_{10}). However, in cases where reproductive endpoints are expected to be much more sensitive than vegetative endpoints, reproduction studies may be necessary. A definitive analysis of the complete dataset, re-assessing the statistics behind the analysis is necessary[11]. In this final analysis, in particular, the following aspects should be considered:

- Evaluate whether it is possible to predict which PPP have a much larger impact on reproduction than on vegetative growth.
- Derive extrapolation factors from vegetative to reproductive endpoints.
- Evaluate whether in most of the cases early stage vegetative vigour testing covers also reproductive effects at the same effect level (e.g. ER_{50}, ER_{25} or ER_{10}). The generation of additional data may be proposed to cover potential effects on reproduction for example for specific Modes of Action.

- Develop modelling tools:

Modelling studies should be conducted to translate reductions in seed production or germination to effects on population size and thus link the test endpoints to the specific protection goals.

The definitive analyses of the complete dataset (including all the relevant publications that could be identified) requested at the workshop have been addressed in the report Christl (2017b).

[11] The revised analyses are described in Christl 2017b. The extended abstract of this report is presented in Appendix 3.

Chapter 3: How to conduct higher-tier tests (field studies) for non-target terrestrial plants and what options are available?

Gertie Arts and Franz Streissl

3.1 Introduction

In the new data requirements for Active Substances and Plant Protection Products, screening data with six terrestrial plant species from six different families including mono- and dicotyledons is required (EU 2009 a,b). In the case of herbicidal or plant growth regulatory activity, toxicity tests generating ER_{50} values shall be conducted on a selection of 6 to 10 monocotyledon and dicotyledon plant species (EU 2009 a,b). Routine testing with NTTP is performed in highly standardized greenhouse studies. Extended laboratory studies with more realistic exposures may be conducted if a high risk has been identified in standardized tests. Semi-field and field tests measuring effects on plant abundance and biomass production at different distances from the treated crop can be performed as a further refinement. As there are no standardized tests available for extended laboratory tests or semi-field and field tests, the type and conditions of these studies are be discussed with the national authorities.

There is limited experience and a lack of standardization of multispecies and field studies (Krueger, 2015; unpublished report that served as input for the workshop). Higher-tier studies might vary from semi-field outdoor experiments (outdoor potted plants, field plots) via field experiments (vegetation plots, field strips) to the landscape level, where monitoring and modelling might become useful approaches. There is a need to clarify the role of such studies and when they are appropriate to address a regulatory question or outcome (Krueger, 2015; unpublished report that served as input for the workshop).

How to conduct higher-tier testing for NTTP, was one of the three charge questions of the second non-target terrestrial plant workshop. The breakout groups were asked to discuss higher-tier NTTP testing in order to provide guidance on species, endpoints, test design and modelling in view of the agreed protection goals. The breakout groups were specifically asked to discuss the following questions:

- For higher-tier studies consider which species, endpoints, testing methods, or modelling approaches could be appropriate to address such a situation?
- Do you consider the approaches discussed the previous question to be at an equivalent tier in the risk assessment? Justify your answer.

This chapter provides the abstracts of the keynotes, a reflection of the discussions during the break-out groups and the plenary sessions, a summary and conclusions.

3.2 Abstracts of keynote presentations

3.2.1 Regulatory perspective on what needs to be addressed in NTTP risk assessment

Carmen Schweikert (UBA)

Protection goals and higher tier testing

The legal background of plant protection product regulation as well as the consistency with other guidelines and documents on risk assessment for other non-target organisms need to be taken into account in a regulatory context.

Regulation (EC) No 1107/2009 sets the preconditions for the approval of plant protection products i.e. the absence of unacceptable effects on diversity and abundance of non-target organisms, biodiversity and the ecosystem. Consequently, also the updated data requirements in Regulation (EC) No 283/2013 point out that “the potential impact of the active substance on biodiversity and the ecosystem, including potential indirect effects via alteration of the food web, shall be considered”.

Non-target terrestrial plants (NTTP) are all plants growing outside fields and those growing within fields that are not the intended pesticide target. Terrestrial higher plants provide habitat and food for other groups of organisms (arthropods, birds and mamals), resources for microorganisms and sometimes for water and soil retention. In-field areas are generally much larger than off-field areas in agricultural landscapes. Hence, a risk assessment limited to the off-field areas lacks addressing the direct and indirect risk at landscape level arising from the removal of non-target plants from the in-field areas. Inevitable effects on the abundance of plant species in-field due to targeted pesticide application may have major ecological impact.

Removal of non-target plants may affect i) specific plant-animal-interactions on species level; ii) general use of plants as food and habitat by animals on functional level (biomass, structure); iii) survival of rare arable plant species that depend on cultivation practices; and iv) abundance of NTTP within large fields, which is important for restoring the meta-population of plants facing an increasing fragmentation of their native off-field habitats.

It has been shown in recent years that changes in agricultural management, including pesticide use, has led to loss of biodiversity in agricultural areas. Against this background, availability of food and habitat in the agrosphere is crucial for farmland species.

The ‘Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target terrestrial plants’ (EFSA, 2010; 2014) proposed 4 Specific Protection Goals (SPG) for key non-target terrestrial plants, covering in-field and off–field areas.

- **Specific protection Goals for off-field areas** refer to food web support; aesthetic values; biodiversity; genetic resources; nutrient cycling; water regulation.
- **Specific protection Goals for in-field areas** refer to food web support; aesthetic values; biodiversity/genetic resources; endangered arable species.

The use of field studies to define a reference tier and to calibrate lower tiers in the risk assessment

In the ‘Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target terrestrial plants’ an attempt was made to match the SPGs with NTTP test endpoints in order to address the protection goals, since these proposed measurement endpoints are based on a limited number of single species tested in the laboratory. The protection goal “negligible effects on plant populations at the edge of the field” refers to the field situation. The uncertainties from extrapolating from laboratory derived endpoints with single plant species to field effects need to be addressed with appropriate assessment factors.

In order to establish the assessment factors for lower tier assessments it is necessary to calibrate them with a so-called “Reference Tier”. The “Reference Tier” is defined as the field exposure and effect tier. The “Surrogate Reference Tier” is defined as the best available practicable test system as an approximation to the “Reference Tier” and it could be mirrored in well-performed field studies with diverse plant communities. The main challenge is that no community studies are available in the dossiers at present and therefore such calibration of Tier 1 outcome is not feasible with “Reference tier” (actual ecosystem) or “Surrogate reference tier”.

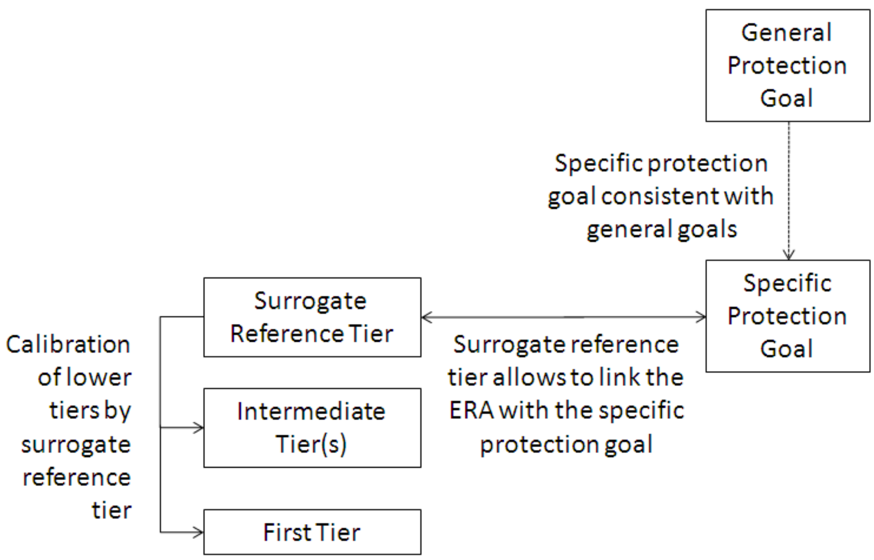


Figure 3.1: Illustration of the relationship between tiers of the risk assessment process and protection goals used by the PPR Panel (EFSA PPR Panel, 2010)

It is apparent that one field study cannot cover all plant communities. A set of different community studies is needed to perform such a calibration. Modelling might help to interpret the results of community studies and to extrapolate study outcomes to other conditions.

The need of multispecies field studies for non-target terrestrial plants also becomes apparent considering the options for a risk assessment beyond Tier 1.

Options to address a risk that was determined at Tier 1 are:

- To provide more field studies on specific plant protection products to reach a safe use in higher Tier. This might lead to an increase in the pool of field studies which might result in generic calibration in the long run.
- To perform generic studies independent from authorisation process of individual plant protection products concerning the ecology of NTTP (e.g. distribution, competition between NTTP).
- To adapt the intended use implying risk mitigation and risk management options.

In consistency with other fields of risk assessment (e.g. aquatic risk assessment) remaining uncertainties have to be addressed also in higher tier risk assessments.

At present, no guidance is given on how to conduct field studies for higher tier assessments of the risk for NTTP exposed to plant protection products. For the choice of sites and design of the studies criteria agreed between Member states are essential, e.g. choice of coenosis actually present in the Member States. Adequate plant communities might be an established off-field coenosis not yet adapted to adjacent agricultural practice (i.e. not exposed to overspray, spray drift, etc.).

How to address in-field/ indirect effects?

The acceptability of effects on biodiversity and the ecosystem could be addressed by defining percentages of food and habitat loss causing non-acceptable impact. This could also be investigated in higher tier studies.

For the protection goal ‘food web support’ driven by NTTP, it can be assumed that there is no recovery potential in- field in relevant time frames after application of broad-band herbicides. Since indirect effects in-field are inevitable when pesticides with specific mode-of-action are applied, prospective risk regulation needs to provide options for risk managers on how to prevent unacceptable effects on biodiversity and the ecosystem.

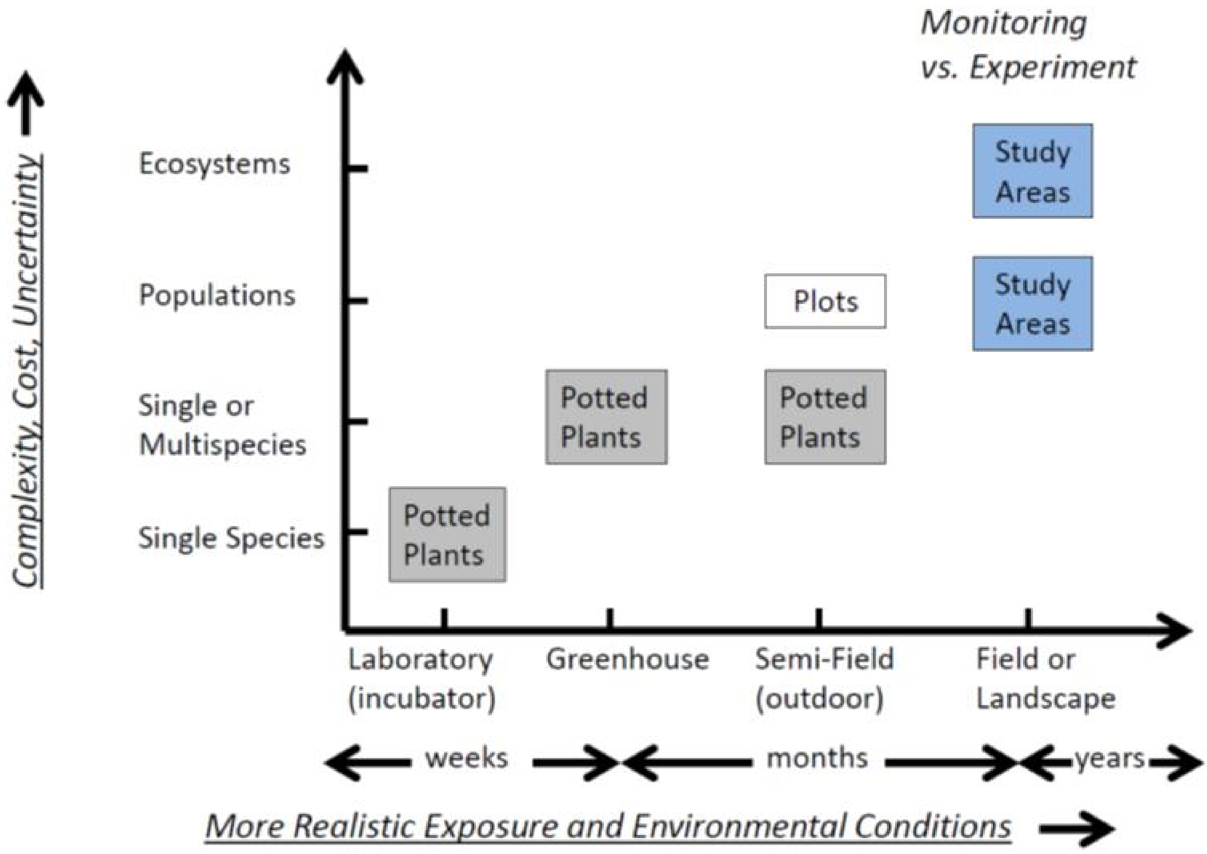
The current management measures aim at reducing exposure in off-field areas (buffer zones, wind breaks, drift-reducing application techniques) to mitigate risk for non-target plants. However this does not mitigate in-field risks and the provision of additional management options is essential to reduce effects on non-target organisms caused by the reduction of non-target plant abundance and diversity within the cropped area. Uncropped areas without pesticide exposure that compensate indirect effects of pesticides on higher trophic levels could be one solution to ensure in-field SPGs. For rare arable weeds, which need disturbance, uncropped but tilled field edges with no herbicide exposure which allow these species to grow, appear to be a suitable management measure.

3.2.2 Higher-tier testing overview: review of multispecies and field testing in assessing risk of chemicals to non-target plants in agricultural landscapes

H.O. Krueger, Wildlife International, Ltd. / Aquatic Plant and Insect Toxicology

A literature search was performed to pull together information and make recommendations on how multispecies studies and field studies can be used in a tiered approach to evaluate risk to non-target terrestrial plants. This project was part of a SETAC endorsed technical workshop on Non-Target Terrestrial Plants with funding by ECPA in April 2014 at Wageningen, NL. Field studies or other multispecies studies are larger scale experiments that maybe used to answer specific questions that arise from guideline laboratory studies or incidents observed on the landscape and reported to companies and regulators. However, there is limited experiences and a lack of standardization of multispecies and field studies which pose several challenges. There is a need to clarify the role of such studies and when they are appropriate to address a regulatory question or outcome. The purpose of this review is to present types of approaches that may be used in field and multispecies studies and describe how such studies may be conducted. Recommendations have been made that may aid in preparation of risk assessment guidelines.

Multispecies studies can refer to multiple plant species or also be more comprehensive and look at interactions among populations of both plants and animals, as well as evaluate trophic level effects in the ecosystem (e.g. aquatic mesocosm studies). One important consideration is whether the study has multiple species with only one species per pot or plot, or if there are multiple species in one pot or plot. The former follows conventional testing, while the latter is directly looking at interaction and competition between or among species. They can be conducted in the lab or greenhouse as well as outdoors in the field. Field studies can be set up to measure effects on existing plant communities or can be more experimental and set up plots of planted species in more traditional randomized plot experimental designs. The latter is usually referred to as a simulated field study. One difficulty of longer term tests in a greenhouse is that there is the problem of plants becoming pot bound which can effect growth and reproduction. The illustration below shows relationships between types of lab and field studies in reference to the experimental units, levels of biological organization, and experimental conditions that trend towards more realistic exposure and environmental conditions.



Krueger (2015) summarizes the results of the literature search as a number of recommendations:

Multispecies testing allows us to analyze effects of plant protection products on NTTP under more realistic conditions with respect to exposure, growth conditions, competition, recovery and indirect effects than single species tests. Experimental options of different complexity and realism are available which have to be selected carefully based on the specific question to be addressed.

Until now there is not much experience and no guidance on how to conduct multispecies tests with NTTP. However, they should be considered as a higher tier option similar as for soil invertebrates, non-target arthropods and aquatic communities.

Studies must address concerns identified from lower tier tests or from incident reports from the field.

There is a need to simulate realistic exposures from drift events in large scale field or monitoring studies or to set concentrations in plot experiments. Current technologies in spray drift application equipment (including drift reduction nozzles) should also be considered in the experimental design.

Experiments need to have a proper statistical experimental design with adequate statistical power to insure that there is high certainty in the validity of the results. Do not expect field studies to produce the precision and accuracy achievable in the lab. Definition of reference sites or controls must be established in conducting field studies. This is important for the interpretation of the results of the study in which patchiness of plants occurring under natural growth conditions need to be taken into account. It is to be preferred that the minimum detectable differences are reported.

Determining effects of a single chemical may be confounded by other chemicals or nutrients (fertilizers) used in the study area. Care must be taken to minimize the effects of these variables in field test maintaining similarity among plots. When conducting large scale field or monitoring studies these are variables that need to be measured.

Higher tier tests may need to be conducted in multiple sites depending on the geographic range in which a chemical is used. Here, modelling of environmental conditions and traits can help.

There is a need to address sub-lethal endpoints of plant growth and reproduction to determine population level effects on multiple species.

There is a need to address how changes in plant population dynamics are affecting ecosystem level endpoints such as biodiversity, nutrient cycling, energy transfer, primary production, and habitat. Alterations to plant community structure may result in adverse effects on organisms at higher trophic levels that represent other key drivers of the ecosystem. However, if there are little or no adverse effects on growth and reproduction on plant populations one might be able to assume that the ecosystem services are also protected.

There is a need to collate and exchange understanding, knowledge and protocols for field studies and multispecies studies, as well as develop new approaches to the problem.

There is no reason to perform multispecies tests without knowing how the results will be used by regulators in their evaluation of risk and whether the results will be used to offset the hazardous effects that may have been identified in lower tier tests.

3.3 Higher-tier testing: report of workshop discussions (keynote presentations, breakout groups and plenaries)

3.3.1 Discussion of keynotes

The direct effects of PPPs on NTTP can be assessed in a wide range of possible tests. A challenge however, is how to address the risks of indirect effects of PPPs that may affect NTTP in the field and in the off-field habitats. For example, the effects of insecticides on pollinators and consequences for the pollination of flowering non-target terrestrial plants. Another challenge is to address indirect effects of pesticide use on biodiversity. For example the indirect effects of plant extinction due to herbicide use on other species were discussed. A wide variety of ecological studies on non-target terrestrial plant communities is available in the open literature, however studies of non-target terrestrial plants linked to pesticide use and applications are limited. There are no methods available for assessing the indirect effects of PPPs on NTTP nor are there any data to underpin the development of a method for extrapolating from single species to community endpoints. When developing such higher-tier methods it is important to ensure that protection goals are still achieved at all levels.

Before conducting higher-tier studies, it is necessary to clearly define the question to be answered. This will then dictate the study design including the vegetation type to be studied (e.g. herbs, shrubs, woodlands), the species to be studied, the geographic location of the studies (i.e. should studies be conducted in each country/ geographical region?), the approach to be used, etc. Workshop participants questioned whether higher-tier studies should consider plant-animal interactions. For example, should the presence of herbivorous species (vertebrates and invertebrates) or pollinators be taken into account?

Higher-tier studies may be performed in glasshouses or in the field and possible endpoints of higher-tier studies are species abundance, species number and biomass (fresh weight and dry weight) per square meter. In greenhouses, tests are conducted in plant pots, which may constrain the study design. The size of the plant plots and the duration of the observations should be defined prior to the study. It was noted that recovery may be observed after weeks or months following initial effects. Field studies are more realistic than glasshouse studies, but are also more variable as more uncontrolled environmental factors influence the test plants. Field studies on NTTP may be difficult to design and interpret as the community may not be known at the beginning of the study. There is therefore a need to discuss the study protocol, including scale, before starting the study in order to define the question and to be sure that the study addresses the key regulatory concerns. Another category of higher-tier plant studies are monitoring studies, which can be conducted over several years covering different seasons. Monitoring studies can also assess woody vegetation, which may be difficult to establish in field or greenhouse studies and can provide information on multiparameter causes of effects. However, monitoring approaches need to be standardised. The areas monitored may vary depending on the landscape and selection of monitoring sites is critical. The number, location and type of sites monitored depend on the focus of the monitoring study.

3.3.2 Breakout groups and plenaries

The data required for NTTP testing are well defined and include screening data, efficacy trials, dose-response information and standardized toxicity testing (COMMISSION REGULATION (EU) No 284/2013 of 1 March 2013, COMMISSION REGULATION (EU) No 283/2013 of 1 March 2013, OECD, 2006a;b; EPA guidelines for testing of terrestrial plants). All information from the Tier 1 and greenhouse studies is needed to set up higher-tier studies. Also efficacy trials might provide information about relative sensitivity. Modified exposure tests could be performed with the sensitive species from Tier 1.

Workshop participants agreed on the need for a tiered approach for risk assessment for NTTP and therefore for the development of higher-tier studies. Breakout groups discussed number of topics in relation to higher-tier studies including: species selection, endpoints, testing methods, and modelling approaches could be appropriate to address such a situation. Workshop participants considered whether species selection should be based on the geographic region (or zone) of interest or based on plants traits or based on plants potentially present in the adjacent off-field area. They agreed that feasibility of testing must be one of the selection criteria. A trait approach was considered a promising approach, with test species selected based on their trait profile. For multispecies testing, the question to be addressed is: what are the correct combinations of plant species to combine in a field test considering the below- and aboveground variability and competition?

Endpoints need to be linked to the protections goals and potential endpoints include biomass, reproduction, number of species, biodiversity indices. Distinction in higher-tier studies between community- and population-level responses will determine the endpoints to measure. Endpoints are also linked to the traits to be considered. When studying vegetation / field strips over time, there might be a preference for non-destructive endpoints: e.g. species composition (see e.g. Kleijn and Snoeijing, 1997; Kleijn and Verbeek, 2000), number of individuals, plant cover (taking the structure of the community into consideration), plant height. However, validation is required if non-destructive endpoints are to be used as a proxy for biomass. Community-level endpoints include trait composition, species composition or biodiversity indices (e.g. evenness or Simpson index). Reproductive endpoints need not necessary to be addressed, but need to be accounted for in the risk assessment (see Chapter 2).

In higher-tier field-testing methods, attention should be given to the method of application of the test compound (e.g. as an overspray, or as side-ward overspray to mimic differences in intensitiy, height and droplet size to which NTTP are exposed). Exposure concentrations in off-field spray drift is different from overspray and could be higher or lower. Some compounds react differently under greenhouse conditions compared to natural sunlight in the field. Therefore going from laboratory to greenhouse and from greenhouse to field makes exposure more realistic.

Higher-tier testing, and specifically field testing, involves a higher variability than Tier 1 laboratory studies and the power of higher-tier studies to detect effects should be quantified. The length of higher-tier studies can also be highly variable depending on the test species and endpoints of interest. For example, studies with wild plants might be conducted over two years and the results could be quite different from those obtained from short-term studies (e.g. 8 weeks) with fast growing species, due to other factors that vary over such a long study period.

Modelling approaches might include a set of model communities based on particular habitats (e.g. field edges, meadows). In such models, multiple landscape-scale exposure can be combined with species sensitivity and individually-based plant community models. A challenge for using modelling approaches in the risk assessment is to explain how model ouputs can be interpreted and used to assess risk. The implementation of modelling approaches in risk assessment also requires field datasets for the calibration and validation of individual models, which is a challenge as such datasets are rarely available. One of the applications of models is to extrapolate to different / representative landscapes. Modelling approaches might also provide insight on the magnitude of effects that can be expected based in information from field studies. Currently, the number of available models is low.

Monitoring studies may be useful for the calibration of risk assessement. One of the approaches for monitoring NTTP is the quadrat approach in which vegetation cover, taxonomy, and biodiversity at the community level are recorded. Monitoring data provide information about the actual field situation and possible multistressor conditions and is therefore needed as input for defining the reference Tier. However, defining a reference tier is difficult because of the gradient in the exposure in field situations. Workshop participants questioned whether a field study be used as a surrogate reference tier for calibrating Tier 1 and if a modeling approach could help.

3.4 Summary of higher tier testing and field studies

The workshop participants agreed that higher-tier tests have a number of benefits. Results of an ERA based on higher-tier studies might be used to calibrate the ERA based on lower-tier testing (in order to investigate what level of protection it presents). Field studies are also more realistic than Tier 1 studies. However, they also include higher variability. One of the benefits of field studies is that they offer options for refinements. Dissipation of the test compound in the field is expected to result in lower exposure and consequently reduced effects than in laboratory or glasshouse studies. Another benefit of field studies is that inter-species competition might be addressed in field studies, but not in greenhouse studies. Finally, field studies may allow the assessment of the effects of multiple stressors (e.g. several compounds in a crop approach and applied according to good agricultural practices).

The participants also agreed that higher-tier tests have a number of limitations. There is the problem of identifying appropriate reference sites, and the selection of key species according to the ecosystem services they provide (food source for birds and mammals, habitat, source for pollinators). Acceptability criteria for field studies also need to be defined and clarification is needed on their design and interpretation. Finally, higher-tier tests have limited representativeness of different field situations and therefore a combination of experiments and modelling approaches could be beneficial.

The reference community for NTTP risk assessment needs to be defined more precisely. Species traits must be considered when looking at assessment endpoints and communities to be used in higher-tier tests. For example, risk assessment might focus on specific species or group of species (key species) in the community based on the ecosystem services they provide (e.g. if we consider pollination, flowering plants might be considered as key species). It is important to assess effects on communities and to consider the landscape-scale context in the risk assessment.

All participants agreed that there should be a tiered approach to assessing risk to NTTP. However, one group of workshop participants preferred to have different higher-tier options available and to choose the best one with respect to the question raised and to the appropriate protection goals at stake (i.e. population at edge of field vs community at landscape level). The other group of participants preferred to consider a tiered approach based on a tiered system of single species studies > modified exposure tests > community tests > field studies.

In case of monitoring, participants highlighted the fact that there is no control in a monitoring study. Trends (correlations) can be identified but causation cannot be ascribed. Moreover, multiple stressors are automatically assessed in a monitoring study. Participants also highlighted that if field studies have Minimum Detectable Differences (MDD) between observations and measurements lower than 50 %, these MDDs are considered as good in an appropriate experimental context (EFSA, 2013). However, if the specific protection goals are formulated as effects of less than 10% then those differences are hardly ever measureable in field tests. Research is needed to quantify the power of higher-tier field tests bearing in mind that power is also dependent on the specific endpoint measured in the test.

3.5 Conclusions

The workshop participants agreed the following recommendations:

- If a refined effects assessment is required, potential higher-tier approaches include:
  - single species tests with refined exposure;

- o testing additional species (to allow SSD approach) and/or growth stages;
- o testing single species, multispecies or performing plot experiments (greenhouse, semi-field or field-testing);
- o perform population / community modelling.

Experimental studies include studies in the greenhouse (e.g. potted plants) and outdoor studies at semi-field or field level, e.g. plot experiments.

- There is a need for criteria to evaluate and interpret field studies in the context of specific protection goals.
- There is a need for defining a (surrogate) reference tier in order to calibrate the tiered approach.
- For defining a reference community for an EU level risk assessment, a trait-based approach seems promising in the view of some participants. Traits could be linked to the ecosystem services provided by NTTP.
- The following general tiered approach is appropriate for NTTP.

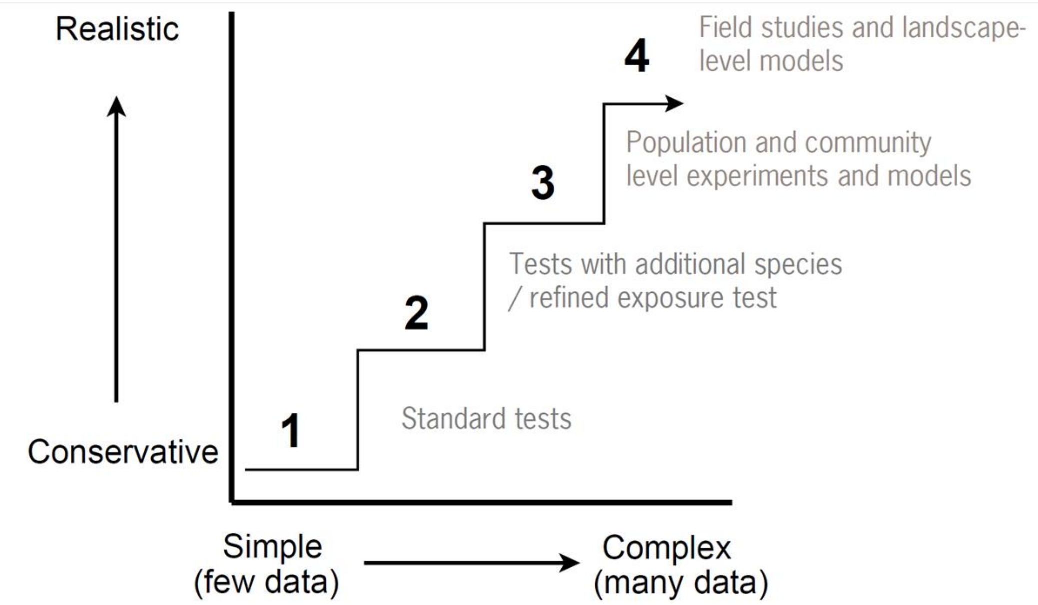


Figure 3.2: Tiered approach in NTTP effects assessment

Chapter 4: How to mitigate risks for non-target terrestrial plants

Eva Kohlschmid and Christoph Mayer

4.1 Introduction

Mitigation options that reduce the exposure of non-target organisms to plant protection products (PPPs) are an important refinement tool in cases where the standard risk assessment can not exclude unacceptable effects on non-target organisms. Mitigation measures could be part of the field management strategy (e.g. vegetated or non-vegetated buffer zones where no spraying of particular products is allowed) or they may be technological modifications of spraying equipment (e.g. drift reducing nozzles or shielding devices).

The topic of risk mitigation was discussed by all participants in three different breakout groups, which addressed the same charge questions:

a) Considering the MAgPIE workshop mitigation recommendations: which mitigation measures are appropriate for NTTP risk assessment?
b) What are the regulations / policies available for the protection goals of NTTP in agricultural landscapes? And how can they be used in concert with the pesticide regulation?

Each breakout group briefly presented the outcome of their discussion to the other participants as basis for further discussion in a plenary session. An introduction to the topic was provided with by two keynote presentations:

1. Common weed management practices (Martina Keller)
2. Outcomes of MAgPIE workshop (Véronique Poulsen, Jan van de Zande)

The first presentation by Martina Keller provided an overview on the general need and the general management options that are available for weed control, including non-chemical alternatives. The ‘take home message’ was that weed control is essential to sustain agricultural productivity in conventional and organic growing systems. Non-chemical weed control options, such as mechanical or thermal control, were generally unspecific with regard to the weed species affected. Hand weeding could target specific plant species if applied at later growth stages. However, hand and mechanical weeding may require a high frequency of interventions since, like chemical weed control, they are most effective when the target plants are in early growth stages.

The second presentation focused on the outcomes of a series of SETAC Workshops on risk mitigation of effects from PPPs (MAgPie; Alix et al., 2017; MAgPIE proceedings can be downloaded from https://www.setac.org/magpie). The presentation consisted of two parts. The first part, presented by Véronique Poulsen, introduced the MAgPie toolbox for farmer-focused mitigation options. Some of these risk management tools (e.g. buffer zones and spray drift reducing technology (SDRT)) are also important tools for the exposure refinement in risk assessments (e.g. for non-target plants, non-target arthropods and aquatic organisms). The second part of the kenote, presented by Jan van de Zande, provided an extensive overview of SDRT options as implemented in the Dutch national risk management process (for methodology and background information see Van de Zande et al., 2012; 2017).

4.2 Abstracts of keynote presentations

4.2.1 General weed management practices

Martina Keller (Agroscope, CH)

There are several definitions of the term "weed". A commonly used definition of weed scientists is: "a weed is a plant growing where it is not desired" (Zimdahl 1993). It is important to notice that this definition is completely anthropocentric (Zimdahl 1993) and today this definition and even the term "weed" (especially the German term "Unkraut") is rather debated in "non-weed science" communities.

Independent of the definition or the term used, weeds affected considerably and still affect humankind. A sign of this is that weeds are for example even mentioned in Genesis as a punishment to Adam and Eve: "thorns and thistles it shall bring forth for you; and you shall eat the plants of the field" (Genesis 3: 18; English Standard Version Bible) (Zimdahl 1993). Weeds are plants, which compete directly with our crop plants for light, water, nutrients and for space and consequently cause the highest potential yield losses compared to pathogens and animal pests (Kraehmer 2012a, Oerke 2006). Oerke found a potential yield loss of 34 % and an actual yield loss of 8.5 % due to weeds for the globally important crops wheat, rice, maize, potatoes, soybean, and cotton (2006). Apart from quantitative yield losses, weeds can reduce the quality of the produce or even cause food safety issues: For example chamomile florescence in canned peas or *Senecio* species in baby leaf salads (e.g. Anonymous 2013). Weeds can also harbor other pests and they do serve as hosts when the crop host is not present and thus undermine crop rotation efforts (Zimdahl 1993).

There are many management options to control weeds. One can distinguish between indirect control methods and direct methods (Pallutt 2002, Müller-Schärer 2002, Verschwele & Zerger 2002, Berger 2002 chapters in Zwerger and Ammon 2002). Examples of the formers are crop rotation, the cultivation of well-adapted crops, soil tillage, competitive varieties, seeding date and seeding density etc. (Pallut 2002).

Many weed species can produce very large numbers of seeds per single plant e.g. sheperdspurse 38'500 seeds per plant or redroot pigweed 117'400 seeds per plant (Zimdahl 1993). Therefore, prevention of seeding and the prevention of the spread by vegetative propagation means (e.g. rhizomes) is of great importance (Pallut 2002).

Direct control methods are biological (currently of minor importance), mechanical, thermal and chemical control methods (Pallutt 2002, Müller-Schärer 2002, Verschwele & Zerger 2002, Berger 2002 chapters in Zwerger & Ammon 2002). Mechanical weed control can be achieved by hand weeding or with equipment such as harrows or hoes etc. (Pallut 2002). The latter two can be controlled by cameras today and allow precise and rather fast weed control (Rueda-Ayla 2013, Melander 2014). "Flaming" of weeds is an example of a thermal control method (Verschwele & Zwerger 2002).

Weed control with herbicides was predominant in Europe in the last decades and still is (Kraehmer 2012a). Herbicides provide rather cheap, effective and selective weed control in many crops and allowed a considerable yield increase and considerable reduction of labor in agriculture (less hand weeding and less mechanical weed control in general) (e.g. Zimdahl 1993, Oerke 2006, Kraehmer 2012b, Keller et al 2015). Herbicide application in relation to the crop can take place as pre-planting application with or without incorporation and post-planting. The term pre-emergence and post-emergence application can be used in relation to the crop or the weeds. Generally, one distinguishes between foliar active, soil active and foliar as soil active herbicides. Another way to distinguish between herbicides is whether they are contact or systemic herbicides. Selectivity of a herbicide is of importance. It can range from non-selective (e.g. glyphosate or glufosinate, non GMO situation) to highly selective herbicides (Zimdahl 1993) (e.g. fluazifop-P-butyl applied in spinach to control grass weeds http://www.psm.admin.ch/psm/). While highly selective herbicides are useful to control a targeted spectra of weed species mainly in crops, the non selective glyphosate is a crucial tool for sustainable soil management in low and no-till cropping systems (Kraehmer 2012a). Furthermore, weed spectra of active ingredients vary, depending also on the dose and the time of application. Due to their many positive characteristics, herbicides are viable tools for integrated weed management.

Important concepts of integrated weed control are the economic thresholds (ETs) and the critical period of/for weed control (CPWC) (Harker and O'Donovan, 2013, Anonymous 2015a). Economic thresholds are of relevance in competitive crops such as cereals. In the case of weeds and herbicides, the use of ETs implies that herbicides are only applied if the loss caused by the weeds present in the field is equal or higher than the costs for the herbicides and the application costs (Coble and Mortensen 1992). For winter cereals, ETs were determined and established by Niemann (1981), and Gerowitt and Heitefuss (1990) in Germany. They suggested following thresholds: 40-50 plants m^{-2} for broad-leaved weeds, 20-30 plants m^{-2} for grass weeds, 0.1-0.5 plants m^{-2} for *Galium aparine* and 2.0 plants m^{-2} for *Fallopia convolvulus*. For the former 5-10% ground cover were also suggested as threshold (Gerowitt and Heitefuss 1990). In contrast, in non-competitive crops such as maize or vegetables the concept of the CPWC is used. This concept suggests that there is a period in the crop growth during which no competition by weeds is acceptable at all (Coble and Mortensen 1992; Hall et al 1992, Knezevic et al 2002). E.g. for rather "competitive" vegetable crops there is the rule of thumb that during the first half of the growth period the crop has to be kept weed-free, for the second part weeds can be accepted (if they will not interfere with harvest). For non-competitive vegetables, the CPWC is even longer. In leafy vegetables, no weeds can be tolerated at all during the complete growing period (Imhof et al. 2015). There are also cases in which weeds put consumers at risk. For example, *Solanum nigrum* in spinach, or *Senecio* species in baby leaf salads as mentioned above. Concluding effective weed control is important for food security and food safety.

Current challenges in weed control are resistant weeds, and especially the decreasing number of available active ingredients due to

i) the increasing number of active ingredients lost due to the review process of pesticides in the European Union and
ii) the decreasing number of launches of new active ingredients by chemical companies (Rüegg 2007, Kraehmer 2012b).

During the review process of existing pesticides as laid down in Directive 91/414/EEC about 1'000 active substances of pesticides were reviewed (main phase 2001-2009). Thereof, only about 250 active ingredients passed the assessment (Anonymous 2009, Karabelas et al 2009). Roughly, about 73 % of the herbicidal active ingredients (162 from 221) were withdrawn (based on the observation period from 2000-2008). Additionally, the use of active ingredients, which are still available, is more and more restricted (amount applied per area, or applied per area and time e.g. metazachlor http://www.psm.admin.ch/psm/). The renewal process of authorized active substances under the Regulation EC/1107/2009 is a continuing process and thus further withdrawals are likely in the coming years. In contrast to the many withdrawals, Stübler et al (2016) estimate that only about 3 to 4 new herbicidal active ingredients will be introduced into the market in the near future (2016-2020).

Regarding risk mitigation options. For the **in-crop/in-field** situation, it is as rather difficult to protect non- target plants. One could even argue that in the field all plants except the crop plants are (potential) targets. Yet, already the ETs and the CPWC provide some risk mitigation, as weeds are not controlled if they do not or do not anymore cause unacceptable losses. Site-specific weed control could allow even more specific and precise application in the field. However, these technologies have not been (widely) adopted yet (Christensen et al 2009). Further, since many, interactive factors affect yield, it is difficult (for most weeds) to link their occurrence with exact values for yield losses. Furthermore, the impact of a weed species on yield strongly depends on its density, the crop, the variety, the management system, other weeds present as the respective year and site (when and where the crop is grown). The weediness of a species does not need to be a trait appearing every year in every crop at every site. Therefore, it is difficult to determine, which species would qualify as non-target plant in a certain field and which not. Last, it needs to be considered that such a distinction is equally difficult/impossible if non-chemical weed control methods (hoeing, harrowing etc.) are applied. However, mechanical methods are expected to have negligible effects off-crop and off-field on non-target plants.

For the **off-crop/in-field** situation it is generally desirable to reduce/minimize drift. Apart from protecting non-target plants off-crop/in-field and off-field, drift reduction improves the image of farmers and agriculture in general, e.g. less pesticides in surface waters and avoidance of non-registered residues in adjacent crops (which is a big issue for vegetable production). Untreated field areas without crop need to be managed as well, as otherwise weeds establish and set seeds. Furthermore, they provide shelter for pests. In addition, if adjacent to the crop flowering plants are present, some insecticides cannot be applied and thus crop protection in-field/in-crop is hampered (e.g. thiamethoxam in lettuce in Switzerland http://www.psm.admin.ch/psm/).

Loss of diversity of occurring weed species has many causes (e.g. fertilizer, seed cleaning, adjusting the soil pH-value by liming, less crops grown, herbicides etc.) (e.g. Meyer 2013). Thus, only abandoning the use of herbicides will most likely not bring diversity back per se. However, there are **interesting projects** such as “100 Äcker für die Vielfalt” (Germany), “Projekt zur Förderung der natürlich vorkommenden Ackerbegleitflora” (Switzerland), which adjust the management system to enhance the reestablishment of rare weeds (Anonymous 2015b, Uebersax 2015). With such projects rare weeds might be more easily and more effectively protected than in untreated small stripes in the fields.

4.2.2 Outcomes of MAgPIE workshop: mitigation of environmental risks of pesticides in agriculture

Véronique Poulsen (ANSES, FR), Jan van de Zande (Wageningen Plant Research, NL)

Environmental risk mitigation measures are a key component in defining the conditions of use of pesticides in crop protection (EC, 2009a; EC, 2011). These risk mitigation measures derive directly from the evaluation of pesticide products and the risk assessment conducted for each use, and are reported in the approval regulations for an implementation in European Member States (EC, 2015). The recommended measures are specific of the type of risk they intend to mitigate. Hence, they range from the adjustment of the conditions of use to minimize transfers to groundwater to the setting of buffer zones at the edge of the crop. Eventually, they are reported on the labelling in the form of Safety Precaution Phrases, according to Regulation (EU) No 547/2011 (EC, 2011). The implementation of risk mitigation measures has resulted in the genesis of a wide variety of approaches, implemented at national levels as good agricultural practices or legislative measures. Therefore, the national mitigation options provide significant divergences between countries.

In this context, a 2-steps workshop was organized in April and November 2013, under the umbrella of the Society of Environmental Toxicology And Chemistry (SETAC) and the European Commission. The workshop aimed to develop a risk mitigation toolbox suitable to Member States’ needs in the context of Regulation (EC) No 1107/2009 that is taking into account farming practices and the legal frameworks.

Among the participants were risk assessors and risk managers of 24 European countries representing stakeholders from industry, academia and agronomical advisors/extension services. The first workshop (Rome, 22nd to 24th April 2013) was preceded by inventory of the environmental risk mitigation measures, which are used in European countries. During the workshop discussions focused on ground water protection, surface water protection (including the protection of aquatic organisms), protection of off-field areas and Protection of the in-field area. For these areas the collected information on mitigation tools was used to rank the tools according to criteria like acceptability, suitability, practicability and efficiency.

During the second workshop (Madrid, 13th to 15th November 2013) the content of the risk mitigation toolbox gathered for each group was discussed in detail. Mitigation measures initially ranked as principally suitable for further promotion were further evaluated concerning options to measure their effectiveness, overlaps with other regulatory frameworks, multifunctional aspects and potentials to optimize the measures.

This presentation was focused on mitigation measures applicable for non-target terrestrial plant (NTTP) risk assessment (Table 4.1). Based on an example for field margin types it was shown that promoting effects for NTTP may vary according to the type of field margin available in the field (Table 4.2). It was also discussed, that some effective mitigation options overlap with other European policies and regulations, e.g. CAP promoting 5% of farm holding to be reserved for (generic) recovery area or DIR 2009/128 on sustainable (EC, 2009b) use requiring the reduction of overall risk and impact of PPP-use.

Table 4.1: Mitigation measures according MAgPIE 2 applicable for NTTP risk assessment

Category	Risk Mitigation Measure	Category of risks that may be reduced
Buffer Zone	No spray zone, wind-dependant no spray zone, bare soil, landscape dependant buffer zones, aerial treatments	All organisms from exposure to spray drift
Field margin	Vegetated buffer zone	All organisms from exposure to spray drift / runoff Provides habitat and food resource
	Multifunctional field margin	All organisms from exposure to spray drift / runoff Provides habitat and food resource
Compensation areas	Recovery areas (ecological focus areas)	All organisms from exposure to spray drift / runoff (pending on location) Provides habitat and food resource
Spray drift reduction technologies	Nozzles, equipped sprayers, directed spray, precision treatments	All organisms from exposure to spray drift
Dust drift reduction technologies	High quality coating, low dust drillers	All organisms from exposure to dust drift
Conditions of application	Application rate and frequency management	All organisms from exposure to drift/runoff

Table 4.2: Effect of different field margin types on NTTP taking into perspective also parameter like value-cost-balance, practicality and efficiency of spray drift reduction (high values: high impact, low values: low impact). NR: Natural regeneration, GR: Grass sown, WF: Wildflower sown, P&N: Pollen and Nectar mix, WBS: Wild bird seed mix, AC: Wild bird seed mix, CH: Conservation headland. -1 = Negative impact for the environmental benefit 0 = No positive impact for the environmental benefit 1 = Some benefits for the environmental benefit 2 = Major benefits for the environmental benefit 3 = Most beneficial of all field margin types for the environmental benefit

Environmental benefit	Attribute	NR	GR	WF	P&N	WBS	AC	CH
Management	Value (AES or crop) vs costs	1	1	2	1	1	1	3
	Practicality	3	3	2	1	1	3	3
Plants	Overall	2	1	2	1	1	3	3
	Annual arable weeds	1	-1	-1	1	2	3	3
	Perennial wildflowers	3	2	3	1	1	1	1
Spray drift	Pesticides	3	3	3	2	2	2	2

Spray drift reducing techniques

Technological options to reduce spray drift of PPPs to the off-field area are e.g.: Drift reducing spray nozzles, Special equipment, Directed spray techniques, Precision treatment, Dose of product, etc. All these technological options can be combined with the use of buffer zones to maintain safety areas around the treated area.

In MAgPIE an inventory was made available of Spray Drift Reducing Technology (SDRT) used in the EU member states. Most implemented SDRT in the different countries was the use of Spray Drift Reducing Nozzles (DRN). SDRT and DRN can be categorised in level of spray drift reduction of 25%, 50%, 75%, 90% and 95% (ISO22369). The drift reduction potential of spray nozzles has to do with the amount of fine droplets in the spray. Coarsening the spray quality and reducing spray drift potential is based on drop size measurements of the spray fan and wind tunnel measurements. DRN are used for spray applications in arable crops, fruit crops and vines.

Spray Drift Reducing Technology can be based on crop-adapted and changed spray directions, the use of air assistance to guide the spray drops to the crop canopy or shielding the spray process (tunnel sprayers). But also sprayer boom height and sprayer speed influence spray drift potential to a high level resp. reducing spray drift with reduced boom height and increasing spray drift with increasing forward speed of the sprayer. All of these SDRT can be equipped with DRN to give higher levels of spray drift reduction than the technique equipped with standard nozzle types. In e.g. DE, NL, BE, FR, UK these SDRT are categorised and listed in official publications. Some PPP are only allowed to be applied with minimal levels of drift reduction specifying the minimal use of SDRT of e.g. 50% or higher. In NL the situation exists that applying PPP alongside waterways needs a minimal SDRT of 50% in the outside spray swath (>14 m) and depending on the toxicity of the PPP a higher level of SDRT (e.g. 90%).

SDRT are classified in drift reduction classes (ISO22369) following a standardised spray drift measuring protocol (ISO22866). However the defined reference spray technique, spray nozzle type and its used pressure and the evaluation distance at which the spray drift reduction is determined varies per country. This needs further exchange of information and harmonisation.

With the increasing use of new technologies in agriculture as of GPS, sensors and robotics new options for applying PPP are developed. Crop-adapted spraying, Canopy Density Spraying, Plant specific spraying, Map-based spraying are some of these new developments. They all have in common that PPP dose is adapted to what is detected in the field, e.g. plant position, crop size or plant canopy density. All these systems lead to a lower input of PPP and with the realised spray drift reduction of the application technology itself using these features the level of emission reduction to outside the treated area will be increased.

The development of SDRT is taken up by industry but the take up by farmers and regulators may be constrained by perceptions. Questions raised are often dealing with: the impact of

SDRT on product efficacy; the practicality of the implementation; the general availability on the market; the costs, benefits and economics; the awareness; and the enforceability. In general, these concerns are over-stated and suggest lesser awareness. They can however be addressed through support for best management practice awareness raising campaigns, technology inventories and user-directed tools to support customised assessments.

The MAgPIE workgroup concluded with some identified problems and some recommendations to come to their possible solutions.

Problem	Recommendations
SDRT: Developing independent nozzle classification schemes for all MS is impractical/inefficient	• MS Regulators may draw upon extensive experience and sound scientific foundation associated with schemes in place in other countries (NL, D, UK) supported by readily accessed SDRT database
Lack of harmonisation leads to inconsistency in role and take-up of SDRT and other techniques across Europe	• A greater degree of harmonization would: o encourage technological advances by farmers. o encourage submission of consistent risk assessments o increase zonal regulatory efficiency • Recommendation to adopt ***basic*** harmonized framework of SDRT efficacy thresholds (e.g. 50%, 75%, 90% and 95% effectiveness) • Expansion to include 99% would: o anticipate further technological advances o allow for options of compounded mitigation
Need to increase awareness amongst users	• Encourage support for user-friendly drift management tools; o Swedish “Hjälpreda” o TOPPS-Prowadis drift evaluation tool
Need to allow for greater flexibility in label S-phrases	• Present proposals that would support; o More effective / flexible communication of mitigation options o Allow for wide range of options o Accommodate local constraints and circumstances

In general it was concluded that SDRTs are a valuable and important measure for exposure refinement in the risk assessment, also for non-target plants.

4.3 Risk mitigation for non-target terrestrial plants: Report of workshop discussions (plenary and breakouts)

4.3.1 Considering MAgPIE recommendations, which mitigation measures are appropriate for NTTP risk assessment?

Three different breakout groups discussed which mitigation measures could be appropriate for NTTP risk assessment. Two groups considered that, in principle, all MAgPIE techniques were useful to mitigate risk to NTTP including compensation areas - either as field margin or recovery area. The third group, however, considered that the lack of detailed definitions for the different categories of vegetated buffer zones, multifunctional field margins, and recovery areas made it difficult to discuss their potential usefulness and implementation. This group proposed some ideas on how to proceed if the protection goal was not met on a landscape level. For example, management of buffer strips in such a way that the ecosystem service was provided to the magnitude defined in the specific protection goal. They further identified the need to evaluate the efficiency of mitigation measures for different non-target organisms. An overview of the efficacy of mitigation measures for different types of plants was provided in the second keynote presentation (see Table 4.2 in abstract above), but workshop participants requested more clarity in description, definition and aim of mitigation measures.

Other options proposed included specific management of fields in such a way that the specific protection goals are achieved at the landscape level (set aside areas). The group emphasized that the placing of certain mitigation measures depends on their purpose(s), the type of field management and the field surroundings. Buffer zones and multifunctional field margins buffer the influence of PPPs from the field to off-field areas, they are most efficient adjacent to non-cropped off-field areas, especially if these already contain high biodiversity. In contrast, such mitigations may be less effective if the off-field area is also a cropped field.

One group concluded that the primary role of field margins was risk mitigation (e.g. no spray zones/no crop zones) rather than a refuge for non-target species. This group also considered uncropped or completely non-vegetated buffer zones easier to regulate, as there was a lack of trust in farmers complying with no spray buffer zones when the field margins are cropped. Sweden apply wind-dependent no spray zones, that is, buffers that are required only on the downwind edge of the field. Wind-dependent no spray buffer zones, combined with an optional spraying of this area after the wind has turned, was considered as an optimal, pragmatic approach.

It was noted that the adoption of risk-reducing technologies/nozzles was not harmonized among Member States (MS). This is apparently due to the existence of different technical calibration systems for spray drift reducing technologies (SDRT) and a generally high variance in national acceptance of different mitigation tools. Accounting for this national variability in mitigation tools it was suggested that risk assessors on EU or zonal level should limit the definition of required mitigation to a relative scale, which could be expressed, for example, as percentage exposure reduction. The final definition of applicable mitigation measures are left to the MS, which may then chose the nationally/regionally appropriate mix of tools from a matrix of buffer, SDRT and other tools. Benefits and limitations of this approach were identified. The relative freedom of MS to choose the mitigation mix appropriate for their territory (flexibility) was considered positive. Whereas, the highly limited simultaneous acceptability of different options from the toolbox by farmer, risk manager and risk assessor (i.e. no pragmatic mitigation possible) was considered negative. As a general rule, farmers should be encouraged to conduct agriculture with adequate drift reducing technologies (DRT) and precision technology (GPS) to enable switching off nozzles adjacent to vulnerable/relevant areas.

Two breakout groups discussed compensation areas. One group was concerned about the control aspect in case compensation for applying certain products would be paid for within the CAP greening policy. The group was also unsure on what scale - farm-level, unified farm-level, etc. - such compensation areas should be implemented. Discussing how to combine compensation areas with risk assessment, they proposed to move to a landscape risk assessment, where compensation would need to be implemented for areas of low biodiversity. This group was also in favour of defining specific protection goals for in-crop areas and compensating for unavoidable effects on NTTP. The group acknowledged that such measured would be the responsibility of the MS and that the need for compensation would be product specific. For example, if a broad spectrum herbicide is applied compensation may be needed, in contrast, for a selective herbicide this may not be the case. As predicting the exact spray schemes needed for all fields and crops is impossible, farmers may be forced to provide compensation areas for a realistic worst case application scenario. However, as the second group pointed out, compensation areas are currently not a viable tool within the PPP regulation (1107/2009) and, as a consequence, assessment of in-crop risk probably fits better under the Sustainable Use and/or Habitats directives.

Both groups agreed that compensation measures generally require a significant level of management to ensure the intended benefits of the maintenance or enhancement of biodiversity and ecosystem services are achieved. There are different options on how to implement compensation areas (e.g. set aside fields, field margins and specific in-field areas) and their management depends on the primary focus of the compensation measure (e.g. cropped and unsprayed or seeded with mixture of flowers). This additional management was acknowledged as a cost for the farmer. The attractiveness and acceptability of policy measure to farmers are important for the effective implementation of compensation measures. For example, it was mentioned that in Germany that three quarter of the implemented EFA (ecological focus area of CAP, ‘greening’) are made up of catch crops and thus, the intended benefit of the ´greening´ for biodiversity and ecosystem services maintanance can be considered as marginal. A workshop participant also highlighted a specific issue of German policy where long-term compensation areas, which would have high biodiversity, are vulnerable to be taken out of use after several years of set aside. This practice would subsequently reduce the effective farming acreage and may be a reason why long-term programs are facing low acceptability amongst farmers[12].

4.3.2 What are the regulations and what policies are available for the protection goals of NTTP in agricultural landscapes? And how can they be used in concert with the pesticide regulation?

The breakout groups identified several policies addressing biodiversity, conservation, and environmental farm management topics. These include the Sustainable Use Directive and the National Action Plans linked to it, the Habitats Directive and the greening measures and agro-environmental schemes under the EU Common Agricultural Policy. It was unclear to most participants how these schemes and policies are interacting, thus a need for increased alignment between the EU legislations as well as national schemes was identified. Concerning the decisions according to the Sustainable Use Directive (128/2009/EC) and Regulation 1107/2009, it was explained that those are agreed on in joint meetings within DG Santé.

Given that biodiversity is addressed in policies at different geographical scales (e.g. local, national, EU) the question was raised: what is the appropriate level at which to address biodiversity issues related to PPPs? The national scale US Conservation Reserve Program, which pays farmers for removing environmentally sensitive land from agricultural production, was suggested as a possible example here. The need for evaluating the efficiency of the different risk mitigation options in different landscape contexts was again highlighted, as this would allow risk managers to choose the appropriate measures at the national level (depending on how the directives are implemented in MS, which may include compensation by farmers). The development of a toolbox for evaluation of the efficiency of different risk mitigation options was proposed.

4.3.3 Plenary discussion of mitigation options for NTTP.

The plenary discussion focused mainly on the applicability of the MAgPie toolbox for specific and multifunctional purposes, the interaction of risk mitigation with other regulation and policies and the spatial scale on which such mitigations should be applied.

---

[12] A post-workshop check revealed that the topic is apparently not restricted to Germany as it refers to the European Court Case C-47/13 (http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62013CJ0047&from=EN), in which the European Court ruled that any part of arable land, which was planted for more than five consequetive years with grass or herbaceous forage and has not been part of crop rotation, will be considered as 'permanent pasture' even if the land was ploughed or the soil otherwise worked during or after these 5 years. This means that such land is lost from a farmer's perspective for planting crops. It seems that arable set asides, which are funded by EU greening programmes, are exempt from this rule. However, the ruling is causing some confusion within farmer communities and lowering acceptance for longer-term greening programmes.

Concerning the mitigation measure toolbox from MAgPIE, there was broad consensus that it is appropriate. There was also agreement that risk assessors define the degree of risk reduction, whereas risk managers specify the specific tools needed to implement the risk reduction at a national level. A general need for harmonization between regulations and policies was identified. Although it was acknowledged that the risk assessment of PPPs could be informative for other regulations/policies, it remained unclear how the outcome of a PPP risk assessment could be incorporated into the respective management decision schemes. However, during the discussion it was evident that management decision schemes are relevant to the problem formulation phase of risk assessment and to the consideration of which mitigation options are appropriate for the different spatial scales of the risk assessment.

The key question in the discussion of the multi-functionality of mitigation measures (e.g. managing buffer zones for both spray drift mitigation and food web support) was: Would multi-functionality go beyond the Pesticide Regulation? The answer may be a qualified yes for the majority of mitigation options, since issues concerning such multi-functionality of mitigation options seem to be linked more to the landscape than to the specific PPP use on a certain field.

Some participants proposed that different policies could work alongside each other under an umbrella of sustainable agriculture, especially at the landscape scale. This may facilitate the translation of the risk assessment outcome into workable instructions/guidances to be applied under different policies (e.g. Sustainable Use directive, Habitat directive, local red lists, CAP set aside), in order to reach the most efficient compromise between agriculture and the environment.

The question was raised whether indirect in-crop effects are an issue at the landscape scale or at the farm scale? And, where is the boundary between PPP regulation and landscape-relevant policies? One proposal was that farmers who use PPPs in landscape where indirect effects were manifested needed to compensate for biodiversity loss in low biodiversity landscapes, but not in landscapes of high structural and biological diversity. For example, it was suggested that the re-establishment and maintenance of particular species (e.g. partridge populations) could be set as a protection goal that requires both in-field considerations and landscape level assessments. Other participants also indicated that they viewed the landscape context playing an important role on the outcome of the risk assessment of certain compounds and thus should be placed under the PPP regulation (1107/2009). However, this perspective was not shared by the majority of participants. The example the maintenance of species (partridge) populations, for instance, was considered a general problem of the landscape structure and not an issue of one single active substance. It was questioned whether the definition of scenarios that require the establishment of compensation areas were appropriate to the current PPP regulation (1107/2009). For example, measures to enhance landscape structures are not specific to PPPs or active substances and are therefore subject to other policies. They would, therefore be more appropriately be handled within, for example, the

Sustainable Use Directive and/or the Habitats Directive, including the national policies linked to these directives.

Mitigations linked to the inevitable in-crop effects caused by weed control were discussed in the 1st NTTP workshop and revisited in this workshop. The arguments were similar or the same as for the landscape aspect above. It was concluded that in-field effects on NTTP might (under certain landscape conditions) lead to unacceptable effects on other non-target species. For example, reduced in-field NTTP biodiversity may cascade through the food web reducing of the quality of habitat needed for the maintainance of certain farmland bird populations. A concern, which in principle, was acknowleged by all participants. However, disagreement remained on how to address this issue. Some particpants expressed the opinion that risk assessment should describe all risks, including risks via food web effects, and that risk management should consider all these risks under 1107/2009. However, most particpants did not agree that in-field food web effects are a unique feature of chemical weed (pest) control alone. On the contrary, there was a majority consensus that in-field reduction of plant biodiversity is inevitable within an agricultural production site were the totality of the agricultural management practices of a farmer are targeted to favour crop over non-crop plants in fields. This principle applies to both conventional and organic farming. As discussed in the first keynot presentation by Martina Keller, all weed control practices including non-chemical methods (e.g. soil steaming for seed bed preparation, flaming, frequent tilling, grubbing or hoeing) have a significant impact non-crop plant abundance in the fields. It was questioned whether the method of weed control is important if the impact is comparable. Proponents of specific in-field protection goals argued that the magnitude of these inevitable in-crop effects caused by the application of a herbicide may vary greatly depending on the surrounding landscape. However, this view was challenged by a risk regulator who posed the question: *Are the effects clearly an issue of the active substance or are they an issue of the surrounding landscape?* The first case would require application of the PPP regulation, whereas in the latter case other agro-environmental policies would apply instead.

The nature conservation concepts of land sparing vs. land sharing were discussed briefly. These concepts are another aspect of landscape management and clearly within the responsibility of the MS. The appropriateness of one or the other approach is clearly dependent on the focal species and their spatial habitat requirements. The general, in-field protection of NTTP would make the land sharing mandatory, reducing options for land sparing. Many of the currently rare arable weeds could benefit from measures within a land sharing approach, provided the economic pressures associated with such an approach do not force it to be undermined by changing management practices or land use (e.g. transformation of arable land into sites for infrastructure and real estate). On the other hand, there are species that do not tolerate arable management at all or need at least a considerable size of non-arable habitats, such species could benefit from measures within the land sparing approach (e.g. permanent or long-term set-asides as refuges or corridors). It was agreed that the connectivity between landscapes is very important and it is not only the patch size that matters. It was also agreed that neither land sparing nor land sharing are superior as both approaches are required to provide adequate protection of habitat for all species. Since elements of both approaches are essential for adequate environmental management of the agricultural landscape, there is a need to optimise the two approaches at a MS or lower local levels.

The participants agreed that the different regulations/policies were developed to address environment impacts at different geographical and temporal scales caused by multiple anthropogenic activities. The alignment of policy options for an efficient and sustainable balance between agricultural productivity and maintenance of the environment was perceived as the main challenge. This requires a clear definition and agreement of the potential protection needs and a cross reference to other existing policies to assess whether these protection needs are addressed elsewhere.

4.4 Summary

The majority of tools collated in MAgPIE aim to reduce the potential for off-field exposure. The off-field, in contrast to the in-field, is the area where a clear difference in environmental impact potential can be identified for the different weed control practices. While the non-chemical methods are considered to have no direct effects on plants growing outside the treated area, effects from off-field exposure to chemical weed control products are the main area of concern for NTTP. Drift was considered one of the main exposure routes for off-field NTTP in the first NTTP workshop. Spray drift is an important exposure route for other non-target organisms and was discussed in detail in a specific SETAC workshop on Spray Drift Exposure in February 2016, the topic was therefore not elaborated much further here.

In the in-field/in-crop area, all weed control practices, whether chemical or non-chemical, are aiming to reduce in-field weed density as much as needed in order to promote crop productivity and to reduce the number of occasions when weed control is required. The significant but unselective effects of all control practices on non-crop plants in-field was discussed intensively during the workshop. It was acknowledged that any suitable weed control practice is inevitably having an effect on in-field NTTP plant diversity and abundance. Workshop participants had differing views on the implications of the trade-off between effective weed control and potential in-field protection goals for NTTP. Some participants argued that compensation areas for in-field effects must be clearly linked to the application of a certain product, independent of the landscape structure. Other participants argued that the need for compensation areas must be clearly defined and subsequently it should be checked if the landscape already provides these compensation areas (e.g. adjacent nature reserves, rows of trees in the agricultural landscape, borders with (grassy) vegetation bordering ditches and streams) or if new compensation areas are necessary. As a general point, it should be noted that the MAgPie toolbox is considering mitigation measures that go beyond the basic PPP Regulation according to 1107/2009/EC. Most mitigation tools, which are linked to specific farm or field management, are falling under other European and/or national policies such as the CAP and the Sustainable Use Directive 128/2009/EC. The majority of participants were of the opinion that this would also be the most sensible approach for any compensation measure.

4.5 Conclusions

Mitigation measures as described in the report of the MAgPIE workshops are considered appropriate for non-target terrestrial plants (see the toolbox from MAgPIE, Table 4.2). Furthermore, mitigation measures should be considered in the context of the surrounding landscape. Further tools, like field margins and vegetated buffer strips could also be beneficial for non-crop plants growing in-field in two ways: field margins provide a habitat for NTTP and non-crop plants provide a habitat and shelter for birds or pollinators. Consequently, any in-field impact on NTTP can be compensated for in the field margins or vegetated buffer zones. However, vegetated strips need to be managed and the question is what to manage them for: reducing run-off, to protect annual / perennial NTTP, to provide habitat for other species. The management strategy for vegetated strips should be related to the specific protection goals, but needs to be regulated in a broader context. It was further concluded that the risk assessment indicates what proportion of risk reduction is required, but how to achieve this (i.e. implementation) is up to individual Member States. Management of vegetated strips can be part of a landscape-level management strategy for which a broader context of the landscape level is needed. Because of the trade-off between field productivity and the protection of in-field non-crop plants, compensation is the only option identified to make up for the inevitable level of effects on NTTP in the in-field area and should be defined by risk managers in the light of the specific protection goals. It was further concluded that several pieces of legislation may be relevant in concert with the pesticide regulation (EC) 1107/2009 when considering compensation (e.g. sustainable use directive; CAP; habitat directive).

Chapter 5: General conclusions, recommendations and outlook

Based on additional work performed between and after the two SETAC Europe NTTP workshops, including extended analyses of datasets from the open literature, from Good Laboratory Practice studies undertaken by Christl (2015a,b[13]) and the dataset used in the EFSA NTTP study (EFSA, 2014), the following conclusions could be drawn:

- Based on an analysis of sensitivities of wild versus crop species (Christl, 2015b[13]), testing with standard crop species appears to be protective of wild species (see Appendix 2 and Chapter 2).

- Comparative analysis of reproductive and vegetative endpoints indicates that the reproductive endpoints (mainly seed number and seed biomass as measures of seed production) are on average less than a factor of 2 more sensitive than the vegetative endpoints (vegetative vigour, biomass) when comparing the same point estimate (i.e. ER_{10}, ER_{25} or ER_{50} each). This conclusion was independent of dataset (EFSA, 2014 vs. an extended dataset containing published and unpublished information from Good Laboratory Practice studies (Christl, 2015a[14]).

    - Based on the recently completed re-analysis of a larger dataset (Christl, 2017b), the factors to extrapolate from vegetative endpoints for young plants to reproductive endpoints, maintaining the same effect level (i.e. ER_{50}, ER_{25}, ER_{10}), were calculated to range from 0.74 to 1.43. Additionally the extrapolation factors from an ER_{50} based on vegetative endpoints in juvenile plants to an ER_{10} or ER_{25} for reproduction ranged from 6.25 to 8.68 and from 2.32 to 3.69, respectively.

- Most data show that early-stage vegetative vigour testing also covers reproductive effects. Additional data may be proposed to cover potential effects on reproduction, for example for specific toxicological modes of action. There is an outstanding action to assess whether it is possible to predict which plant protection products have a much larger impact on reproduction than on vegetative growth

- Workshop participants agreed on a tiered approach for higher-tier risk assessment for NTTP. Workshop participants identified a number of benefits and limitations of higher-tier studies for NTTP. Potential higher tier approaches for NTTP include:

---

[13] Christl (2015b) has been replaced by Christl (2017a)

[14] Christl (2015a) has been replaced by Christl (2017b)

  - single species tests with refined exposure;
  - testing additional species (to allow SSD approach) and/or growth stages;
  - testing single species, multispecies or performing plot experiments (greenhouse, semi-field or field-testing);
  - perform population / community modelling.
- A number of research needs were identified related to higher-tier approaches for NTTP:
  - Modelling studies should be conducted to translate reductions in seed production/germination to effects on population size to link the test endpoints to the specific protection goals.
  - There is a need for criteria to evaluate and interpret field studies in the context of Specific Protection Goals.
  - There is a need for defining a (surrogate) reference tier in order to calibrate the tiered approach.
  - Basic research is needed for quantification of the power of higher tier field tests.
- Risk assessment should indicate what proportion of risk reduction is required, but how to achieve this (i.e. implementation) is up to individual Member States.
- Mitigation measures as described in the report of the MAgPIE workshops are considered appropriate for NTTP and should be considered in the context of the surrounding landscape.
- Field margins and vegetated buffer strips could provide benefit for non-crop plants growing in-field. However, vegetated strips need to be managed and the question is what to manage them for: reducing run-off, to protect annual / perennial NTTP, to provide habitat for other species). Management strategy of vegetated strips should be related to the specific protection goals, but need to be regulated in a broader context.
- Compensation is the only option identified to make up for the inevitable level of effects on NTTP in the in-field area and should be defined by risk managers in the light of the specific protection goals. Several pieces of legislation may be relevant in concert with the pesticide regulation (EC) 1107/2009 when considering compensation (e.g. sustainable use directive; CAP; habitat directive).

Acknowledgments

We thank the European Crop Protection Association for their support of the workshop. We also thank the Dutch Ministry of Economic Affairs for supporting the work in the Steering Committee. We also thank all participants for their contributions to the workshop and covering their workshop costs.

Chapter 6: References

Alix, A. C. Brown, E. Capri, G. Goerlitz, B. Golla, K. Knauer, V. Laabs, N. Mackay, A. Marchis, V. Poulsen, E. Alonso Prados, W. Reinert & M. Streloke (eds.), 2017. Mitigating the Risks of Plant Protection Products in the Environment: MAgPIE. SETAC, 2017. ISBN: 978-1-880611-99-9. Open Access e-book. To be downloaded from https://www.setac.org/magpie

Anonymous, 2009. http://ec.europa.eu/dgs/health_food-safety/information_sources/docs/plant/factsheet_pesticides_en.pdf Accessed: 5 Oct 2016

Anonymous, 2013. http://www.hortipendium.de/Gemeines_Kreuzkraut) Accessed: 19 Sep 2015

Anonymous, 2015a. https://www.nap-pflanzenschutz.de/de/praxis/integrierter-pflanzenschutz/grundsaetze-ips/ Accessed: 19 Sep 2015

Anonymous, 2015b. http://www.schutzacker.de/ Accessed: 19 Sep 2015

Arts G.H.P, Dollinger M., Kohlschmid, Maltby L., Ochoa-Acuña H., Poulsen V., 2015a. An ecosystem services approach to pesticide risk assessment and risk management of non-target terrestrial plants: recommendations from a SETAC Europe workshop, Environ Sci Pollut Res, 22:2350–2355.

Arts, G.H.P., Dollinger M., Kohlschmid, Maltby L., Ochoa-Acuña H., Poulsen V., 2015b. An ecosystem services approach to pesticide risk assessment and risk management of non-target terrestrial plants: recommendations from a SETAC Europe workshop. 56 pp. The report can be downloaded from the SETAC website and the Plant expert group page.

Boutin C., Aya K.L., Carpenter D., Thomas P.J., Rowland O., 2012. Phytotoxicity testing for herbicide regulation: Shortcomings in relation to biodiversity and ecosystem services in agrarian systems. Science of the Total Environment 415: 79-92.

Carpenter A., Boutin C., Allison J., 2013. Effects of chlorimuron ethyl on terrestrial and wetland plants: Levels of, and time to recovery following sub lethal exposure. Environmental Pollution, 172: 275-282.

Carpenter D. and Boutin C., 2010. Sublethal effects of the herbicide glufosinate ammonium on crops and wild plants: short-term effects compared to vegetative recovery and plant reproduction. Ecotoxicology, 19, 1322–1336.

Christensen S., Søgaard H. T., Kudsk P., Nørremark M., Lund I., Nadimi E. S., Jørgensen R., 2009. Site-specific weed control technologies. Weed Research 49, 233–241.

Christl, H., 2015a. Sensitivity of crops and wild plant species, and of vegetative and reproductive endpoints – Literature review and analysis for SETAC AG Plants. Keynote presentation at the second SETAC tripartite workshop on Terrestrial plants, held in Wageningen, September 21-22, 2015. [preliminary work of Christl, 2017b].

Christl, H., 2015b. Comparative assessment of the sensitivity of wild plant and crop species to plant protection products and their active substances, evaluated in laboratory and field tests, published data and regulatory (unpublished) studies, in context of Regulation 1107/2009 and the upcoming new Terrestrial Guidance document. An initiative of the SETAC tripartite workshop on Terrestrial plants, held in Wageningen, April 1-3, 2014. Tier 3 Solutions. Report B14037. March 2015. 142 pp. [preliminary report of Christl, 2017a].

Christl, H., 2017a. Comparative assessment of the sensitivity of wild plant and crop species to plant protection products and their active substances, evaluated in laboratory and field tests, published data and regulatory (unpublished) studies, in context of Regulation 1107/2009 and the upcoming new Terrestrial Guidance document. An initiative of the SETAC tripartite workshop on Terrestrial plants, held in Wageningen, April 1-3, 2014 (and extended after the second workshop 21-22 Sept 2015). Tier 3 Solutions. Report B14037-B15060. 6 April 2017. 227 pp. To be downloaded from the Tier 3 Homepage and the SETAC Plant Interest Group page.

Christl, H., 2017b. Non-target plants - comparative assessment of vegetative and reproductive endpoints - Literature review and analysis for SETAC AG Plants to plant protection products and their active substances, in context of Regulation 1107/2009 and the upcoming new Terrestrial Guidance document. An initiative of the SETAC tripartite workshop on Terrestrial plants held in Wageningen, April 1-3, 2014 (and extended after the second workshop 21-22 Sept 2015). Tier 3 Solutions. Report B15062. 6 April 2017. 244 pp. To be downloaded from the Tier 3 Homepage and the SETAC Plant Interest Group page.

Coble H. D., Mortensen D. A., 1992. The Threshold Concept and its Application to Weed Science. Weed Technology 6, 191-195.

COMMISSION REGULATION (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market

COMMISSION REGULATION (EU) No 284/2013 of 1 March 2013 setting out the data requirements for plant protection products, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market

COMMISSION REGULATION (EU) No 546/2011 of 10 June 2011 implementing Regulation (EC) No 1107/2009 of the European Parliament and of the Council as regards uniform principles for evaluation and authorisation of plant protection products, Official Journal of the European Union, 11.6.2011

EC, 2002. Guidance Document on Terrestrial Ecotoxicology in the context of the Directive 91/414/EEC. SANCO/10329/2002, rev.2 final, 17 October 2002, pp. 1–39.

EC, 2009. Regulation (ec) no 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC.

EC, 2009b. Directive 2009/128/ec of the European Parliament and of the Council of 21 October 2009 establishing a framework for Community action to achieve the sustainable use of pesticides.

EC, 2011. Commission Regulation (EU) No 547/2011 of 8 June 2011 implementing Regulation (EC) No 1107/2009 of the European Parliament and of the Council as regards labelling requirements for plant protection products.

EC, 2015. Commission implementing regulation (EU) No 540/2011 of 25 May 2011 implementing Regulation (EC) No 1107/2009 of the European Parliament and of the Council as regards the list of approved active substances, updated 01.01.2015.

EPA, 2001. Review of Non-Target Plant Toxicity Tests Under the North American Free Trade Agreement (NAFTA). Final Report of the FIFRA Scientific Advisory Panel (SAP) Meeting Held on June 27-29, 2001

European Food Safety Authority (EFSA), 2010. Scientific Opinion on the development of specific protection goal options for environmental risk assessment of pesticides, in particular in relation to the revision of the Guidance Documents on Aquatic and Terrestrial Ecotoxicology (SANCO/3268/2001 and SANCO/10329/2002). EFSA Journal, 8(10): 1821 (55 pp) doi: 10.2903/j.efsa.2010.1821.

European Food Safety Authority PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFSA Journal 2013;11(7):3290, 268 pp. doi:10.2903/j.efsa.2013.3290.

European Food Safety Authority (EFSA) PPR Panel (Panel on Plant Protection Products and their Residues), 2014. Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target terrestrial plants. EFSA Journal 2014;12(7):3800, 163 pp. doi:10.2903/j.efsa.2014.3800

Gerowitt B. and Heitefuss R., 1990. Weed economic thresholds in cereals in the Federal Republic of Germany. Crop Protection 9, 323-331.

Hahn, M., Lenhardt, P. P., & Brühl, C. A., 2014. Characterization of field margins in intensified agro-ecosystems—why narrow margins should matter in terrestrial pesticide risk assessment and management. Integrated environmental assessment and management. Environmental Toxicology and Chemistry (in press)

Harker K. N., O'Donovan J. T., 2013. Recent Weed Control, Weed Management, and Integrated Weed Management. Weed Technology 27, 1-11.

Imhof T., Kling J., Kuenzler R., Scheider M., Schnieper S., Thomas N., Wohler C., 2015. Wahlfach Gemüsebau (course vegetable production). Edition lmz, Zollikofen, Switzerland.

Karabelas A. J., Plakas K. V., Solomou E. S., Drossou V., Sarigiannis D. A., 2009. Impact of European legislation on marketed pesticides – A view from the standpoint of health impact assessment studies. Environment International 35, 1096-1107.

Keller M., Böhringer N., Möhring J., Rueda-Ayala V., Gutjahr C., Gerhards R., 2015. Changes in Weed Communities, Herbicides, Yield Levels and Effect of Weeds on Yield in Winter Cereals Based on Three Decades of Field Experiments in South-Western Germany. Gesunde Pflanzen 67, 11-20.

Kleijn D. and Snoeijing I.J., 1997. Field boundary vegetation and the effects of agrochemical drift: botanical change caused by low levels of herbicide and fertilizer. Journal of Applied Ecology, 34: 1413-1425.

Kleijn, D., and Verbeek, M., 2000. Factors affecting the species composition of arable field boundary vegetation. Journal of Applied Ecology, 37(2), 256-266.

Knezevic S. Z., Evans S. P., Blankenship E. E., Van Acker R. C., Lindquist J. L., 2002. Critical period for weed control: the concept and data analysis. Weed Science 50, 773-786.

Kraehmer H., 2012a. Innovation: Changing trends in herbicide discovery. Outlooks on Pest Management 23: 115-118.

Kraehmer H., 2012b. Technical demands and political restrictions for weed control. 25th German Conference on Weed Biology and Weed Control, March 13-15, 2012, Braunschweig, Germany.

Melander B., 2014. Physical and cultural weed control –status and future directions. Keynote presentation at: XXIX Brazilian Weed Science Conference, 1-4 September 2014, Gramado, Brazil.

Meyer S., 2013. Impoverishment of the arable flora of Central Germany during the past 50 years: a multiple-scale analysis. Biodiversity and Ecology Series B 9 Göttingen Centre for Biodiversity and Ecology, Göttingen, Germany.

Niemann P., 1981. Schadschwellen bei der Unkrautbekämpfung (Thresholds for weed control), Volume 257, Reihe A: Angewandte Wissenschaft–Schriftenreihe des Bundesministers für Ernährung, Landwirtschaft + Forsten. Landwirtschaftsverlag, Münster, Germany.

Norberg-King T.J., 1993. A linear interpolation method for sublethal toxicity: the inhibition concentration (ICp) approach (version 2.0). Technical Report 03-93. US Environmental Protection Agency, Environmental Research Laboratory, Duluth, MN.

OECD GUIDELINES FOR THE TESTING OF CHEMICALS, 227. Terrestrial Plant Test: Vegetative Vigour Test, 2006a.

OECD GUIDELINES FOR THE TESTING OF CHEMICALS, 208. Terrestrial Plant Test: Seedling Emergence and Seedling Growth Test, 2006b.

Oerke E-C., 2006. Centenary Review Crop losses to pests. Journal of Agricultural Sciences 144, 31-43.

Olszyk D., Pfleeger T., Lee E.H. and Plocher M., 2009. Pea (*Pisum sativum*) seed production as an assay for reproductive effects due to herbicides. Environmental Toxicology and Chemistry, 28, 1920–1929.

Reuter S., Siemoneit-Gast S., 2007. Entwicklung einer weiterführenden Methode zur Bewertung des Risikos für terrestrische Pflanzen durch Exposition mit Pflanzenschutzmitteln und ihren Wirkstoffen (Development of an extended method for assessing the risk to terrestrial plants exposed to plant protection products and their active ingredients). Report (in German with Abstract in English). For the UBA. UFOPLAN-Nr. FKZ 202 67 431. Institut für Agrarökologie. RLP AgroScience GmbH, Germany, 203p

Rotchés-Ribalta R., Boutin C., Blanco-Moreno J.M., Carpenter D.J. and Sans X.F., 2012. Effects of herbicide on ecological patterns of characteristic and rare weed species. British Ecological Society, Birmingham, UK.

Rueda-Ayala V.; Weis M., Keller M., Andújar D., Gerhards R., 2013. Development and Testing of a Decision Making Based Method to Adjust Automatically the Harrowing Intensity. Sensors 2013, 13, 6254-6271.

Rüegg W. T., Quadranti M., Zoschke A., 2007. Herbicide research and development: challenges and opportunities. Weed Research 47, 271–275.

SANCO/10329/2002, Guidance Document on Terrestrial Ecotoxicology Under Council Directive 91/414/EEC, 17 October 2002.

Schmitz, J., Schäfer, K., & Brühl, C. A., 2013. Agrochemicals in field margins—Assessing the impacts of herbicides, insecticides, and fertilizer on the common buttercup (Ranunculus acris). Environmental Toxicology and Chemistry, 32(5), 1124-1131.

Schmitz, J., Schäfer, K., & Brühl, C. A., 2014a. Agrochemicals in field margins—Field evaluation of plant reproduction effects. Agriculture, Ecosystems & Environment, 189, 82-91.

Schmitz, J., Hahn, M., & Brühl, C. A., 2014b. Agrochemicals in field margins–An experimental field study to assess the impacts of pesticides and fertilizers on a natural plant community. Agriculture, Ecosystems & Environment, 193, 60-69.

Seefeldt S.S., Jensen J.E. and Fürst E.P., 1995. Log-logistic analysis of dose–response relationships. Weed Technology, 9, 218–227.

Sheppard S.C., Evenden W.G., Abboud S.A., Stephenson M., 1993. A plant life-cycle bioassay for contaminated soil, with comparison to other bioassays: mercury and zinc. Arch. Environ. Contam. Toxicol. 25: 27-35.

Sterling, T.D., 1959. Publication decisions and their possible effects on inferences drawn from tests of significance - or vice versa. Journal of the American Statistical Association. 54 (285): 30–34.

Strandberg B., Mathiassen S.K., Bruus M., Kjær C., Damgaard C., Andersen H.V., Bossi R., Løfstrøm P., Larsen S.E., Bak J. and Kudsk P., 2012. Effects of herbicides on non-target plants: How do effects in standard plant tests relate to effects in natural habitats? Pesticide Research No. 137, 2012. Danish Ministry of the Environment, Environmental Protection Agency, 114 pp.

Stübler H., Busch M., Strek H., 2016. Invited plenary presentation: Weed control at the cross roads – which innovations are on the horizon? "Weed Science and Management to Feed the Planet" Proceedings of the 7th International Weed Science Congress, June 19-25, Prague, Czech Republic, Czech University of Life Sciences Prague, 8-12.

Uebersax A., 2015. Schlummernde Raritäten auf dem Acker (rare weeds „napping" in the fields). Schweizer Bauer 26 Sep 2015, 7. http://www.agrofutura.ch/projekt/ressourcenprojekt-zur-erhaltung-und-f%C3%B6rderung-gef%C3%A4hrdeter-schweizer-ackerbegleitflora-kanton Accessed: 1 Oct 2015.

Zande, J.C. van de, H.J. Holterman and J.F.M. Huijsmans, 2012. Spray drift assessment of exposure of aquatic organisms to plant protection products in the Netherlands. Part 1: Field crops and downward spraying. Wageningen UR Plant Research International, Plant Research International Report 419, Wageningen. 2012. 84p.

Zande, J.C. van de, H.J. Holterman, J.F.M. Huijsmans and M. Wenneker. 2017. Spray drift for the assessment of exposure of aquatic organisms to plant protection products in the Netherlands. Part 2: Sideways and upward sprayed fruit and tree crops. Wageningen UR, Wageningen Plant Research Report 564, Wageningen. 2017. 66p.

Zeegers M., 2016. Science integrity and publication bias. Keynote presentation and seminar held at the EU SETAC meeting. 22^{th} – 26^{th} May 2016, Nantes, France.

Zimdahl R. L., 1993. Fundamentals of Weed Science. Academic Press, Inc. California, United States of America.

Zwerger P., Ammon H. U., 2002. Unkraut – Ökologie und Bekämpfung (Weeds – Ecology and Control) Verlag Eugen Ulmer, Stuttgart, Germany.

Appendix 1: Names and affiliations of participants

First name	Surname	Affiliation
Udo	Hommen	Fraunhofer Institut
Lorraine	Maltby	The University of Sheffield
Jette	Reeg	University of Potsdam
Jan	van de Zande	Plant Research International
Beate	Strandberg	Department of Bioscience, University of Aarhus
Sandra	Siemoneit	RLP AgroScience GmbH
Gertie	Arts	Wageningen Environmental Research (Alterra)
Eva	Kohlschmid	Agroscope
Steffen	Matezki	UBA
Veronique	Poulsen	Anses
Peter	Van Vliet	Ctgb
Robert	Luttik	Retired
Carmen	Schweikert	Umweltbundesamt (UBA)
Streissl	Franz	EFSA staff
Martina	Keller	Agroscope
Karin	Nienstedt	EU Commission
Jose	Tarazona	EFSA
Joanna	Davies[1]	Syngenta
Margit	Dollinger	Bayer Cropscience
Hank	Krueger	Wildlife International
Giovanna	Meregalli[1]	Dow
Christoph	Mayer	BASF
Thorsten	Schad	Bayer Cropscience
Anne	Thompson[1]	Dow
John	Wright	Monsanto
Holger	Teresiak	Agro-check
Heino	Christl	Tier 3 solutions
Kees	Romijn	ECPA EEG chair
Stefania	Loutsetii	DuPont
Chen	Teel	DuPont

[1] unable to attend the meeting due to unforeseen circumstances.

Appendix 2: Abstract of Christl, H., 2017a. Sensitivity of wild plant and crop species to plant protection products. Literature review and analysis for SETAC AG Plants

Heino Christl[1,2]

[1] on behalf of the first SETAC workshop on “Non-target terrestrial plants”
[2]Tier3 solutions GmbH, Kolberger Str. 61-63, 51381 Leverkusen, DE

E-mail contact: Heino.Christl@tier3.de.

Introduction

Following recommendations of the 1st SETAC workshop on “Non-target terrestrial plants” a literature review was performed to compare the sensitivity of terrestrial plant species (crop species and wild species), aiming to test the hypothesis that wild plant species could be more sensitive to plant protection products (PPPs) than the standard test species - which generally, though not exclusively, are crop species. Published literature and unpublished data generated for the registration of PPPs were searched for this review, and endpoints from crop species and wild species were compared. An EFSA expert group recently worked on a Scientific Opinion also touching this point [1] and their concerns were also considered.

Materials & methods

Available Data

A formal literature search was performed, papers with ERx-endpoints of non-target plant species were assessed and the endpoints included in a database. Based on an initial evaluation of published data, substances were identified that were most likely to permit analysis, i.e. those where 6 or more wild plant species endpoints were available.

The second source of data were regulatory studies (GLP, standard guideline, i.e. OECD 208 and 227 or OPPTS 850.4250 Vegetative vigour Tier 2, and OPPTS 850.4225 Seedling Emergence Tier 2, and corresponding TIER 1 documents. The members of the Crop Life International NTP group were asked to provide the confidential endpoints of their regulatory studies (GLP, seedling emergence (SE) and vegetative vigour (VV)), in particular for the substances identified during the initial evaluation of the published literature. These data were incorporated in the data base. Subsequently further data collated for a different project could be included and the database was extended. On request of commenting members of the workshop, also ER_{10} were included in the evaluation (previously analysis focused only on ER_{25} and ER_{50}), and we assessed all substances for which crop and wild data were available.

Evaluation

Each endpoint is characterised by three key elements: the study type (e.g. SE, VV, or corresponding field tests with incorporation or with spray application), the measured variable (e.g. seedling emergence, shoot height, biomass wet- or dry weight, etc.) and the effect level (e.g. ER_{10}, ER_{25}, ER_{50}, ER_{75}, etc.). Further known variables considered in the database were the mode of action, the test material (e.g. active substance, formulation), the test species, the taxonomy (e.g. family, monocot/dicot), the test system (i.e. lab, field) and the ‘anthropinistic affiliation’ (i.e. crop species, wild plant species) etc. Confounding data were also initially considered, but their subsequent analysis was impeded by incomplete reporting in the published papers. Subsequently, the database was searched for subsets of data where matching datasets both of crop or wild plant species were available, active substance by active substance (a.s.).

The fundamental rule in the analysis was to compare like with like, e.g. ER_{25} endpoints of a.s. ”X” based on biomass from SE studies, or ER_{50} endpoints of a.s. ”Z” based on shoot height from VV/foliage applied field studies. For the graphical presentation, the selection of endpoints was normalized by dividing them by their common geometric mean, so that no data could be attributed to any particular a.s. (important step to guarantee data confidentiality). In the graphical presentation differentiation between monocots and dicots was also possible. The selected data was split into crop species and wild plant species and, for each subgroup, the lowest endpoint (minimum – most sensitive species per group) and the average endpoint (geometric mean, average sensitivity of the subgroup) calculated: the former as it is the regulatory most relevant endpoint, the latter as it is less affected by different and extreme values, so better reflects the overall sensitivity of the whole group.

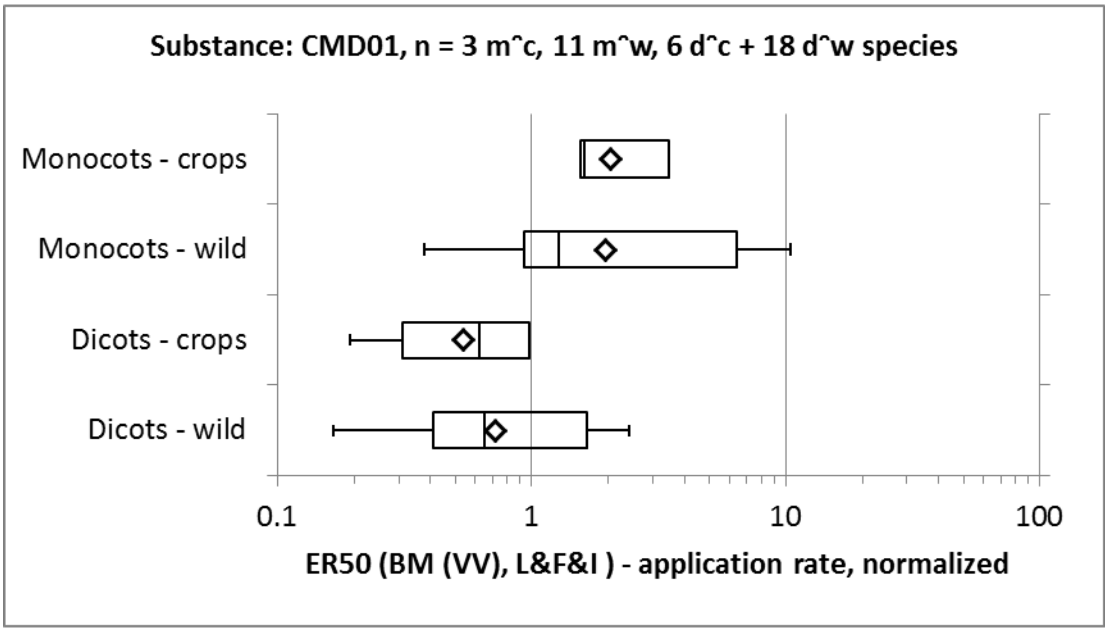
Based on either the most sensitive species, or the mean sensitivity of the two groups, quotients were calculated, expressing the difference between crop species and wild species. In the revised assessment (incorporating comments from the Steering Committee and workshop participants) also ER_{10} were considered, (the previous version considered only ER_{25} and ER_{50} endpoints, as these were more often reported and as ER_{10} are considerably less reliable than ER_{25} or ER_{50} values for mathematical reasons). The main assessment was based on vegetative vigour studies and biomass data, this being the subset of data with most observations. Additional statistical analysis considered also other vegetative endpoints and seedling emergence data where available.

Results

A total of 3057 species/a.s. test combinations were included in the data base, comprising 50 herbicides. For 42 of these both crop and wild species numeric endpoints were available (with at least one effect level in both groups, e.g. a crop-ER_{25} and a wild-species ER_{25}, allowing calculation of a quotient). A total of 74 quotients could be calculated (based on ER_{10}, ER_{25} or ER_{50} each).

Individual comparisons of crop and wild species endpoints

Example plot visualising one individual set of endpoints (one a.s., one endpoint). The same type of figure was generated for all active substances with matching crop and wild species data (here e.g. tests with substance CMD01, ER_{50} endpoints (biomass VV, all study types (lab, field or intermediate)) of wild plant and crop species).

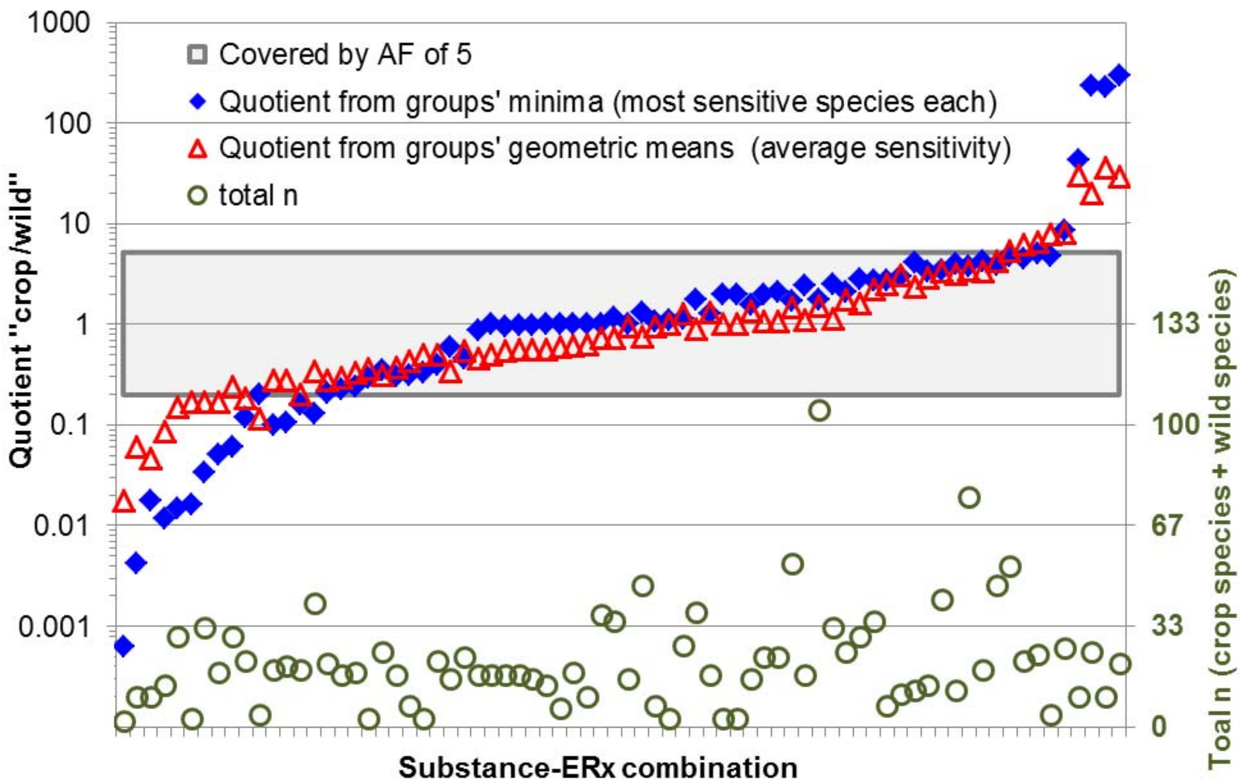


Appendix Figure 1: Example: Substance CMDx - Distribution of ER_{50} endpoints (biomass) of wild plant and crop species only biomass data from vegetative vigour studies. The rhombi mark the geometric mean of data points, the box 25, 50 and 75%iles, the whiskers minima and maxima.

Of the resulting distributions the following key parameters were used for the quotient approach, the geometric means as central estimates, and the minimum as regulatory relevant estimate.

Overall comparison - quotients

The example above allows to illustrate how the final evaluation may either be based on the minima (i.e. the most sensitive species of each group) or on the geometric means, which are a measure of the average sensitivity. Based on the most sensitive species (left whisker bars), the most sensitive was a wild dicot species. However, based on the geometric mean of dicots (rhombi), the crops appeared to be more sensitive than the wild species. These differences in sensitivity were expressed by means of a quotient, dividing the crop endpoint by the wild species endpoint. Quotients below "1" indicate cases where crop species were more sensitive than wild species (i.e. crop endpoints lower than wild species' endpoints), quotients above "1" indicate that wild species were more sensitive than crop species. The calculated quotients are displayed in the following Appendix Figure 2.



Appendix Figure 2: Quotients (triangles and rhombi) of the individual cases (combinations of substances and endpoints) on a log scale, sorted by average quotients (of minima- and geometric means-quotients) in ascending order. Cases above 1 indicate that wild species were more sensitive than crop species, cases below 1 that crop species were more sensitive; either based on their central value (geometric mean) = triangles, or on the groups' minima (comparison of the most sensitive species of each group) = rhombi.. The green circles at the bottom indicate the number of species (lower n of the compared groups, secondary ordinate).

In this chart it is apparent that there are about as many low quotients (below 1) as above 1.

The circles at the bottom indicate the number of species (lower n of the compared groups). With the updated database there are higher 'n' i.e. more species per substance-ERx-combination at the right side where quotients are above 1, but there are fewer of these substance-ERx-combinations.

In addition it is apparent that quotients based on the minima i.e. the two most sensitive species of each dataset (blue rhombi) show larger deviations from the central value (which is close to one) than the quotients based on the average sensitivity of the groups (red triangles). 73% of the quotients based on minima were within the rectangle indicating the area covered by an assessment factor of 5. Basing the assessment on average sensitivities, 74.3% of the quotients are within the grey rectangle. Regarding the quotients outside of the grey rectangle, i.e. outside the area covered by an assessment factor of 5, 8.1% based on minima and 12.2% based on average sensitivity were greater than 5 while 18.9% based on minima and only 13.5% of those based on average sensitivity were smaller than 0.2. Also the extent of deviation was larger for the quotients based on the group's minima, in particular at the left tail.

The overall outcome is thus a strong support that by and large there is no consistent difference in sensitivity between crop and wild plant species (at least for the systematic groups for which data were available), if matching endpoints are assessed i.e. only like with like is compared.

The overall average of quotients based on the most sensitive species of each group was 0.84, and on mean group sensitivity was 0.87. Weighted averages (considering total n) were 1.11 for quotients based on minima, and 1.09 for those based on mean group sensitivity, respectively.

Further statistical methods

On request of workshop participants additional statistical methods were applied, namely (A) a multiple regression analysis of all data performed in-house, and (B) comparisons of distributions including censored values (the assessments above considered only numeric endpoints), followed by ANOVA analyses, which was provided by John W. Green, DuPont. This assessment also included seedling emergence data and tentatively also endpoints based on other measurements than biomass (e.g. shoot height). Both assessments were subsequently repeated and extended, including ER_{10} endpoints and additional datasets.

In assessment (A) we found with the standard model (no interactions) and including all effect levels (pooled model) only slight differences between crops and wild plants in both directions (depending on the effect level assessed), and with the model including interactions (mathematically defined) between confounding parameters that, based on all data (including vegetative ER_{10}, ER_{25} and ER_{50} data), wild plant endpoints appear – depending on the model

– either to be similar or to be slightly, albeit significantly, higher than crop endpoints (up to a factor of 1.9). With the Mixed Effect Model wild plant endpoints appeared – depending on the model – either to be similar or to be slightly, albeit significantly, lower than crop endpoints (up to a factor of -1.6). So, overall, the outcome indicates high statistical power and sometimes significant deviations in either direction, but no consistent difference between crops and wild plants, let alone by a large margin.

In analysis (B), following generation of adjusted distributions (also considering censored values) several ANOVAs were performed considering the fundamental explanatory variables, including mode of action and plant families. The analysis found for certain modes of action and plant family-combinations significant differences between crops and wild species, but again in both directions, and not consistently for all effect levels, but e.g. only for ER_{25}. One of the main findings was that it might be worth to test Poaceae regularly as they had deviating endpoints for a number of modes of action. This finding might reflect the fact that there is a lot of Poaceae data available, both for crops and for wild species. However, considering that Poaceae are already tested regularly in standard lab tests generated for risk assessment purposes, this finding does not trigger any additional testing recommendation. Overall analysis (B) also did not retrieve conspicuous differences in sensitivity between crops and wild plants.

The hypothesis that wild species are per se more sensitive to agrochemicals than wild crops was thus not supported by any of these additional statistical analyses.

Conclusion

The overall conclusion was that, based on ER_{10} ER_{25} and ER_{50} endpoints, focusing on vegetative biomass measurements and on grasses or shrubs, i.e. annual, biennial or perennial herbaceous plants, the available data sets do not indicate consistent differences in sensitivity between crop species and wild plant species. As such crop species, typically tested in studies used for regulatory purposes, can be regarded as adequate surrogates for non-target terrestrial plants, including wild species.

References

[1] EFSA 2014. Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target terrestrial plants. EFSA Journal 2014;12(7):3800, 163 pp. doi:10.2903/j.efsa.2014.3800

Acknowledgement – This review was sponsored by members of the CLI NTP group and ECPA. We thank the Steering Committed and the peer-reviewers from the 1st SETAC workshop on “Non-target terrestrial plants” and the members of the CLI NTP group for their comments.

Appendix 3: Abstract of Christl, H., 2017b. Sensitivity of vegetative and reproductive endpoints to plant protection products. Literature review and analysis for SETAC AG Plants

Heino Christl[1,2]

[1] on behalf of the first SETAC workshop on "Non-target terrestrial plants"
[2]Tier3 solutions GmbH, Kolberger Str. 61-63, 51381 Leverkusen, DE

E-mail contact: Heino.Christl@tier3.de.

Introduction

Following recommendations of the 1st SETAC workshop on "Non-target terrestrial plants" and considering comments made during the 2nd SETAC workshop a literature review was performed to compare the sensitivity of endpoint groups, i.e. vegetative endpoints and reproductive endpoints of terrestrial plant species. The goal was to test the hypothesis that reproductive endpoints are generally more sensitive than the vegetative endpoints that are available from standard regulatory tests. Published literature and unpublished data generated for the registration of PPPs were searched for this review. An EFSA expert group recently produced a Scientific Opinion covering these topics (EFSA, 2014) and the data on which their analysis was based were also assessed in this review.

Materials and methods

Available Data (focussing on reproductive vs vegetative vigour endpoints)

Formal literature searches were performed, papers with ERx-endpoints for non-target plant species were assessed and the endpoints included in a database. Based on an initial evaluation of published data, the species-substances combinations for which reproductive endpoints were available were identified. For these combinations the database (including datasets from the species-sensitivity evaluations, and EU data (DAR or Review endpoints)) was searched for the vegetative endpoints of matching test-substance - species combinations. The data available was mainly concerned with grasses or shrubs, i.e. annual, biennial or perennial herbaceous plants.

The data listed in the Appendix of EFSA's Scientific Opinion (EFSA, 2014) were assessed in parallel in order to check whether using their database would lead to different results. Any data listed in the EFSA 2014 Scientific Opinion but until then not included, were also incorporated into our database. Further confounding factors such as testing conditions and test design were retrieved from the original papers where available. Some of the endpoints listed in the EFSA 2014 Scientific Opinion were recalculated by the authors of the Scientific Opinion and not included in the original publication. In these cases, both the original endpoint and the endpoint recalculated by the authors of the EFSA Scientific Opinion were included, but with the same experiment number[15]. Consequently for the comparison based on minima, the lowest endpoint was considered, while for assessments based on central estimates the geometric mean was selected (this applies not only to the EFSA 2014 Scientific Opinion data but also to the whole database).

Evaluation

Most of the papers generating reproductive endpoints also reported vegetative endpoints. In a first step these data pairs were assessed visually, and quotients were calculated (dividing the vegetative by the reproductive endpoint), thus specifying if, in a given test, vegetative endpoints were lower (quotients <1) or higher (quotients >1) than the corresponding reproductive endpoints. At this stage there were often multiple observations for a given substance - species combination. In a second step, we aimed to consolidate data by generating average (geometric mean) or worst-case (minimum) descriptors for the sensitivity of the two parameters, so that just one quotient per endpoint type (ER_{10}, ER_{25}, ER_{50}) and substance-species-combination was obtained. This consolidation was performed in two steps: first by experiment (i.e. selecting either the lowest or the geometric mean of all endpoints reported, to obtain a single value for each physical experiment), and second by substance-species combination (again selecting the overall lowest or the geometric mean), so that ultimately four combinations were obtained (see columns in Appendix Table 1). In the EFSA 2014 Scientific Opinion just one set of endpoints is listed per substance-species combination; the authors do not discuss how they selected the one displayed out of any further endpoints if such were available. We have assumed, the authors of the EFSA Scientific Opinion (2014) chose always the lowest endpoint.

Based on the results of a previous assessment[16] (in which differences in sensitivity between greenhouse and field data were found to be less pronounced than expected), field and greenhouse data were considered together, i.e. based on minima the lowest reproductive endpoint was compared with the lowest vegetative endpoint, irrespective of whether it had been generated in a field study or in a lab/greenhouse study. Also, the huge variety of different reproductive variables measured (e.g. number of inflorescences or flowers, number of pods, number or weight of seeds, germination success of the F1-generation[17]) were

---

[15] The experiment number is a unique number attributed to every experiment to prevent one experiment being considered twice or more, even if several endpoints of the same experiment were reported.

[16] See Christl (2017a) and the abstract of this report in Appendix 2.

[17] We are aware that some of these endpoints are more and others less relevant, but the decision on which endpoints to consider and which not would inevitably include a judgemental aspect and lead to discussions in the course of any potential re-evaluation. However, we did exclude some reproduction-related endpoints that are not relevant for a non-target plant population reproduction, such as even colouring of apples or marketability of cucumbers.

considered together at this stage. For the paired approach, all these were used both for a geometric mean and for a worst case (minimum) overall measure of toxicity.

An aspect based on initial assessments considered to be important was the age/growth stage of the plant at application, and the subsequent duration of the observation until evaluation. In the initial assessment of data pairs generally matching endpoints were compared; effects on vegetative endpoints of mature plants with reproductive effects on mature plants. This approach thus answers the scientifically valid question of whether the physiological processes leading to vegetative growth and the more complex processes leading to reproduction are similarly susceptible to exposure to herbicides. This is also one of the questions discussed in the EFSA Scientific Opinion (Appendix B; EFSA, 2014). However, the more relevant question from a regulatory perspective is a different one, namely: how does the sensitivity of reproductive endpoints from mature plants relate to vegetative endpoints of young plants? The latter are the endpoints always available for the risk assessment of herbicidal active substances, and the current discussion aims to answer the question of whether we miss a fundamental protection goal when we use only vegetative endpoints of young plants and not also reproductive endpoints which are inevitably from generally less sensitive mature plants.

To address this question, vegetative endpoints were differentiated by plant age and assigned to two categories: juvenile plants (i.e. all standard lab/greenhouse test data) or vegetative from older plants (generally assessed together with the reproductive endpoints). These two sets of vegetative endpoints were generally assessed separately. However, for some approaches, we merged all vegetative endpoints so that vegetative and reproductive endpoints could be related no matter what age the plants were when the vegetative endpoint was determined.

This paired approach was primarily based on numeric endpoints only. In parallel, assessments were also performed considering censored values.

In addition the data were assessed by an independent evaluator (John W. Green, DuPont) applying different statistical methods. For these the original database was used. Green included also censored endpoints in his comparison of distributions. The main confounding factors that were included as explanatory variables were: substance/formulation tested; mode of action; the measured variable (e.g. shoot height, biomass wet- or dry weight, number of seeds etc.); and effect level (i.e. ER_{10}, ER_{25}, ER_{50}); the test species; higher taxon (e.g. family, monocot/dicot) and its 'anthropinistic affiliation' (i.e. crop species/wild plant species); the test system (i.e. lab/field). Further confounding data were also considered; their subsequent analysis was however impeded by incomplete reporting in the published papers and this is therefore not included in the final statistical analysis. The main analysis focussed on comparing vegetative endpoints of juvenile plants with reproductive endpoints.

Results

A total of 2873 datasets with vegetative endpoints of juvenile species are included, 1058 listing vegetative endpoints of older plants and 1260 reproductive datasets. Because in some instances several effect levels could be retrieved, the numbers of endpoints are higher, and still higher if censored endpoints are also considered. A total of 94 herbicides[18] were included, but matching vegetative and reproductive endpoints (allowing calculation of quotients) were not available for all of them. A total of 428 substance-species combinations appeared in the database, but only 65 of them included numeric reproductive data, 78 when including censored endpoints. In this abstract we focus on the assessment based on numeric endpoints (non-censored) only.

Paired approach (quotients for individual substance-species-combinations)

In the table below we list the overall quotients for different effect-level and endpoint comparisons. As detailed further up, the quotients displayed are means based on the individual quotients of individual substance-species-combinations (SSC). Quotients can either be based on the minima of each substance-species-combination (by experiment and by substance-species-combination), on the geometric means, or by combinations taking the minima by experiment but the geometric mean by substance-species-combination, or the other way round. In the EFSA 2014 Scientific Opinion just one set of endpoints is listed per substance-species combination; the authors do not discuss how they selected the one displayed out of any further endpoints if such were available. We have assumed, the authors of the EFSA Scientific Opinion chose always the lowest endpoint.

[18] This includes a few mixture formulations for which vegetative and reproductive endpoints were available.

Appendix Table 1: Quotients calculated by active substance-species-combination (SSC), comparing vegetative ER_{50} with reproductive ER_{10}, either based on juvenile or on older plants. Quotients based on numeric endpoints (censored values excluded). The leftmost data column is based on overall minima, the rightmost on overall average sensitivity (geometric means at both consolidation steps), those in-between list the outcomes for intermediate approaches. n indicates the number of substance-species combinations with matching endpoint pairs. Blue colour-grades visualize quotients below 1, red colour grades those above.

		Main data: Overall quotients based on					EFSA data	
Variant	Subst.-spec.-comb.	Min	Geo	Min	Geo	n	(one value	n
'f<>' = 0	Experiment	Min	Min	Geo	Geo		per SSC)	
I. Comparing vegetative and reproductive endpoints at the same effect level								
Veg. (juv.) ER_{10} / Reproduct. ER_{10}		1.43	1.33	1.36	1.32	44*	1.43	49*
Veg. (juv.) ER_{25} / Reproduct. ER_{25}		0.84	1.11	0.74	1.18	24	no data	
Veg. (juv.) ER_{50} / Reproduct. ER_{50}		0.97	1.18	1.14	1.38	71*	1.67	38*
Veg. (old) ER_{10} / Reproduct. ER_{10}		1.08	1.12	1.40	1.42	52*	no data	*
Veg. (old) ER_{25} / Reproduct. ER_{25}		1.20	1.48	1.33	1.60	54	no data	
Veg. (old) ER_{50} / Reproduct. ER_{50}		1.45	1.46	1.76	1.71	68*	no data	*
II. Comparing different effect levels (ER_{50} with ER_{25} etc.) for the same parameter								
Veg. (juv.) ER_{50} / Veg. (juv.) ER_{10}		4.38	5.14	4.20	4.91	119	7.20	45
Veg. (old) ER_{50} / Veg. (old) ER_{10}		5.74	5.29	5.61	5.14	68	no data	
Reproduct. ER_{50}/ Reproduct. ER_{10}		5.27	5.51	5.06	5.42	89	6.06	44
Veg. (juv.) ER_{50} / Veg. (juv.) ER_{25}		2.16	2.29	2.08	2.22	235	no data	
Veg. (old) ER_{50} / Veg. (old) ER_{25}		2.78	2.62	2.62	2.53	72	no data	
Reproduct. ER_{50}/ Reproduct. ER_{25}		2.19	2.36	2.05	2.31	55	no data	
III. Simultaneous comparison of vegetative and reproductive endpoints and different effect levels (ER50 with ER25 etc.)"as proposed by EFSA								
Veg. (old) ER_{50} / Reproduct. ER_{10}		6.85	7.11	8.81	9.03	64	no data*	
Veg. (juv.) ER_{50} / Reproduct. ER_{10}		6.25	7.42	7.32	8.68	65	8.34	41
Veg. (old) ER_{50} / Reproduct. ER_{25}		2.10	2.46	2.30	2.60	37	no data	
Veg. (juv.) ER_{50} / Reproduct. ER_{25}		2.32	3.01	2.63	3.69	37	no data	

* EFSA listed 12 vegetative datapoints that we considered as vegetative endpoints of **old** plants.

Comparing vegetative and reproductive endpoints by the same effect level

The quotients obtained by dividing vegetative by reproductive endpoints by the same substance-species-combination (listed in the six uppermost data rows of the table above) indicated clearly that differences were marginal, no matter which effect levels were compared and which consolidation types. Reproductive endpoints were generally lower than vegetative ones on juvenile plants by a factor only slightly greater than one (i.e. 1.11 – 1.43). However, in three cases ratios were below 1 (range: 0.74 - 0.97), indicating that sometimes even the averages of vegetative endpoints for juvenile plants were slightly lower than the averages of the corresponding reproductive endpoints. All the averaging quotients calculated for vegetative endpoints for mature plants and their corresponding reproductive endpoint were above 1 (range: 1.08 – 1.76).

The quotients from the EFSA Scientific Opinion data alone (EFSA, 2014) are fundamentally consistent, being within the same range (1.43 and 1.67).

Comparing different effect levels (ER_{50} with ER_{10} etc.) for the same parameter

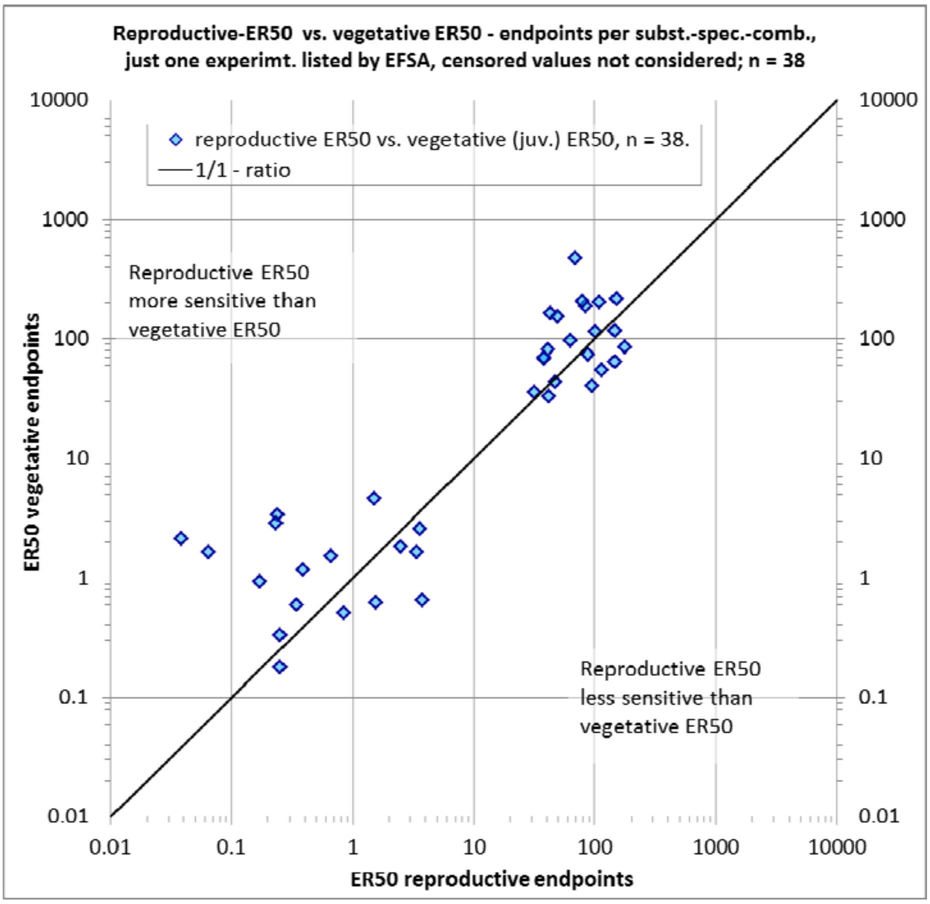
The quotients listed in the following six data rows indicated that ER_{50} were a factor of ca. 5 higher than the corresponding ER_{10} values, no matter which endpoint types they were based on (comparing like-with-like), varying between 4.20 and 5.74. The quotients calculated from the EFSA Scientific Opinion data alone were slightly higher (6.06 and 7.20) but again fundamentally agree. This comparison thus shows that a change of the effect levels has a much larger impact on the overall conservatism than a change from vegetative to reproductive endpoints. Comparisons of ER_{25} and ER_{50} values resulted in lower quotients, as to be expected, ranging between 2.05 and 2.78 when based on numeric endpoints only.

Simultaneous changes (steps I. and II.), as proposed in the EFSA Scientific Opinion

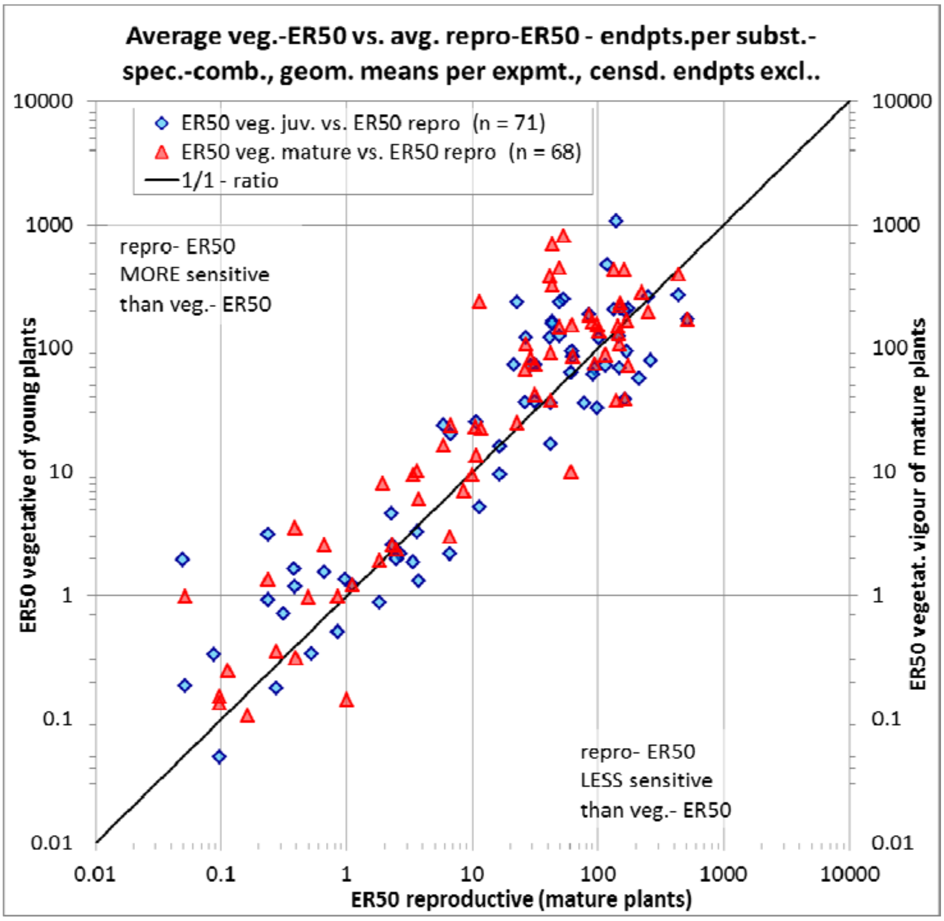
The EFSA Panel proposed moving away from vegetative ER_{50} to reproductive ER_{10}, i.e. changing both the endpoint type and effect level in one step, with the use of an extrapolation factor (EFSA, 2014). Evaluating solely EFSA Scientific Opinion data the resulting quotient was 8.34. Based on the whole data set both steps in one would change the protection level by a factor of between 6.25 and 9.03: a range that includes the estimate based on EFSA-data alone (8.34) so, again, this is fundamentally in agreement with the present analysis. However, in their Scientific Opinion (EFSA, 2014), the EFSA Panel proposed a correction factor of 35 for both steps, claiming increased uncertainty that may require a higher correction factor (they used a 90th percentile). The current data analysis based on a comparison of 5420 numeric data points and 139 quotients (SSC*x) comparing reproductive ERx with vegetative ERx of juvenile plants does not support the need for such a correction factor.

For completeness, we also assessed the average difference between vegetative ER_{50} and reproductive ER_{25} (two lowermost rows of Appendix Table 1. In these cases, quotients ranging from 2.10 to 3.69 were derived. Comparing these quotients to those in the two rows immediately above and with those of the first set of comparisons (i.e. vegetative to reproductive endpoints), it is evident that, where change of endpoint type and change of effect level are combined, by far the largest fraction of the total increase in conservatism is due to the change of the effect level, not due to the change from vegetative to reproductive endpoints.

As example plots we display vegetative and reproductive data pairs on scatter plots, here based on ER_{50} from geometric means per experiment and per substance-species combination. Appendix Figure 3 is based on data from EFSA Scientific Opinion only; Appendix Figure 4 is based on the whole database. When comparing vegetative and reproductive endpoints, all points scatter largely around the 1/1 ratio, further supporting that overall there is no indication for reproductive endpoints being distinctly lower than vegetative endpoints.



Appendix Figure 3: EFSA's data pairs displaying reproductive ER_{50} endpoints (abscissa) and vegetative ER_{50} endpoints of juvenile plants (ordinate), (EFSA listed just one endpoint per experiment; EFSA, 2014), each point signature stands for one substance-species-combination for which both vegetative and reproductive ER_{50} data were available. EFSA considered only numeric endpoints; n VVj ∩ RPo = 38; overall quotient (geometric mean) = 1.67, see Appendix Table 1.



Appendix Figure 4: All data pairs displaying reproductive ER_{50} endpoints (abscissa) and vegetative ER_{50} endpoints of juvenile plants (ordinates), based on geometric means per experiment, geometric means per substance-species combination. Each point signature stands for one substance-species-combination for which both vegetative and reproductive ER_{50} data were available. Only numeric endpoints; n veg. juv. ∩ repro = 71, n veg. mature ∩ repro = 68. Overall quotient (geometric mean) = 1.38, see Apeendix Table 1.

In the report illustrating the analyses made to compare vegetative and reproductive endpoints (Christl, 2017b), matching figures and tables are presented also for evaluations including censored values. Obviously n is higher if these are included, but also the uncertainty of quotients based on them is larger. The resulting patterns and quotients were however quite similar and largely overlap the ranges of quotients and clouds of points displayed here. Hence the overall outcome is the same.

Further assessments were performed at family level, and for individual modes of action. While there were individual families where extreme quotients (based on single or very few substance-species-combination) indicated that reproductive endpoints were much lower than vegetative endpoints, this was only true for individual cases, and it was mitigated by other substance-species combinations of the same family where reproductive endpoints were similar to vegetative endpoints. It was not possible to pinpoint any plant families that would require specific testing for reproductive endpoints.

Analysis at the mode of action level was considered useful to detect any mode of action where reproductive endpoints were regularly lower than the vegetative counterparts. However there was no mode of action that stood out in this respect. The species-substance combinations with extreme differences resulting in very high quotients belonged to AASI (Amino Acid Synthesis Inhibitors), CMD (Cell Membrane Disrupters) and GW (Growth Regulators). However, all these three modes of action also occurred with very low quotients. Therefore no individual mode of action could be recommended for regular testing for reproductive endpoints.

Further statistical analysis (comparison of distributions)

The additional statistical analysis provided by John W. Green, DuPont, found significant differences between vegetative and reproductive endpoints for certain combinations of mode of action and plant family, but again in both directions, and not consistently for all effect levels. Overall this analysis did not show any conspicuous differences in sensitivity between vegetative and reproductive endpoints.

All these comparisons of distributions are based on the species for which endpoints happened to have been reported. The results may thus be affected by chance, particularly sensitive plants could have been tested on vegetative endpoints but other – less sensitive – species of the same family on reproductive endpoints, or vice versa. Results must therefore be interpreted with care.

Observations are most reliable when confirmed by the paired approach (as detailed further up) where vegetative and reproductive endpoints were assessed for each substance-species-combination.

Discussion

The hypothesis that reproductive endpoints are generally distinctly lower than vegetative endpoints was not supported by the databases assessed here. This is in contrast to the perception conveyed by some of the studies included in the database, and also in contrast with the interpretation of the data in the EFSA Scientific Opinion (EFSA, 2014). In this context it is maybe worth considering the potential for a two-fold bias: firstly substance-species-combinations with already low vegetative endpoints are more likely to be tested also for reproductive endpoints than those with higher vegetative endpoints (i.e. the more sensitive ones are re-tested); secondly the trend in scientific publications to present results supporting the tested hypothesis (e.g. reproductive endpoints are lower than vegetative ones), while “null” results not supporting or even contradicting it are less likely to be publishable. This possible lack of publication of null results may be a major source of bias in the published literature (e.g. Sterling, 1959; Zeegers, 2016) but there seems to be no easily applicable remedy. It is up to the scientist to publish both inconclusive and negative data, although this is increasingly difficult as it is less appealing for peer-reviewed journals. Therefore, in a database that is mainly built on published data, we have to consider that it may be biased by the absence of null results. From a regulatory or conservation perspective the bias is towards conservatism, i.e. would rather lead to a false positive than to a false negative. The databases compiled by the EFSA Panel and the databases underlying the reports of Christl (2017a,b) are useful for partially overcoming this bias, as they combine vegetative and reproductive data from different sources.

Based on this initial evaluation and including all data available to date, reproductive endpoints were generally only slightly lower than vegetative endpoints for juvenile plants, i.e. the former were more sensitive by a factor of 0.74 to 1.43. However there were exceptions, in individual cases vegetative and reproductive endpoints were reported to differ by several orders of magnitude. From information contained in the papers assessed it was not possible to explain these cases, neither there was sufficient evidence to exclude them as being not reliable. It must however be assumed that many of the more extreme differences between vegetative and reproductive endpoints (in both directions) might not stand up to experimental re-investigation.

The change of the effect level as proposed in the EFSA Scientific Opinion (EFSA, 2014) would have a much more pronounced impact on the conservatism of the risk assessment than the change from vegetative to reproductive endpoint. Unfortunately the ER_{10}, which is very difficult to measure given the natural variability and accuracy of plant testing, bears much more uncertainty (confidence intervals much wider) and is also more difficult to measure (given the natural variability and accuracy of plant testing) than an ER_{25} or ER_{50} which are more amenable to accurate statistical analysis. We do not think that loss of stochastic certainty should be accepted due to change to a less suitable effect level, just to change the level of protection.

Last but not least, there is doubt that ER_{10} is ecologically relevant. There is ample evidence that reproductive success of typical edge-of-field species is extremely variable, depending on weather, timing and type of agricultural practice such as fertilisation, weeding, inter- and intraspecific competition. A reduction in reproductive success of only 10% would be hard to detect and would be well within the naturally occurring variability.

Therefore, if there is a need for changing the level of protection in the risk assessment for terrestrial non-target plants, an additional assessment factor - while maintaining ER_{50} - or maybe moving to the ER_{25} would be more expedient than a move to an ER_{10} endpoint based on reproduction.

Conclusions

Based on two terrestrial plant databases and the analysis of 5420 endpoints, expected clear differences between ER_{10} and ER_{50} values were detected, but hardly any between reproductive and vegetative endpoints.

The use of reproductive endpoints as the basis for the risk assessment, instead of the currently used vegetative endpoints for juvenile plants, would not increase the margin of safety. In contrast, it would cause a multitude of problems (e.g. invalid or unacceptable data) due to increased complexity and also the absence of a standardised testing methodology.

A move from ER_{50} to ER_{10} as a basis for the risk assessment would increase conservatism but also increase stochastic uncertainty as it is difficult to assess such small effects in plant testing. There is little evidence that ER_{10} endpoints are ecologically relevant for plants.

The combined changes proposed by the EFSA Panel would increase the conservatism of the standard risk assessment for herbicides in the EU by a factor of 6.3 – 9.0, not by a factor of 35, as indicated in the Scientific Opinion (EFSA, 2014). A move to the ER_{25} as basis for the risk assessment could utilise data generated to address US regulatory requirements, increase conservatism and avoid the uncertainty inherent to ER_{10}. Undoubtedly using either the vegetative ER_{50} or ER_{25} as basis for the risk assessment would offer a more robust risk assessment than a change to reproductive ER_{10}. Reasonably adapted assessment factors may be more appropriate to address increase in conservatism than a change to ER_{10} values.

Overall, based on vegetative and reproductive ER_{10}, ER_{25} and ER_{50} endpoints and on the available data sets (vascular plants, largely annual, biennial or perennial herbaceous plants but also some trees), there were no consistent differences in sensitivity between vegetative and reproductive ecotoxicological endpoints when the same effect level is compared (e.g. ER_{50}, ER_{25} or ER_{10}).