RIVM report 601516009/2002

Factsheets for the (eco)toxicological risk assessment strategy of the National Institute for Public Health and the Environment (RIVM), Part II

Editors: R. Luttik and S.M.G.J. Pelgrom

This investigation has been performed by order and for the account of the Board of Directors of the National Institute for Public Health and the Environment, within the framework of project 601516, Risk Assessment of Substances: Science and Market. In addition parts of this investigation have also been performed as part of the RIVM project 650210, Expert advisering carcinogene, mutagene en reproductie toxische stoffen, by order of the Ministry of Health, Welfare and Sport.

RIVM, P.O. Box 1, 3720 BA Bilthoven, telephone: 31 - 30 - 274 91 11; telefax: 31 - 30 - 274 29 71

Authors

Chapter 1: **G.H. Turkstra and M.T.M. van Raaij**
Chapter 2: **M.T.M. van Raaij**
Chapter 3: **T. van der Velde-Koerts**
Chapter 4: **S. Ciarelli and M.H.M.M. Montforts**
Chapter 5: **B.J.W.G. Mensink, C.E. Smit and F.M.W. de Jong**

Abstract

Five factsheets describing risk assessment methods used at the Centre of Substances and Risk assessment (CSR) of the National Institute for Public Health and the Environment (RIVM) are presented here with the main aim of promoting greater transparency in the risk assessment methods used at the Institute in general and within the Centre in particular. The factsheets, listed below, reflect a state-of-the-art approach; they are also meant to function as a platform for discussion.

1. $Alpha_{2u}$-globulin associated nephropathy and renal-cell neoplasms
2. Follicular thyroid tumours in rodents
3. Pesticide residue analysis in plant and animal products
4. Sediment risk assessment for pesticides
5. How to evaluate and use ecotoxicological field tests for regulatory purposes

The first 3 factsheets are related to human risk assessment and the last 2 to environmental issues. Remarks, omissions or additional information sent to the editors (first name) will be appreciated.

Preface

This report was written within the framework of the project ‘Risk Assessment of Substances: Science and Market’. The results as presented in this report have been discussed by members of the human and environmental peer review groups of the Centre of Substances and Risk assessment (CSR), and in some cases experts were consulted, all are acknowledged for their contribution. These members and experts are: M.E. van Apeldoorn, R.A. Baumann, P. van Beelen, R.B. Beems, J. Janus, A.G.A.C. Knaap, F.X.R. van Leeuwen, J.B.H.J. Linders, R. Luttik, W.C. Mennes, M.H.M.M. Montforts, P. van Noort, B.C. Ossendorp, M.T.M. van Raaij, W. Slooff, G.J.A. Speijers, H. Stevenson, T.P. Traas, P.W. Wester, G.J. Schefferlie, G.J.A. Speijers, A.G.A.C. Knaap, A. van Wezel and P. van Zoonen.

Contents

Samenvatting

In dit rapport worden 5 factsheets gepresenteerd, die voor de risicoschatting van stoffen binnen het Centrum voor Stoffen en Risicobeoordeling (CSR) gehanteerd worden. De eerste 3 factsheets hebben betrekking op de humane risicoschatting en de overige 2 factsheets op risicoschatting voor het milieu.
In de factsheet "**Alpha$_{2u}$-globulin associated nephropathy and renal-cell neoplasms**" wordt de toxicologische relevantie van α_{2u}-globuline-gerelateerde nefropathie en het optreden van niertumoren in de mannelijke rat beoordeeld.
Sommige chemische stoffen zijn in staat om een specifieke nefropathie te induceren gerelateerd aan een accumulatie van α_{2u}-globuline in de nier. Deze nefropathie is geassocieerd met het optreden van niertumoren. Er zijn aanwijzingen dat deze nefropathie specifiek is voor de mannelijke rat.
In de factsheet "**Follicular Thyroid Tumours in Rodents**" wordt de relevantie van folliculaire schildkliertumoren in knaagdieren besproken. Het vóórkomen van dergelijke tumoren gaat, met name in ratten, vaak gepaard aan een fysiologische verstoring van de hypothalamus-hyposfyse-schildklier as door bepaalde stoffen. Hormonale veranderingen en de verhoogde activiteit van de schildklier kunnen zorgen voor een toename in schildkliertumoren. In deze factsheet wordt ingegaan op de diverse mechanismen die kunnen leiden tot fysiologische verstoring, verschillen in gevoeligheid tussen rat en mens voor dergelijke effecten en op basis van welke gegevens een uitspraak kan worden gedaan met betrekking tot relevantie van schildkliertumoren in de mens.
In de factsheet "**Pesticide residue analysis in plant and animal products**" worden criteria gegeven wanneer een analysemethode en de daarmee verkregen analyseresultaten als valide worden beschouwd voor residubeoordelingen van bestrijdingsmiddelen.
In de factsheet "**Sediment risk assessment for Pesticides**" wordt een risicobeoordelingsmethode beschreven waarmee de effecten van bestrijdingsmiddelen op sedimentbewonende organismen kunnen worden ingeschat. Naast een beschrijving van de methoden voor de bepaling van de blootstellingsconcentraties in het sediment en de standaard toxiciteitstesten wordt een beslisboom gepresenteerd waarin wordt aangegeven wanneer bepaalde testen moeten worden uitgevoerd en hoe hiermee het risico kan worden bepaald.
De factsheet "**How to evaluate and use ecotoxicological field tests for regulatory purposes**" biedt een overzicht van de momenteel aanwezige leidraden op het gebied van het uitvoeren en evalueren van veldstudies. Om de evaluatie van studies te ondersteunen wordt ingegaan op een aantal specifieke technisch inhoudelijke punten. De minimum eisen waaraan een studie moet voldoen worden behandeld, de onderdelen die bijdragen aan de wetenschappelijke betrouwbaarheid worden geïdentificeerd en er worden criteria voor de evaluatie aangedragen. Speciale aandacht wordt gegeven aan de interpretatie van de resultaten, de statistische onderbouwing van de conclusies en het afleiden van eindpunten. Verder wordt ingegaan op de bruikbaarheid van testen voor de risicobeoordeling en wordt een toelichting gegeven op manier waarop een eindpunt voor deze risicobeoordeling kan worden gebruikt.

Summary

This report presents 5 factsheets for the risk assessment methods used in the Centre for Substances and Risk assessment (CSR). The first 3 of these factsheets are dealing with issues related to human risk assessment and the other 2 with environmental risk assessment.
The factsheet "**Alpha$_{2u}$-globulin associated nephropathy and renal-cell neoplasms**" discusses the toxicological relevance of α_{2u}-globulin-associated nephropathy and renal tumours in male rats for human risk assessment. Some chemicals have been found to induce a specific nephropathy associated with the accumulation of α_{2u}-globulin. This nephropathy has been related to the occurence of kidney tumours. There is evidence that indicates this nephropathy as specific to the male rat.
In the factsheet "**Follicular Thyroid Tumours in Rodents**" the relevance of follicular thyroid tumors in rodents is discussed. The presence of such tumors is, primarily in rats, related to a physiological disturbance of the hypothalamus-petuitary-thyroid axis upon exposure to chemical substances. Hormonal changes and an increased activity of the thyroid may result in an increased occurrence of thyroid tumors. In this factsheet it is discussed which mechanisms of action for physiological disturbances and which species differences in sensitivity are present, and on the basis of which information a conclusion on the relevance for humans can be established.
In fact sheet "**Pesticide residue analysis in plant and animal products**" criteria are provided when an analytical method and the analytical results produced by this method are considered as valid for pesticide residue assessments.
In the factsheet "**Sediment risk assessment for Pesticides**" a risk assessment method is described for estimating the risk of plant protection products to sediment-dwelling organisms. After a description of the methods for calculating an exposure concentration in the sediment and the standard toxicity tests for sediment dwelling organisms a tiered approach is presented for the risk characterisation. Criteria are presented when to carry out a certain type of toxicity tests and how the results will be used in the risk characterisation.
The factsheet "**How to evaluate and use ecotoxicological field tests for regulatory purposes**" provides risk evaluators and assessors with an overview of the existing guidance on the performance and evaluation of ecotoxicological field tests. Specific technical guidance is given to facilitate the evaluation of field studies in the process of pesticide registration. The minimum package of requirements to which a field test should comply is discussed, the test items that contribute to the overall scientific quality are identified and evaluation criteria are listed. The statistical substantiation and the interpretation of test results and the derivation of a suitable endpoint are considered. Criteria to determine the usefulness of a test for the purpose of the risk assessment are given and the way an endpoint can be used for risk assessment is discussed.

Introduction

One of the main tasks of the Centre for Substances and Risk assessment (CSR) of the National Institute of Public Health and the Environment (RIVM) is to assess the risk of compounds on public health and the environment. To carry out risk assessments it is of the highest importance that adequate and up-to-date risk assessment methods are available. Some of these methods are taken over (adopted) from other organisations, but many are, for a large part, developed within the RIVM. These risk assessment methods are not rigid procedures but can be adapted based on new/developing scientific information, possible triggered by questions from policy makers or by developments in national or international organisations. For specific problems or gaps in the assessment of (eco)toxicological effects, 'factsheets' are written by employees of CSR in co-operation with experts. In these factsheets the assessment strategy of RIVM/CSR is described. After adoption of the factsheet by the advisory board and the head of the laboratory of CSR all employees of CSR have to follow the risk assessment method described in the factsheet.

In 2001 the first eight factsheets were published in the RIVM report 601516007 (Factsheets for the (eco)toxicological risk assessment strategy of the National Institute of Public Health and the Environment, edited by Luttik and Van Raaij).

In the report of 2002 five new factsheets are presented (3 factsheets related to public health issues and 2 factheets related to environmental issues):
Factsheets concerning public health
1. Alpha$_{2u}$-globulin associated nephropathy and renal-cell neoplasms
2. Follicular thyroid tumours in rodents
3. Pesticide residue analysis in plant and animal products

Factsheets concerning the environment
4. Sediment risk assessment for pesticides
5. How to evaluate and use ecotoxicological field tests for regulatory purposes

We hope that by publishing these factsheets, the risk asessment methods followed by RIVM/CSR will become more transparent. The authors of each factsheet have tried to describe the state of the art of their subject. Remarks, omissions or supplementary information will be appreciated and can be send to Robert.Luttik@RIVM.NL and will be passed on to the responsible authors.

1 Alpha$_{2u}$-globulin associated nephropathy and renal-cell neoplasms

Factsheet FSV-006/00 date 13-04-2001

Authors:

G.H. Turkstra and M.T.M. van Raaij.

1.1 Introduction and problem definition

For the evaluation of toxic substances, the potential effects after chronic exposure are usually determined in chronic toxicity studies with rodents. Potential effects after chronic exposure also include the possibility of the substance to induce carcinogenic effects. The occurrence of some types of tumours in rodents and their relevance for human risk assessment are sometimes the subject of extensive debate [1]. For example, exposure to certain chemicals results in nephropathy and renal adenoma/carcinomas in male rats, while in female rats and mice of either sex no renal effects are found [2, 3]. The incidence in nephropathy and tumours in male rats seems to be related to α_{2u}-globulin, a small protein which is only observed in male rats and not in female rats, mice of either sex, or in humans. α_{2u}-Globulin is an urinary globulin and not a micro (μ) globulin, as the widespread used term $\alpha_{2\mu}$-globulin would suggest [4]. This latter designation, which has sometimes been used for the urinary globulin, should be avoided. Examples of chemicals that induce α_{2u}-globulin-associated renal tumours in male rats are for instance unleaded gasoline [3] and d-limonene [5]. This factsheet establishes the toxicological relevance of α_{2u}-globulin-associated nephropathy and renal tumours in male rats for human risk assessment.

1.2 Background information and mechanism of α_{2u}-globulin

Mechanisms for tumour induction
In the rodent kidney, several mechanisms of chemically induced carcinogenesis have been identified. These mechanisms can be categorised as follows (according to Hard, 1998 [6]):

A. Direct DNA reactivity.
Some genotoxic substances, particularly certain N-nitroso compounds (or their metabolites), are known to interact directly with DNA of renal tubule cells, causing genomic alterations resulting in carcinomas [7, 8].

B. Indirect DNA damage mediated by oxidative stress.
At least two compounds (potassium bromate [9] and ferric nitrilotriacetate [10]) have been shown to generate reactive oxygen species in rodent kidneys. Reactive oxygen species may cause genomic alterations resulting in carcinomas. Thus, potassium bromate and ferric nitrilotriacetate cause DNA damage via an indirect mechanism [6].

C. Sustained stimulation of cell proliferation.
A number of chemicals appear to induce the development of renal cell tumours through a process involving prolonged renal tubule cell injury coupled with regenerative cell proliferation. This mechanistic pathway can be further subcategorised in two mechanisms:
- Cytotoxicity induced directly by a chemical itself (e.g. chloroform [11]).
- Indirect cytotoxicity resulting from the impairment of a physiological process, induced by a chemical. This is the proposed mechanism for α_{2u}-globulin nephropathy and the associated renal carcinogenesis.

Kidney tumours induced by chemicals in category C tend to occur with a low incidence (usually less than 30%, even at high doses), with a long latency, and may exhibit sex-dependent differences and in some cases, species-specificity. This is in contrast to chemicals

representing mechanistic categories A and B that can induce high (up to 100%) incidences of renal tumours which may have relatively short latent periods, and are not necessarily sex-specific.

Normal physiology of α_{2u}-globulin
In male rats, α_{2u}-globulin is primarily produced in the liver under the stimulus of testosterone [12]. The molecular weight of α_{2u}-globulin is approximately 18.5 kDa. In male rat kidneys, α_{2u}-globulin, as well as other naturally occurring low-molecular-weight proteins, is transferred from the plasma into the urine by glomerular filtration. The proteins are then partially reabsorbed from the glomerular filtrate into the renal tubule cells of the kidney where they are eventually broken down. Of the proteins excreted by male rats, approximately 35% is α_{2u}-globulin. α_{2u}-Globulin is a member of a superfamily of proteins that bind and transport small hydrophobic molecules. Many proteins of this superfamily are synthesised in mammalian species, including humans. It is unknown whether a specific endogenous ligand exists for α_{2u}-globulin and its specific physiological function is presently unknown.

α_{2u}-Globulin-associated nephropathy
Investigations regarding impairment of normal α_{2u}-globulin physiology (by chemicals) have been done mainly with d-limonene, 2,2,4-trimethyl pentane, mineral oils and unleaded gasoline [12]. The agents identified so far as α_{2u}-globulin-inducers are non-genotoxic and do not depend on direct genetic injury for the production of renal tumours. When some chemicals bind reversibly and non-covalently to α_{2u}-globulin, it appears that a complex is formed which is more resistant to lysosomal degradation than the unreacted protein itself. This may result in lysosomal accumulation of α_{2u}-globulin in the P2 segment of the proximal tubule.

The induction of α_{2u}-globulin-associated nephropathy by certain chemicals progresses through a specific, time-dependent, sequence of pathological changes.

- Within 24 hr of dosing these compounds, rapid accumulation of hyaline droplets is observed in proximal tubule cells (under the microscope seen as spherical inclusions in the cytoplasm). These droplets contain α_{2u}-globulin and are the first morphologic manifestations of α_{2u}-globulin nephropathy.
- After 5 days of continuous chemical exposure, the next characteristic lesions occur, namely single-cell necrosis and exfoliation in the P2 segment epithelium.
- Following 3 to 6 weeks of continuous chemical exposure, granular casts accumulate which are formed from cellular debris. Subsequently, tubule dilation occurs at the junction of the P3 segment and the thinner loop of Henle. Furthermore, enhanced cell replication in response to cell death can be seen as increased cell division or as increased DNA synthesis demonstrated by labelling techniques.
- After prolonged chemical exposure (such as in chronic laboratory animal studies), tubule hyperplasia, linear mineralisation in the renal papilla (possibly representing remnants of debris from disintegrating granular casts), and renal tubular epithelial cell tumours are eventually observed.

However, if treatment is stopped after the first 3 weeks of exposure, recovery will occur and normal renal architecture will be restored. A schematic representation of the changes initiated by the binding of a chemical to α_{2u}-globulin is shown in Figure 1.1 [12].

Several agents that induce α_{2u}-globulin-associated nephropathy have been shown to promote both spontaneously and chemically initiated preneoplastic and neoplastic lesions in tubule epithelial cells of the male rat kidney [13]. Furthermore, a relationship between sustained renal-cell proliferation and the promotion of preneoplastic and/or neoplastic lesions has been established, providing support for the conclusion that sustained cell proliferation is causally related to the development of renal tumours in male rats [13]. Appendix A shows a list of agents which have been shown to result in accumulation of α_{2u}-globulin in renal tubule cells.

The pivotal role of α_{2u}-globulin in the nephropathy observed in male rats has been shown in several ways:

- It has been determined that rats of the NCI-Black Reiter strain do not synthesise α_{2u}-globulin in the liver, do not develop α_{2u}-globulin nephropathy and are not susceptible to renal tumour promotion by agents that induce renal tumours in other common rat strains [14].

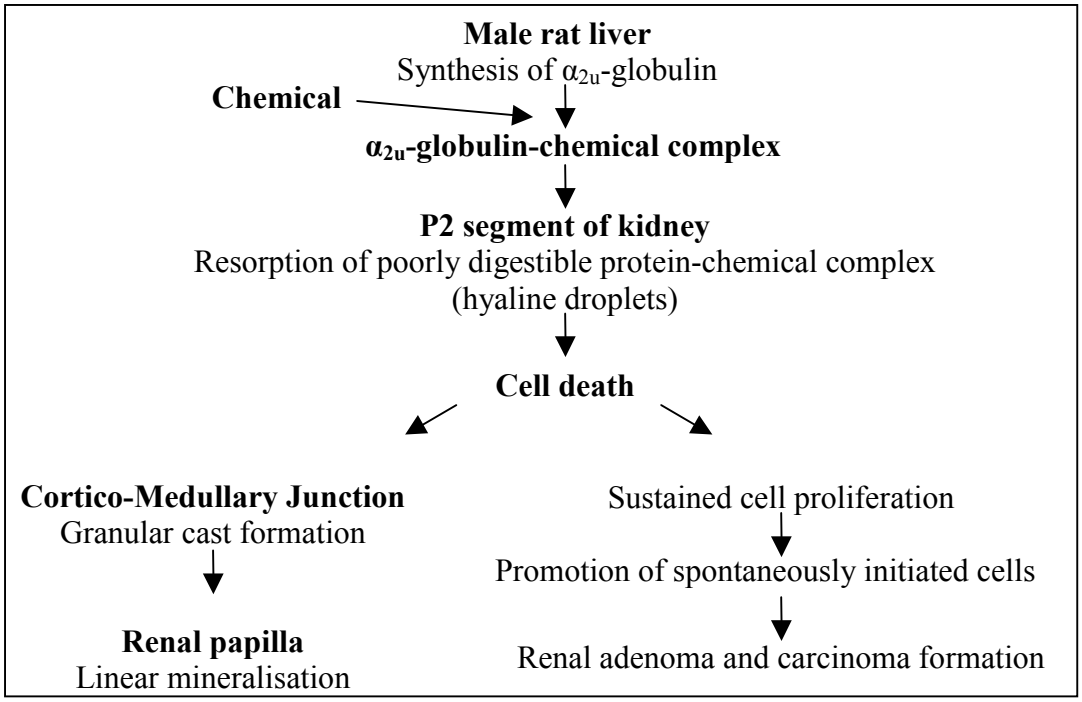


Figure 1.1: Schematic representation of the continuum of changes initiated by the binding of a xenobiotic to α_{2u}-globulin (according to Swenberg and Lehman-McKeeman, 1999 [12]).

- Mice do not synthesise α_{2u}-globulin and are therefore resistant to renal toxicity following exposure to agents that induce α_{2u}-globulin-associated nephropathy in male rats. However, when mice are genetically engineered to synthesise α_{2u}-globulin, these mice indeed develop the nephropathy when they are exposed to agents that induce α_{2u}-globulin-associate nephropathy in male rats [15].

Structure-activity relationship

Lipophilic compounds are capable of binding to α_{2u}-globulin within a binding site pocket of specific dimensions [12]. Attempts have been made to establish structure-activity relationships in order to explain or predict the potential of substances to induce α_{2u}-globulin-accumulation. Such predictions have proven to be of value only for a limited range of

compounds (i.e. aliphatic or alicyclic hydrocarbons). However, there is a diversity of compounds that produce α_{2u}-globulin accumulation and no structure-activity relationships have yet been made for other groups of compounds [16, 17].

1.3 Normal values and variation

Renal tumours
Historical control data for renal tumour incidences in untreated mice, rats, and hamsters are presented in Table 1.1 These data indicate that the incidence as well as variation in spontaneous renal tumours in rats and mice are low.

In humans, renal cell adenocarcinoma/ carcinoma is regarded as a serious disease, however it is not one of the most common neoplasms. The incidence rates of renal-cell carcinoma in Northern Europe, Australia and North America are 0.006-0.008 % in men and 0.003-0.004 % in women [20]. Incidence rates have been increasing since the 1950s in most populations although the rate of increase seems to be slowing down in some countries and levelling off in a few others (see also paragraph 4, interspecies differences).

***Table 1.1**: Incidences of spontaneous renal tumours in different species:*

Species	Strain	Sex	Spontaneous incidence (%)	Type of tumours	References
Rat	various strains	m f	<0.2 <0.1	renal parenchymal	[18]
	Fisher 344	m	0-2 0-6	transitional cell tubule cell	[19]
		f	0-2 0-2	transitional cell tubule cell	[19]
Mice	various strains	m f	<0.2 <0.1	renal parenchymal	[18]
	$B6C3F_1$	m	0-2 0-2	transitional cell tubule cell	[19]
		f	0 0-2	transitional cell tubule cell	[19]
Hamster	various strains	m/f	<2.7	renal parenchymal	[18]

m/f=male/female

In humans, renal cell adenocarcinoma/ carcinoma is regarded as a serious disease, however it is not one of the most common neoplasms. The incidence rates of renal-cell carcinoma in Northern Europe, Australia and North America are 0.006-0.008 % in men and 0.003-0.004 % in women [20]. Incidence rates have been increasing since the 1950s in most populations although the rate of increase seems to be slowing down in some countries and levelling off in a few others (see also paragraph 4, interspecies differences).

In humans, 75-85% of the renal carcinomas are renal cell carcinomas and 10-15% are papillary renal carcinomas. Traditional cytohistological criteria for classification of renal carcinomas are replaced by definitions based on cytogenetic and molecular features. It is clear that renal carcinoma is not a homogeneous entity, and that different mechanisms probably apply to each distinct tumour category.

α_{2u}-Globulin
Low levels of α_{2u}-globulin become detectable in the male rat liver under the stimulus of testosterone at 35-40 days, reaching maximum levels in adult rats at 60-80 days of age [21]. Adult male rat kidneys reabsorb about 30 mg of α_{2u}-globulin per day. Due to development of hepatic insensitivity to androgen during ageing, hepatic synthesis of α_{2u}-globulin begins to fall gradually in male rats some time after 5 months of age. By 22 months of age, there has been a drop of over 90 percent, with α_{2u}-globulin being virtually undetectable in senescent animals. Urinary excretion of α_{2u}-globulin reflects the same age-related trends as synthesis in the liver [22]. Female rats synthesise less than 0.3 mg of α_{2u}-globulin per day and no α_{2u}-globulin is detected in the female rat kidney.

1.4 Susceptible species / subpopulations

Intraspecies differences
Male rats produce α_{2u}-globulin and develop nephropathy and renal tumours under influence of certain chemicals. The production of α_{2u}-globulin has been demonstrated in several rat strains, for instance in Sprague-Dawley, Fisher 344, Buffalo and Brown Norway rats [23]. As discussed in paragraph 3, pre-pubertal and aged males show negligible amounts of α_{2u}-globulin. Accordingly, administration of either decalin to immature male rats [24] or unleaded gasoline to aged, 26-month-old male rats [25] failed to produce renal cortical α_{2u}-globulin accumulation or an increase in hyaline droplets. In female rats, α_{2u}-globulin is undetected in the kidney and females are insensitive for chemicals that bind to α_{2u}-globulin in male rats. Females do not develop renal tumours when exposed to these chemicals.

Interspecies differences
α_{2u}-Globulin-associated hyaline droplet accumulation and renal tumours have not been observed in mice of either sex [24]. In addition, α_{2u}-globulin-renal tumour inducers have been tested for toxicity in guinea pigs (decalin), dogs (d-limonene), and monkeys (unleaded gasoline). No renal pathological changes (including no α_{2u}-globulin accumulation) were observed in these species at doses known to cause α_{2u}-globulin accumulation and renal tumours in male rats [24].

With respect to the α_{2u}-globulin superfamily, it has been shown that the mouse urinary protein, which is most similar to α_{2u}-globulin, does not contribute to a similar syndrome in mice. Furthermore, the lack of a response in female rats, which synthesise many other proteins of this superfamily, demonstrates that these proteins are unlikely to contribute to the renal toxicity. The most abundant α_{2u}-globulin superfamily protein in human kidney and plasma is α_1-acid glycoprotein, and this protein does not bind agents that induce α_{2u}-globulin nephropathy in rats. Human urinary protein is predominantly a species of high molecular weight, and the protein content of human urine is very different from that of rat urine, as humans excrete very little protein (about 1% of the urine concentration in male rats) [26]. The last argument is often used to state that α_{2u}-globulin-associated nephropathy and renal tumours are irrelevant for humans. However, this argument is not valid since the α_{2u}-globulin mechanism is not related to urinary composition, but to 1) resorbed protein fraction and 2) the persistence of the chemical-protein complex.

Adequate data regarding the susceptibility of humans to the induction of renal tumours induced by the specific chemicals that induce α_{2u}-globulin-associated tumours in male rats

are lacking [20]. Regarding unleaded gasoline, cohort studies of refinery workers, truck drivers and gas station attendants have found either no increase or a very moderate non-significant increase in risk of renal cell carcinoma [20]. Cohort studies are usually not able to separate the effect of gasoline from that of other hydrocarbons and other substances that may affect risk of renal cell carcinoma such as asbestos or cigarette smoking. Case-control studies, which can adjust for the effects of other factors, have given either negative results or found a non-significant increase in risk but a trend with duration of exposure was often absent. Furthermore, no studies have looked at leaded and unleaded gasoline separately [20]. No epidemiological study has focused on *d*-limonene, but diet studies have not suggested an association between intake of fruit juice or citrus fruit, which contain *d*-limonene, and renal tumours [20]. Because *d*-limonene is currently being evaluated as a cancer chemo-preventive agent, it should be possible to establish whether this agent produces nephrotoxicity in humans at the high doses used in those clinical trials.

1.5 Additional aspects

Alternative hypothesis

An alternative hypothesis for α_{2u}-globulin-associated nephrotoxicity is that, instead of the accumulation of α_{2u}-globulin itself, the protein functions only as a carrier to transport the bound chemical into renal proximal cells [27]. In this way α_{2u}-globulin serves only to accumulate the specific chemical in the renal cells. Slow release of the ligand from the accumulated ligand-α_{2u}-globulin-complex and/or subsequent metabolism of the released ligand may eventually produce cytotoxicity. In this case, α_{2u}-globulin may only cause a left-shift in the renal cancer dose-response curve for such α_{2u}-globulin-binding ligands in male rats relative to responses in female rats or mice of either sex. This implies that when female rats or mice of either sex would be given very high concentrations of these α_{2u}-globulin-binding ligands, which will achieve the same ligand concentrations in renal tubule cells as in male rats, nephrotoxicity and/or renal tumours would be observed indeed. However, one such a compound, 2,4,4-trimethyl pentane, has been tested *in vitro* at high concentrations and not found to be cytotoxic in primary cultures of renal tubule fragments [28]. At present, there are no data supporting this alternative hypothesis. Anyway, it can be concluded that a process involving α_{2u}-globulin as a vector for chemically induced injury would still remain exclusive to the male rat.

Data gaps
Although the mechanism by which certain compounds induce α_{2u}-globulin-associated nephropathy and renal tumours seems to be clear, other factors might be involved in the etiology of α_{2u}-globulin-associated nephropathy and renal tumours.

- Insufficient information is available whether the process, as described in paragraph 2, is valid for all types of chemicals, because the theory is mainly based on research for three chemicals: d-limonene, 2,4,4-trimethyl pentane, and unleaded gasoline.
- It is not clear whether ligand binding is necessary for α_{2u}-globulin-associated nephropathy and renal tumours. α_{2u}-Globulin accumulation may arise by mechanisms unrelated to ligand binding to this protein:
  - α_{2u}-Globulin accumulation without binding is observed with 2,2,4-trimethyl-pentanoic acid (a metabolite of 2,4,4-trimethyl pentane) [29] and leupeptin (an inhibitor of lysosomal proteolysis) [30].
  - Increases in α_{2u}-globulin accumulation in male rats treated with potassium bromate [31] or Fe-NTA [32] may occur secondary to oxidative damage rather than protein binding.
- The relationship between α_{2u}-globulin binding and renal tumours is not directly linear. Some compounds, with weak binding affinity for α_{2u}-globulin, cause hyaline droplet accumulation and induce kidney tumours in male rats but give rise to very small increases in renal concentrations of α_{2u}-globulin.
- There is a lack in knowledge regarding relationships between the various intermediate steps. This knowledge will improve prediction of the carcinogenic response of chemicals operating through the α_{2u}-globulin-associated mechanism.
- There are major quantitative and qualitative differences between male rats and humans in the amounts of protein excreted in urine. However, little is known concerning the relative quantities of low-molecular-weight proteins that are normally filtered by the human glomerulus and reabsorbed by the renal tubules for catabolism (see also paragraph 4, interspecies differences).
- Insufficient information is available regarding the potential binding of α_{2u}-globulin-inducers to other low-molecular-weight proteins in humans. However, the absence of binding of some of these chemicals to other proteins of the superfamily, suggests, but does not conclusively demonstrate that toxicity in humans could not occur via this mechanism.

1.6 Assessment and CSR Strategy

Based on the analysis in the previous paragraphs, two conclusions can be drawn: 1) The sequence of events proposed to link α_{2u}-globulin accumulation to nephropathy and renal tubule tumours in the male rat is scientifically plausible. 2) The α_{2u}-globulin-associated neoplastic response following chemical administration appears to be unique to the male rat. Even though closely related proteins are present in other species, there is no evidence that these species respond in a similar manner as the male rat with respect to α_{2u}-globulin associated renal tumours.
Therefore, the male rat kidney response to chemicals that induce α_{2u}-globulin accumulation is probably not relevant to human risk assessment.

Both U.S. EPA (1991 [3]) and IARC (1999 [2]) have discussed the risk assessment of α_{2u}-globulin-associated nephropathy and renal tumours in male rats. The U.S. EPA approach is that *only some* criteria (2 and 3; numbers corresponding with the criteria as stated below)

must be fulfilled and some additional information (no 1, 4-6, biochemical data, and SARs) could be added. The IARC states that *all* criteria (no's 1-6) must be fulfilled in order to state that the renal tumours observed in male rats are not relevant for human risk assessment. The CSR strategy is that renal-cell tumours in male rats by agents are not considered of carcinogenic hazard to humans when these agents comply with the following criteria.

Essential criteria:

1. *Non genotoxic.*
   The agent and / or metabolites lack genotoxic activity based on an overall evaluation of in-vitro and in-vivo data. When a substance is genotoxic, the renal tumours are considered relevant for risk assessment.

2. *Induction of the characteristic sequence of histopathological changes in rat studies.*
   The abnormal accumulation of hyaline droplets in the P2 segment of the renal tubule is necessary to attribute the nephropathy and / or renal tumours to the α_{2u}-globulin sequence of events. The finding helps to differentiate α_{2u}-globulin-inducers from chemicals that produce nephropathy and / or renal tumours through other processes.
   The induction of α_{2u}-globulin-associated nephropathy by certain chemicals progresses through a specific, time-dependent, sequence of pathological changes.
   - Within a few days of dosing chemicals, rapid accumulation of hyaline droplets is observed in proximal tubule cells (microscopically seen as spherical inclusions in the cytoplasm).
   - After about one week of continuous chemical exposure, the next characteristic lesions may occur, namely single-cell necrosis and exfoliation in the P2 segment epithelium.
   - Following 3 to 6 weeks of continuous chemical exposure, granular casts accumulate which are formed from cellular debris. Subsequently, tubule dilation occurs at the junction of the P3 segment and the thinner loop of Henle. Furthermore, enhanced cell replication in response to cell death can be seen as increased cell division or as increased DNA synthesis demonstrated by labelling techniques.
   - After prolonged chemical exposure (such as in chronic laboratory animal studies), tubule hyperplasia, linear mineralisation in the renal papilla (possibly representing remnants of debris from disintegrating granular casts), and renal tubular epithelial cell tumours are eventually observed.

   If the response is mild, not all of these lesions may be observed. However, some elements, including hyaline droplets, consistent with the sequence of pathological changes must be demonstrated to be present.
   When an agent induces at a certain dose renal tumours without α_{2u}-globulin-associated nephropathy, the renal tumours might be the result of a different process and therefore these renal tumours are considered to be relevant.

3. *Identification of the protein accumulating in tubule cells as α_{2u}-globulin.*
   Hyaline droplet accumulation is a non-specific response to protein overload in the renal tubule and is not necessarily related to α_{2u}-globulin. Therefore, it is necessary to demonstrate that α_{2u}-globulin is identified in the hyaline droplets found in the male rat.

4. *Male rat specificity for nephropathy and renal tumours.*
   The α_{2u}-globulin-associated renal nephropathy and tumours are specific to the male rat. Positive responses in the renal tubule in female rats imply that α_{2u}-globulin does not account solely for the renal pathology response in the male rats. Thus when nephropathy

and/or renal tumours are found in female rats, these pathological changes are considered to be relevant for risk assessment.

When the toxicological data of compounds comply with all four criteria above, the nephropathy and / or renal tumours in the male rat are considered to be not relevant for human risk assessment.

Furthermore, when additional information is available this information must be included in making an overall conclusion for human risk assessment as follows.

Additional information:

5. *No nephropathy or renal tumours were induced in other species than the rat*
   If studies in other species than the rat are available, nephropathy or renal tumours should not be observed in these studies.
   Positive responses in the renal tubule in mice of either sex, or any other laboratory animal imply that α_{2u}-globulin alone does not account for the renal tubule tumour response in the male rats. Thus, when nephropathy and / or renal tumours are found in mice of either sex, or any other laboratory animal all renal effects, including those in rats, are considered to be relevant for risk assessment.

6. *Reversible binding of the chemical or metabolite to α_{2u}-globulin.*
   Binding of an agent to α_{2u}-globulin can be shown in different ways, e.g. with *in vitro* and *in vivo* studies. These *in vitro* and *in vivo* data can help characterise a chemical as one that induces accumulation of α_{2u}-globulin. This could be specified through complex formation of the chemical and α_{2u}-globulin. For evaluation of the results regarding reversible binding of a chemical to α_{2u}-globulin a case by case approach should be applied.

If a compound is considered to induce α_{2u}-globulin-associated nephropathy, the nephropathy and possibly associated effects[1] are considered not to be toxicologically relevant. The NOAEL in such a study is based on other toxicological relevant endpoints. Therefore, the α_{2u}-globulin-associated nephropathy should not be used for setting a toxicological unit value (e.g. ADI) and should not be used for human risk assessment.

When an agent induces tumours at other sites in the male rat or in other laboratory animals, the relevance of these tumours should be evaluated independently of the α_{2u}-globulin-associated renal tumours.

References

1. Alison RH, Capen CC, Prentice DE. 1994. Neoplastic lesions of questionable significance to humans. Toxicol. Pathol. 22: 179-186.
2. Capen. C.C., Dybing E., Rice J.M.,Wilbourn J.D. 1999. Species differences in thyroid, kidney and urinary bladder carcinogenesis. IARC Sci.Publ. 147
3. U.S. EPA. Risk Assessment Forum. Alpha$_{2u}$-globulin: Association with chemically induced renal toxicity and neoplasia in the male rat. Report no. EPA/625/3-91/019F. dd. 1991.
4. Roy A.K., Neuhaus D.W. 1966. Identification of rat urinary proteins by zone and immunoelectrophoresis. Proc. Soc. Exp. Biol. Med. 121: 894-899.

[1] The severe nephropathy might eventually result in changes in other parameters, e.g. changes in body weight or urine volume. Whether these changes are considered to be related to the nephropathy and therefore considered to be not toxicologically relevant, should be evaluated on a case by case approach.

5. IARC. 1999. d-limonene. Some chemicals that cause renal or urinary bladder tumours in rodents, and some other substances. Monogr. Eval. Carcinogen Risk Hum. 73: 307-327.
6. Hard G.C. 1998. Mechanisms of chemically induced renal carcinogenesis in the laboratory rodent. Toxicologic Pathology 26: 104-112.
7. Hard G.C. (1990) Tumours of the kidney, renal pelvis and ureter. In: Pathology of Tumours in Laboratory Animals. Second Edition (Turusov V.S. and Mohr U., eds), Vol. 99, 301-344, IARC Sci. Publ., Lyon.
8. Solé M., Cardesa A., Domingo J., Mohr U. 1992. The carcinogenic effect of 2,2-dioxopropylnitrosamine on the renal pelvic epithelium of Sprague-Dawley rats, after chronic subcutaneous injections. J. Cancer Res. Oncol. 118: 222-227.
9. Umemura T., Sai K., Takagi A., Hasegawa R., Kurokawa Y. 1990. A possible role for oxidative stress in potassium bromate ($KBrO_3$) carcinogenesis. Carcinogenesis 16: 593-597.
10. Toyokuni S., Uchida K., Okamoto K., Hattori-Nakakuki Y., Hiai H.,Stadtman E.R. 1994. Formation of 4-hydroxy-2-nonenal-modified proteins in the renal proximal tubules of rats treated with a renal carcinogen, ferric nitrilotriacetate. Proc. Natl. Acad. Sci. USA 91: 2616-2620.
11. Hard G.C., Boorman G.A., Wolf D.C. 2000. Re-evaluation of the 2-year chloroform drinking water carcinogenicity bioassay in Osborne-Mendel rats supports chronic renal tubule injury as the mode of action underlying the renal tumor response. Toxicol. Sci. 53: 237-244.
12. Swenberg J.A., Lehman-McKeeman L.D. 1999. Alpha 2-urinary globulin-associated nephropathy as a mechanism of renal tubule cell carcinogenesis in male rats. IARC Sci. Publ. 147: 95-118.
13. Dietrich D.R., Swenberg J.A. 1991. The presence of α_{2u}-globulin is necessary for d-limonene promotion of male rat kidney tumors. Cancer Res. 51: 3512-3521.
14. Dietrich D.R., Swenberg J.A. 1991. NCI-Black-Reiter (NBR) male rats fail to develop renal disease following exposure to agents that induce alpha-$_{2u}$-globulin (α_{2u}) nephropathy. Fund. Appl. Toxicol. 16: 749-762.
15. Lehman-McKeeman L.D., Caudill D. 1994. *d*-Limonene-induced hyaline droplet nephropathy in α_{2u}-globulin transgenic mice. Fund. Appl. Toxicol. 23: 562-568.
16. Barratt M.D. 1994. A quantitative structure- activity relationship (QSAR) for prediction of α-2μ-globulin nephropathy. Quant. Struct.-Act. Relat. 13: 275-280.
17. Barratt M.D. 1994. A quantitative structure-activity relationship (QSAR) for prediction of $\alpha_{2\mu}$-globulin nephropathy. Toxic. in Vitro Vol. 4: 885-887.
18. WHO , red. 1991. Environmental Health Criteria 119: Principles and methods for the assessment of nephrotoxicity associated with exposure to chemicals. WHO-Geneva: 152-153.
19. Haseman J.K., Elwell M.R. 1996. Evaluation of false positive and false negative outcomes in NTP long-term rodent carcinogenicity studies. Risk Anal. 16: 813-820.
20. Mellemgaard A. 1999. Human renal-cell carcinoma - epidemiological and mechanistic aspects. IARC Sci. Publ. 147: 69-80.
21. Roy A.K., Chatterjee B. 1983. Sexual dimorphism in the liver. Ann. Rev. Physiol. 45: 37-50.
22. Neuhaus O.W., Flory W. 1978. Age-dependent changes in the excretion of urinary proteins by the rat. Nephron 22: 570-576.
23. Ridder G.M., Von Bargen E.C., Alden C.L., Parker R.D. 1990. Increased hyaline droplet formation in male rats exposed to decalin is dependent on the presence of α_{2u} globulin. Fundam. Appl. Toxicol. 15: 732-743.
24. Alden C.L., Kanerva R.L., Ridder G. and Stone L.C. (1984) The pathogenesis of the nephrotoxicity of volatile hydrocarbons in the male rat. In: Advances in Modern Environmetal Toxicology (Mehlman M.A., Hemstreet G.P., Thorpe J.J. and Weaver N.K., eds), Vol. VII, 107-120, Princeton Scientific Publishers, Inc., Princeton, USA.
25. Murty C.V.R., Olson M.J., Garg B.D., Roy A.K. 1988. Hydrocarbon-induced hyaline droplet nephropathy in male rats during senescence. Toxicol. Appl. Pharmacol. 96: 380-392.
26. Olson M.J., Johnson J.T., Reidy C.A. 1990. A comparison of male rat and human urinary proteins: implications for human resistance to hyaline droplet nephropathy. Toxicol. Appl. Pharmacol. 102: 524-536.
27. Melnick R.L., Kohn M.C. 1999. Possible mechanisms of induction of renal tubule cell neoplasms in rats associated with α_{2u}-globulin: role of protein accumulation versus ligand delivery to the kidney. IARC Sci. Publ. 147: 119-137.
28. Borghoff S.J., Lagarde W.H. 1993. Assessment of binding of 2,4,4-trimethyl-2-pentanol to low molecular weight proteins isolated from kidneys of male rats and humans. Toxicol. Appl. Pharmacol. 119: 228-235.
29. Charbonneau M., Lock E.A., Strasser J, Cox M.G., Turner M.J., Bus J.S. 1987. 2,2,4-trimethylpentane-induced nephrotoxicity. I. Metabolic disposition of TMP in male and female Fischer 344 rats. Toxicol. Appl. Pharmacol. 91: 171-181.
30. Olson M.J., Mancini M.A., Garg B.D., Roy A.K. 1988. Leupeptin-mediated alteration of renal phagolysosomes: similarity to hyaline droplet nephropathy of male rats exposed to unleaded gasoline. Toxicol. Lett. 41: 245-254.

31. Sai K., Tyson C.A., Thomas D.W., Dabbs J.E., Hasegawa R., Kurokawa Y. 1994. Oxidative DNA damage induced by potassium bromate in isolated rat renal proximal tubules and renal nuclei. Cancer Lett. 87: 1-7.
32. Uchida K., Fukuda A., Kawakishi S., Toyokuni S., Hiai H., Ikeda S, Horio F. 1995. Acute nephrotoxicity of a carcinogenic iron chelate. Selective inhibition of a proteolytic conversion of α_{2u}-globulin nephropathy. FEBS Lett. 357: 165-167.
33. IARC. 1999. dichlorobenzenes. Some chemicals that cause renal or urinary bladder tumours in rodents, and some other substances. Monogr. Eval. Carcinogen Risk Hum. 73: 223-276.
34. WHO, red. 1995. Environmental Health Criteria 174: Isophorone. WHO-Geneva:
35. Caldwell D.J, Eldridge S.R., Lington A.W., McKee R.H. 1999. Retrospective evaluation of alpha 2u-globulin accumulation in male rat kidneys following high doses of diisononyl phthalate. Toxicol. Sci. 51: 153-160.
36. IARC. 1999. Methyl tert-butyl ether. Some chemicals that cause renal or urinary bladder tumours in rodents, and some other substances. Monogr. Eval. Carcinogen Risk Hum. 73: 339-383.

Appendix

Chemicals identified to induce renal nephropathy and accumulation of α_{2u}-globulin in the P2-segment of the tubule in male rats[A].

Chemical	All IARC criteria fulfilled[B]	References
d-Limonene	x	[5]
Unleaded gasoline	x	[12]
2,4,4-trimethyl pentane (TMP) + 2,4,4 trimethyl-2-pentanol	x	[12]
Sodium barbital		[12]
Diethylacetyl urea		[12]
1,4 dichlorobenzene + 2,5 dichlorophenol	x	[33]
Isophorone	x	[34]
3,5,5-Trimethyl-hexanoic acid derivatives		[12]
Decalin		[12]
1-Decolone		[12]
2-Decalone		[12]
Tetrachloroethylene (perchloroethylene)		[12]
Pentachloroethane		[12]
C10-C12 isoparaffinic solvent (saturated aliphatic hydrocarbons)		[12]
Levamisole		[12]
Gabapentin		[12]
Tridecyl acetate		[12]
Diisononyl phthalate (DINP)		[35]
Isopropylcyclohexane		[12]
Methyl tert-butyl ether	x	[36]

A. Chemicals that produce renal tumours in female rats or in mice are not included. Furthermore, if an alternate mechanism has been established, as in the case of potassium bromate-induced oxidative stress, it is not included. This list has not the intention to be complete and is for illustration purposes only.

B. Chemicals identified to fulfil all IARC criteria for α_{2u}-globulin associated renal cell neoplasms in male rats. These criteria include all essential and additional criteria of the CSR strategy.

2 Follicular Thyroid Tumours in Rodents

Factsheet FSV-007/00 date 09-08-2001

Author:

M.T.M. van Raaij

2.1 Introduction

For the evaluation of toxic substances, the potential long-term effects are usually determined in chronic toxicity studies with rodents. Potential effects after chronic exposure also include the possibility of the substance to induce carcinogenic effects. The occurrence of some types of tumours in rodents and their relevance for human risk assessment are sometimes the subject of extensive debate [1,2]. One of these types of tumours are follicular cell tumours in the thyroid gland of the rat. Numerous studies have shown that especially the rat thyroid gland shows a high incidence of proliferative lesions such as hyperplasia and adenomas of follicular cells upon long term exposure to various xenobiotics [9]. Opinions on the relevance of rat thyroid gland tumours have been produced by the HSE [3] (to which an RIVM commentary has been formulated [4), the U.S. EPA [5], and the EU commission group of specialised experts on carcinogenicity [6].
In this fact sheet the occurrence of follicular cell tumours (but not C-cell tumours) and the mechanisms by which they can be induced by chemical substances are described. Whether or not thyroid tumours in the rat are relevant for human risk assessment is established in paragraph 6, the RIVM evaluation strategy.

2.2 Mechanism of action and background information

2.2.1 The Hypothalamus-Pituitary-Thyroid (HPT) axis – Normal physiological function

The thyroid gland in mammals is located just in front of the larynx and its characteristic feature is its ability to concentrate iodide from the bloodstream in order to synthesise the iodide containing hormones triiodothyronine (T3) and thyroxine (T4) [7,8,9].

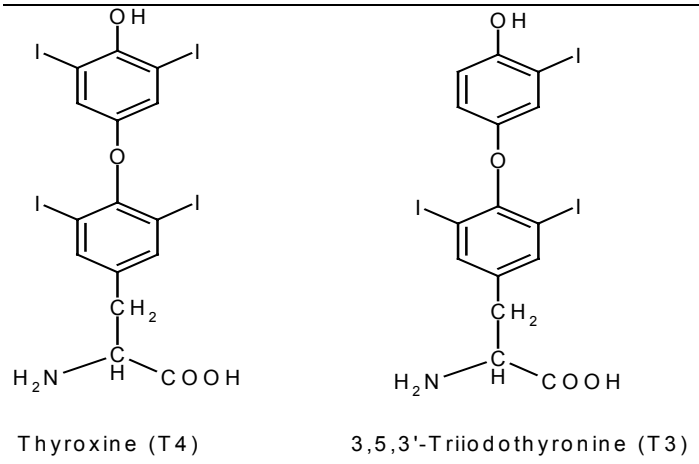


Figure 2.1: Chemical structures of Thyroxine (T4) and Triiodothyronine (T3).

Morphology
The functional unit of the thyroid gland is the thyroid follicle that consists of cuboidal epithelial cells spherically arranged around the lumen of a follicle and surrounded by a basement membrane. The lumen contains colloid almost entirely composed of iodinated glycoprotein called thyroglobulin. In between the follicles, blood capillaries, sympathetic nerve endings and so-called parafollicular C-cells are located. The parafollicular C-cells secrete calcitonin which regulates calcium and phosphorus homeostasis [8].

Thyroid hormone synthesis
The synthesis of thyroid hormones starts by uptake of iodide from the blood by follicle cells by an active transport mechanism. Iodide is subsequently activated in the follicular cells which requires oxidation of inorganic iodide (I^-) to molecular reactive iodine (I_2). This reaction is catalysed by thyroid peroxidase (TPO). Activated molecular iodine is transported to the follicular lumen where it is coupled to tyrosine residues of thyroglobulin. A further step in thyroid hormone synthesis is the coupling of two such iodinated tyrosyl-residues to form thyroxine or triiodothyronine, but these are still bound to thyroglobulin. For secretion, this 'thyroxine-thyroglobulin' is taken up by the follicular cells (colloid droplet formation) and is subsequently broken down into T4, T3, and remaining parts of the thyroglobulin within lysosomes. The thyroid hormones are then secreted into the bloodstream by exocytosis.

Thyroid hormone levels and function
Of the two thyroid hormones, T3 is the most biologically active although (in humans) the total T3 concentration is only about 2% of the T4 concentration [8]. Thyroid hormones in the circulation are in most species bound to plasma proteins such as Thyroid Binding Globulins (TBG) and (pre-)albumin leaving only a small fraction in the free form. Partly because T3 is less firmly bound to carrier proteins than T4, the half-lives of T3 and T4 in man are about 1-3 days and 5-9 days respectively [8,9,10].
Thyroid hormones bind to intracellular nuclear receptors exerting their actions through RNA synthesis and are involved primarily in the regulation of the basal metabolic rate [8].

Metabolism of thyroid hormones
Thyroxine (T4) may be viewed as a 'pro-hormone' which is converted to T3 in various peripheral tissues and to some extent to the biologically inactive reverse T3 (rT3).
Degradation of thyroid hormones occurs primarily in the liver and involves conjugation with glucuronide (mainly T4) or sulphate (mainly T3). These conjugates are excreted via the bile into the intestine. A portion of the conjugated material is hydrolysed in the intestine and the resulting free hormones are reabsorbed into the blood (enterohepatic recirculation), the remainder being excreted through the faeces.

Regulation of thyroid function
Homeostatic control of thyroid function is effected by a sensitive feedback mechanism that responds to changes in circulating levels of T4 and T3. Thyroid Stimulating Hormone (TSH), secreted by the anterior pituitary gland, plays a pivotal role in the regulation of thyroid function. TSH stimulates iodide uptake, thyroid hormone synthesis, iodination of thyroglobulin, and endocytosis and proteolysis of colloid in the epithelial follicular cells of the thyroid gland. The rate of release of TSH by the pituitary is controlled by both Thyrotropin-Releasing-Hormone (TRH) secreted by the hypothalamus and by strong negative feedback of circulating levels of T4 and T3. In figure 2.2 an overview of thyroid function and regulation is presented.

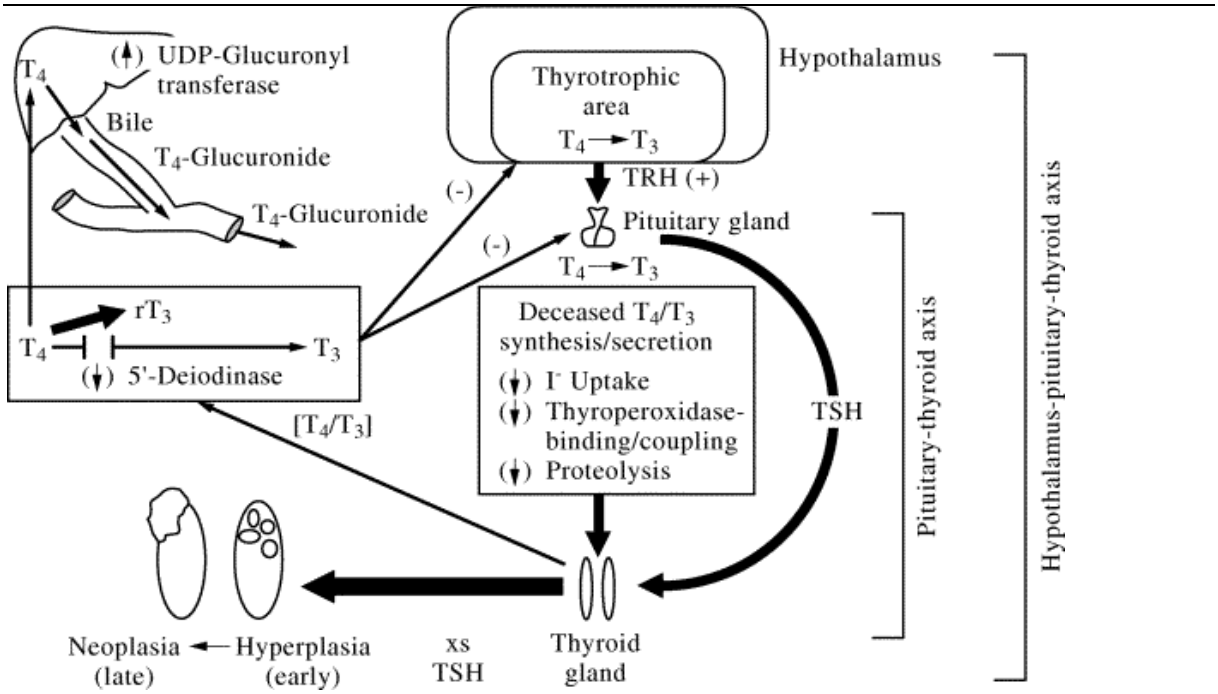


Figure 2.2: The HPT-axis: physiological function and points of disturbance.

2.2.2 The HPT-Axis - points of chemical disturbance

The HPT-axis may be disturbed by xenobiotics at various points [7,9,10,11,13]. In general five levels of disturbance can be observed: 1) inhibition of iodide uptake, 2) inhibition of thyroid hormone synthesis, 3) inhibition of thyroid hormone secretion, 4) stimulation of thyroid hormone degradation, and 5) TSH-receptor mediated effects.

Inhibition of iodide uptake
A number of anions act as competitive inhibitors of iodide uptake, including perchlorate (ClO_4^-) and (iso)thiocyanate (SCN^-). Especially thiocyanate is a potent inhibitor of iodide uptake. These anions have a similar effect on thyroid function as iodide deficiency: the circulating levels of T3 and T4 will decrease.

Inhibition of thyroid hormone synthesis
A number of substances inhibit the incorporation of iodide into the tyrosyl residues of thyroglobulin. Classes of chemicals that inhibit the process of thyroglobulin synthesis include:
1) thionamides (e.g. thiourea, thiouracil, propylthiouracyl, methimazol, carbimazole)
2) aniline-derivates and related compounds (e.g. sulphonamides, p-aminobenzioc acid, p-aminosalicylic acid, amphenone)
3) substituted phenols (e.g. resorcinol, phloroglucinol, tricyanoaminopropene, 2,4-dihydroxy benzoic acid)
4) miscellaneous inhibitors (e.g. aminotriazole, antipyrine, iodopryrine)

Many of these substances act by inhibition of TPO although the coupling of iodinated tyrosyl residues appears to be the most susceptible step in the synthesis of thyroid hormones.

Thiourea acts through a different mechanism, i.e. it stimulates the reduction of I_2 back to I^-, which has eventually the same effect as TPO inhibition.
The goitrogenic effects of sulphonamides on the rat thyroid have been known for about 50 years and it has been shown that sensitive species to the effects of sulphonamides include rat, mouse, dog and swine [11]. Humans, non-human primates, guinea pigs, and chickens have been reported to be resistant to the development of changes in thyroid function by sulphonamides [9,11].
Irrespective of the precise mechanism, inhibition of thyroid hormone synthesis will result in a decrease in circulating levels of thyroid hormones (T3 and T4).

Inhibition of thyroid hormone secretion
Relatively few chemicals inhibit specifically the secretion of thyroid hormones. An excess of iodine is known to result in a lower rate of thyroid hormone secretion. Several mechanisms have been suggested for this effect of high iodide levels including a decrease in lysosomal protease activity (human glands), inhibition of colloid droplet formation (mice and rats), and inhibition of a TSH-mediated increase in cAMP (see below) [10].
Lithium has also been reported to inhibit thyroid hormone synthesis. Its acts by inhibiting colloid droplet formation.
Other xenobiotics (e.g. minocycline) may induce pigmentation in the follicular cells or the colloid. The physico-chemically altered colloid is less able to react with iodine than normal colloid resulting in a decreased thyroid hormone synthesis.

Stimulation of thyroid hormone metabolism/degradation
Some chemicals inhibit the enzyme 5'-(mono)deionidase that converts T4 to T3 in various peripheral tissues, including the pituitary. Via this enzyme the large amount of T4 present in the circulation can be converted to the biologically active T3 in a controlled way. Inhibition of this enzyme results in a decreased T3 level in the peripheral tissues - including the pituitary - resulting in a compensatory increase in TSH levels.
T4 can also be metabolised to the biologically inactive rT3 and a decreased rate of metabolism of T4 to T3 by inhibition of 5'-deiodinase often results in elevated levels of rT3 [7]. An example of a substance inhibiting 5'-deiodinase is FD&C Red no. 3 (=erythrosine) which is widely used as colour additive in foods, cosmetics, and pharmaceuticals.

A number of chemicals disturb thyroid function by their effects on hepatic enzymes (see for extensive reviews on this subject [12,18]). Hepatic phase II biotransformation enzymes play an important role in thyroid hormone economy since glucuronidation and sulfation are the rate limiting steps in biliary excretion of T4 and T3 respectively. Therefore, induction of hepatic phase II biotransformation enzymes may directly lead to an enhanced excretion of T4 and/or T3 resulting in lower circulating levels of thyroid hormones.
Xenobiotics that induce hepatic phase II biotransformation enzymes and disrupt thyroid function in rats include:
1) CNS acting drugs (e.g. phenobarbital, benzodiazepines)
2) Calcium channel blockers (e.g. nicardipine, bepridil)
3) Steroids (e.g. spironolactone)
4) Retinoids
5) Hydrocarbons (e.g. chlordane, DDT, TCDD)
6) Polyhalogenated biphenyls (e.g. PCB, PBB)

Although being potential inducers of thyroid tumours in rats, most of these substance have no apparent carcinogenic activity and produce little or no mutagenicity or DNA damage [7,13].

Modulation of TSH-receptors
Another specific mechanism for influencing thyroid function is modulation of the thyroid TSH receptor. The thyroid response to TSH stimulation is also controlled by autoregulation. Binding of TSH to the TSH receptor results in the activation of the second messenger systems through cAMP and phosphokinase C. Special classes of iodolipids in the thyroid gland attenuate the TSH response. When iodolipids are low (e.g. due to iodide deficiency or inhibition of TPO), attenuation of the receptor response is removed and the thyroid response to TSH is enhanced [3].

2.2.3 Common mechanistic pathway

All the mechanism presented above, have eventually a common mechanistic pathway. Directly or indirectly, circulating levels of thyroid hormones are reduced (either T3, T4 or both) which results in a compensatory increase in TSH secretion by the pituitary. As a consequence, the thyroid gland is stimulated to synthesise and secrete higher amounts of thyroid hormones. Because an increase in circulating TSH levels is the pivotal action in thyroid stimulation, it should be emphasised that TSH is the central marker to be monitored. When this stimulation of the thyroid by TSH is prolonged, 3 phases can be distinguished:
An initial phase (lasting several days): During this phase rapid changes in thyroid morphology occur including resorption of colloid from the follicular lumen, hypertrophy of follicular epithelial cells, and an increase in vascularity.
Second phase of rapid growth: During this phase a sustained increase in thyroid weight and size occurs. Histopathologically follicular hypertrophy and hyperplasia can be detected.
Third phase of accumulation: In this phase the growth of the thyroid slows down as a plateau is reached (increases in thyroid size and weight are limited). Follicular hyperplasia may progress to nodular proliferation of follicular cells and eventually to neoplasia (tumours).

See for an extensive list of substances causing thyroid function disturbances and thyroid tumours, reference no. 19.

2.2.4 Experimental support for the importance of TSH

A prolonged increase in circulating levels of TSH appears to be the pivotal step in chemical-induced thyroid hyperplasia and neoplasia. This central role for TSH is further illustrated by experimental observations.

- For substances that disturb thyroid function, a threshold or a no-effect level on the thyroid gland can be established by determining the dose of the substance that fails to elicit an elevation of circulating TSH levels. In other words, no hyperplasia or increased tumour incidences are observed when TSH concentrations remain at control levels.
- Excessive secretion of TSH alone (i.e. in the absence of chemical exposure) also produces a high incidence of thyroid tumours in rodents. This has been observed in rats fed an iodine-deficient diet and in mice that received TSH-secreting tumour transplants [7,9,10].

- In experiments with phenobarbital (see above, section 2.2), supplemental administration of thyroxine (at doses that returned TSH levels to normal) blocked the thyroid tumour-promoting effects of phenobarbital [10].

2.3 Normal values and variation

In table 2.1 an overview of tumour incidences in non-treated control animals is given for mice, rats, and humans.

Table 2.1. Incidence of thyroid tumours in non-treated mice, rats, and humans.

Species	Type of effect	Normal values (% incidence)	Remarks.	Ref.
Mouse	Follicular tumours	1%	all strains	14
	Follicular tumours	M: 0.15 % (range 0 – 6%) F: 1.94 % (range 0 – 8%)	$B6C3F_1$ strain, NTP program, N=1340	15
Rat	Follicular tumours	<3%	Various strains	14
	Total thyroid tumours (type not specified but presumably incl. C-cell tumours)	5%	Sprague-Dawley, period 1974-1983, N=924	16
	Follicular adenoma Follicular carcinoma	1.3 % (range 0 – 5.3%) 0.2 % (range 0 – 1.5%)	Wistar TNO/W70, 11 exps, N=1533	17
	Follicular tumours	M: 2.08 % (range 0 – 8%) F: 0.89 % (range 0 – 6%)	Fischer 344, NTP program, N=1347	15
Human	Clinically manifest thyroid cancer	0.003%	US population	13
	Thyroid tumours discovered at autopsy (tumour type not specified)	2 %	US population	7

2.4 Susceptible Species / Subpopulations

Interspecies differences
With respect to species differences in thyroid function and potential disturbances of the pituitary-thyroid axis, the focus in the scientific literature has been primarily on differences between rats and humans. Mice and dogs have some form of intermediate position with respect to a number of factors (sometimes appearing to be similar to rats, sometimes to humans). In the following section mainly rat-human comparisons are presented with some additional remarks to other species.

In the blood, thyroid hormones are bound to carrier proteins such as TBG, prealbumin, and albumin [7,9,10]. The binding affinity of TBG for T4 is about 1000 times higher than for prealbumin. In humans, circulating T4 is bound primarily to TBG but this binding protein is not present in rodents, birds and lower vertebrates [7,9,10]. In rats, T4 is bound primarily to albumin and to a lesser extent to pre- and postalbumin. In mice, T4 is carried by both albumin and postalbumin and in dogs by albumin and TBG [7,9,10]. As a consequence of the low binding of T4 in rodents, more thyroid hormone is in the free form and subjected to degradative metabolism. Therefore, the plasma half-life of T4 in rats is considerably shorter (12-14 h) than in man (5-9 days). For T3, these values are 6h (rat) and 24-72h (human)

[7,9,10]. To compensate for the shorter half-lives, the basal levels of TSH are considerably higher in rats (55-65 μU/ml in males; 36-41 μU/ml in females) compared to humans (about 2.5 – 5 μU/ml) [7,13]. It has been suggested that the rat thyroid gland is continuously in a hyperactive state [1]. This is illustrated by the observation that a rat without a functional thyroid requires about 10 times more T4 (20 μg/kg bw) for full substitution than an adult human (2.2 μg/kg bw). The 'hyperactive state' of the rat thyroid is further illustrated by morphological differences in the thyroids of rat and humans. In rats, follicles are relatively small, often surrounded by cuboidal epithelium although it should be noted that thyroid appearance can be influenced by the amount of iodine in the diet. In contrast, humans follicles are normally less active and appear with a large lumen with abundant colloid, surrounded by relatively flattened epithelium [3].
As a consequence of these species differences, it can be concluded that the turnover in thyroid function of the rat (and mouse [1]) is substantially higher compared to humans. Therefore, the rat HPT axis is much more susceptible to physiological disturbances by xenobiotics [3,4,5,7]. The dog has an intermediate status with respect to thyroid activity [1].

Intraspecies differences
Adult male rats have higher circulating TSH levels than females and they are often more sensitive to goitrogenic stimulation and thyroid carcinogenesis [9,13,18]. In addition, follicular cells are often larger in male rats [9]. In humans, there is no sex difference in TSH levels, but females develop thyroid cancer more frequently [13].
In the case of X-radiation (the only demonstrated thyroid carcinogen in humans, see below), children have been shown to be more susceptible. Therefore, when thyroid tumours are the consequence of mutagenic effects, one should consider the possibility that children are more susceptible.

2.5 Additional Observations

A number of observations provide support for a substantial (quantitative) difference between rodents and humans.

- Substances inhibiting TPO activity have been shown to affect thyroid function in sensitive species (mouse, rat, dog) but not or much less in other species (guinea pig, chicken, primates, and humans) [7,9,10]. For example, the IC50 for sulphonamide (concentration necessary for 50% inhibition of TPO) is about 500x higher in monkey compared to rat [1].
- Only few compounds are able to affect circulating TSH levels in humans [1]. Circulating levels of T4 in humans are only affected by very powerful hepatic microsomal enzyme inducers such as rifampicin [1,13]. Little if any effect is found on T3 and TSH levels. Under these circumstances there is no evidence that exposure to such substances may lead to the development of thyroid cancer [10].
- Epidemiological studies, in patients treated with therapeutic doses of phenobarbital (see above, section 2.2), have revealed no indications for an increased risk for thyroid neoplasia [9]. In this respect it may be questioned whether phenobarbital is an enzyme inducer in humans such as it is in rats.
- Conditions which result in a prolonged hyperstimulation of the thyroid gland in humans by increased TSH levels (iodine deficient diet; endemic goitre) provide little if any increase in the incidence of thyroid cancer [10].

- To date, no chemical has been identified as being carcinogenic to the human thyroid gland [13]. The only demonstrated human thyroid carcinogen is X-radiation [3,7], although increased TSH levels may promote the occurrence of X-ray induced thyroid tumours [13].
- Thyroid tumours in humans are histopathologically diagnosed mostly as having a papillary pattern whereas in rodents thyroid tumours are mostly of a follicular pattern. In addition, thyroid tumours (as other tumours) often metastasise in humans but not in rodents [3].
- Increased risk for thyroid neoplasia in humans was indicated only in exceptional circumstances. Patients with dyshormonogenetic goitre (a congenital defect in thyroid hormone synthesis) and patients with Grave's disease (an autoimmune disease) have been indicated (but not proven) to be at greater risk to develop thyroid tumours [11,9,13].

2.6 Assessment and RIVM Strategy

Various organisation or committees have published their views, policies or strategies for interpreting data on thyroid carcinogenesis. The U.S. EPA published a review on this subject in 1989 [7] and an update of their science policy for the assessment of thyroid follicular tumours was released in 1998 [5]. Within the European Union, the HSE has published an opinion on the subject in 1998 [3] which was commented upon by RIVM [4]. These opinions were used in an EU committee group of specialised experts in 1999 [6] which released a decision tree for the classification of substances causing thyroid tumours in rodents (see Appendix). Recently, the IARC published some reviews on thyroid carcinogenesis [19] although clear guidance for the assessment of thyroid tumours were not included.
Roughly, it can be stated that U.S. EPA starts the assessment of thyroid tumours by the default assumption that observations in rats are relevant for humans [5]. The EU committee of specialised experts had to deal with divergent opinions but produced a consensus recommendation on the classification of substances causing thyroid tumours [6, see also Appendix]. In addition, reviews in IARC and the open literature hold the opinion that it is highly unlikely that thyroid tumours occur in humans due to the exposure to non-genotoxic xenobiotics [1,2,9,10,11,14,18,19].

In the following RIVM strategy, the recommendation of the EU specialised expert committee and a previous RIVM opinion on thyroid tumours in rodents [4] are combined.

The following general statements form the (scientific) cornerstones of the RIVM strategy.
- Thyroid tumours in rodents can be induced by either mutagenic effects, by physiological disturbances, or a combination of both.
- Regulation of thyroid function (the HPT-axis) is basically similar in humans and rats.
- With respect to the physiological disturbance of the HPT axis by non-genotoxic xenobiotics, substantial quantitative differences are present between rodents (especially rats) and humans.
- Humans are considerably less sensitive to the development of epithelial follicular thyroid tumours after long-term stimulation than rodents (especially rats).
- A disturbance in the HPT-axis and a concomitant change in thyroid function in humans will trigger various kinds of effects which will urge people to seek medical attention. Therefore, the possible expression of carcinogenicity in humans, caused by prolonged hyperstimulation of the thyroid by TSH, will be a highly improbable possibility.

Assessment of thyroid follicular cell tumours in rats
When a toxicological relevant, dose-related, increase in the incidence of thyroid follicular cell tumours is observed in rat studies, the following steps are made to assess the relevance of the thyroid tumours for human risk assessment.

1. Is the substance genotoxic?

   When the substance is considered to be genotoxic, it cannot be ruled out that genotoxic effects have contributed to the induction of thyroid tumours. Therefore, the observed thyroid tumours are relevant for human risk assessment. In addition, it should be taken into account that children may be more susceptible to thyroid mutagens (see paragraph 4).

2. Does the substance induce a disturbance of the HPT axis in rats?

   To answer this question, information is necessary for all the following aspects. Preferably these aspects are studied at dose levels similar to the lowest-observed-effect-level for carcinogenicity:

   - Evidence for a (histo)pathological sequence of events characteristic for prolonged thyroid stimulation. This includes hypertrophy of follicular cells, increased vascularity, increase in thyroid weight and size, follicular hyperplasia, and eventually nodular proliferation of follicular cells and neoplasia.
   - Evidence for sustained alterations in circulating hormones.
     - It should be demonstrated that exposure to the substance induces elevations in the circulating levels of TSH or in its turnover. Because prolonged (hyper)stimulation of the thyroid by TSH is the ultimate common pathway for the formation of thyroid tumours in rodents, measurement of TSH levels is a prerequisite for a valid evaluation.
     - Additionally, it should be demonstrated that exposure to the substance induces changes in the circulating levels of thyroid hormones such as T4, T3 (and rT3) or in their turnover. Not all substances will affect both T4 and T3 levels. Therefore, measurement of both hormones is necessary for a valid evaluation. However, it should be noted that such changes in hormone levels may occur only temporary because the feedback systems will tend tonoramalise T3 and T4 levels.

     These alterations in circulating hormones (TSH, T3, T4) can be observed relatively quickly after chemical exposure. Therefore, such measurements may be performed in a subacute or semichronic study.
   - Information or experimental evidence on the mode of action by which the HPT axis is disturbed (see categories in paragraph 2.2). The mode of action can be substantiated by general information (e.g. by comparison with other related compounds, SAR's etc…) or by several lines of experimental evidence on the substances itself (e.g. measurement of hepatic phase II biotransformation enzymes involved in thyroid hormone degradation). This should be assessed on a case-by-case basis.

If the substance is non-genotoxic and it has been demonstrated (based on the information indicated under point 2) that the substances induces a prolonged disturbance in the HPT-axis, the thyroid tumours observed in rats are not considered to be relevant for human carcinogenicity risks. This implies that these tumours are not sufficient evidence for considering the substance as potential carcinogenic for humans and hence classification is not indicated. Although the present RIVM strategy does account for it, the EU specialised expert

group included an additional criterion on carcinogenic potency for a final discussion on classification (see [6] and appendix).

In contrast to the risk of carcinogenicity, the disturbance of the HPT-axis itself is considered to be toxicologically relevant for humans since regulation of thyroid function is similar in rats and humans, although the latter are less susceptible to such disturbances. Therefore, disturbance of the HPT-axis is considered to be a hazard indicator for humans and should be taken into account when setting NOAELs and health based limit values. If disturbance in the HPT-axis is the major/critical toxicological endpoint in rats, the interspecies assessment factor to be used for establishing a toxicological limit value may be reduced on a case-by-case basis, because of the fact that humans are substantially less susceptible to disturbances in the HPT-axis than rats. However, the dose levels at which other effects of toxicity become manifest in rats and other species should be taken into account in the derivation of the assessment factor.

Additional remarks:

- Thyroid follicular cell tumours in mice (or possibly dogs) should be evaluated on a case-by-case basis because mice and dogs have an intermediate position with respect to their sensitivity to HPT-axis disturbances and thyroid tumour formation. This will highly depend on the mode of action of the substance (see categories in paragraph 2.2.2).
- C-cell (or parafollicular) tumours are developed from a different group of cells which have a different function then follicular cells. The activity of these cells is regulated by different mechanisms. Therefore, this fact sheet does not cover the assessment of tumours arising from these cells. Such parafollicular tumours should be evaluated separately from follicular tumours.

References

1 Alison, R.H., Capen, C.C., Prentice, D.E. (1994) Neoplastic lesions of questionable significance to humans. Toxicol. Pathol. 22(2); pp. 179-186.

2 Lima, B.S., Van der Laan, J.W. (2000) Mechanisms of nongenotoxic Carcinogenesis and assessment of the human Hazard. Regul. Toxicol. Pharmacol. 32; 153-143.

3 Costigan, M. (1998) The relevance of rat thyroid gland tumours to humans., HSE Toxicology Unit, Bootle, UK.; June 1998.

4 Kroese, E.D., Beems, R.B., Wester, P.W. (1998). Comments to the position paper "The relevance of thyroid gland tumours to humans" by Michael Costigan. National Institute of Public Health and the Enironment (RIVM), Bilthoven, The Netherlands; September 1998.

5 U.S. EPA (1998) Policy for the assessment of thyroid follicular cell tumours. EPA document no. EPA/630/R-97/002; March 1998.

6 EC (1999). European Commission – European Chemicals Bureau. Draft Summary record – Commission group of specialised experts in the fields of carcinogenicity, mutagenicity and reprotoxicity. ECBI/49/99-Add.1 Rev.2; dd. 17-12-1999.

7 Hill, R.N., Erdreich, L.S., Paynter, O.E., Roberts, P.A., Rosenthal, S.L., Wilkinson, C.F. (1989) Thyroid follicular cell carcinogenesis. Fund. Appl. Toxicol. 12; pp. 629-697.

8 O'Riordan, J.L.H., Malan, P.G., Gould, R.P. eds. (1982) Essentials in Endocrinology. Blackwell Scientific Publications, Oxford, UK.

9 Capen, C.C. (1994) Mechanisms of chemical injury of thyroid gland. In: Receptor-mediated Biological Processes: Implications for evaluating carcinogenesis, pp. 173-191. Wiley-Liss, Inc.

10 Capen, C.C. (1997) Mechanistic data and risk assessment of selected toxic end points of the thyroid gland. Toxicol. Pathol. 25(1), pp. 39-48. Special issue on Risk assessment: selected organ and compound-related issues.

11 Capen, C.C. (1998) Correlation of mechanistic data and histopathology in the evaluation of selected toxic endpoints of the endocrine system. Toxicol. Lett. 102-103; pp. 405-409.

12 Curran, P.G., DeGroot, L.J. (1991) The effect of hepatic enzyme-inducing drugs on thyroid hormones and the thyroid gland. Endocrine Reviews 12; pp. 135-150.
13 Hill, R.N., Crisp, T.M., Hurley, P.M., Rosenthal, S.L., Singh, D.V. (1998) Risk assessment of thyroid follicular cell tumours. Environm. Health Perspect. 106(8); pp. 447-457.
14 Thomas, G.A., Williams, E.D. (1999) Thyroid stimulating hormone (TSH)-associated follicular hypertrophy and hyperplasia as a mechanism of thyroid carcinogenesis in mice and rats. In: Species differences in thyroid, kidney and urinary bladder carcinogenesis. Eds. Capen, C.C. et al. IARC scientific publications No. 147, International Agency for Research on Cancer, Lyon, France, 1999.
15 Huff, J. (1999) Chemicals associated with tumours of the kidney, urinary bladder, and thyroid gland in laboratory rodents from 2000 US National Toxicology Program/National Cancer Institute bioassays for carcinogenicity. Appendix 2 in: Species differences in thyroid, kidney and urinary bladder carcinogenesis. Eds. Capen, C.C. et al. IARC scientific publications No. 147, International Agency for Research on Cancer, Lyon, France, 1999.
16 Bode, G., Hartig, F., Hebold, Czerwek, H. (1985) Incidence of spontaneous tumours in laboratory rats. Exp. Path. 28; pp. 235-243.
17 Bomhard, E., Karbe, E., Loeser, E. (1986) Spontaneous tumors of 2000 Wistar TNO/W.70 rats in two year carcinogenicity studies. J. Environ. Pathol. Oncol. 7; pp. 35-52.
18 McClain, R.M. (1989) The significance of hepatic microsomal enzyme induction and altered thyroid function in rats: Implications for thyroid gland neoplasia. Toxicol. Pathol. 17; pp. 294-306.
19 IARC (1999) Species differences in thyroid, kidney and urinary bladder carcinogenesis. IARC scientific publications no. 147. Eds. Capen, C.C., Dybing, E., Rice, J.M., Wilbourn, J.D. International Agency for Research on Cancer, WHO, Lyon, FRance.

Appendix

(Report of the EU committee of specialised experts in the fields of carcinogenicity, mutagenicity and reprotoxicity, Meeting at Arona 1-2 Sept. 1999).

Non-genotoxic thyroid carcinogens in the rodent bioassay

The **Pesticides Working Group** at their meeting in May '99 had initiated and the **CMR WG** at their meeting in June '99 had confirmed the request to the **Specialised Experts** for guidance on the appropriate general approach when undertaking classification of non-genotoxic rodent thyroid carcinogens. Due to short notice, no lead country had been asked, but the **ECB** had collected the relevant documentation with the help and contributions from several **Member States**. A background document drafted by the UK HSE in 1998 and intended to stimulate the discussion within the **Working Group** and to assist in the classification procedure, had already been available (ECBI/22/98 – Add. 1 Rev. 1). This document was taken as the starting point for the issues to be addressed by the **Specialised Experts.**
Comments on the UK HSE draft document for consideration by the **Specialised Experts** had been received from several Member States delegations. Furthermore, recently published relevant evaluations by other groups of experts were tabled, such as the risk assessment for rodent thyroid tumorigenicity by the US EPA, and an evaluation of species differences in thyroid carcinogenesis by IARC.

The **Specialised Experts** confirmed the UK HSE summary of the current scientific knowledge relating to roughly three different mechanisms leading to thyroid tumour formation in rodents, namely genotoxicity, elevation of circulating TSH by several modes of interference with the thyroid-pituitary axis, and local cell proliferation. In addition, substances lacking mechanistic data were described as a fourth group to be considered.

On the question of relevance for humans, the **Specialised Experts** were certain that genotoxicity always has to be regarded relevant and classification of a genotoxic substance forming tumours of the thyroid in rodents with Carc. Cat. 2 or 3 is warranted. Due to controversial databases on genotoxicity for some substances, interpretation problems were anticipated.

Other mechanisms could potentially be species-specific, with humans being less sensitive or not susceptible. Some **Experts** stressed that goitre is proven to represent a preneoplastic change in rodents whereas in humans equivalent evidence is only available for goitre produced in children by a genotoxic agent (observations following exposure to ionising radiation after the Chernobyl accident were cited). Overall, the available data suggest that an increase of TSH in humans does not pose a significant concern regarding potential thyroid carcinogenesis in humans. Therefore, these **Specialised Experts** questioned the significance of non-genotoxic goitrogens for carcinogenesis in humans and defined the difference in susceptibility between rodents and humans to be potentially of a qualitative and not merely quantitative kind. In addition, no contradictory evidence was available for humans. Consequently, if such a mechanism has been identified, no classification was considered to be a possible option. Other **Experts** agreed in that some investigations showed humans to exhibit a much lower sensitivity to substance-induced TSH elevation by disturbance of the pituitary-thyroid axis than rats. However, they remarked that not many studies on human sensitivity are available and definitive studies on whether such agents do or do not produce tumours in the human thyroid are lacking. Furthermore, they said, the substances currently under discussion for thyroid carcinogenicity in rodents by the **Working Group** act by different mechanisms, and it was difficult to have an opinion on all of them in this context. Some **Experts** stressed, knowledge on exposure levels to humans was often insufficient or based on unreliable assumptions. For these **Specialised Experts** an answer in support of a species-specific mechanism not relevant for humans could not be given due to insufficient data.

One **Specialised Expert** presented the draft of a decision chart to the Group summarising the considerations relevant to decide on a classification for thyroid carcinogenicity. The majority of **Experts** welcomed the chart concept as a useful tool for finding a practical common approach although the draft was recognised to require

further elaboration by the Group. Another opinion was that such a chart could be helpful for risk assessment considerations but not for hazard assessment, which was requested from the Group of **Specialised Experts**. One **Specialised Expert** expressed strong dissatisfaction with the limited and superficial discussions possible within one meeting day and the attempt to summarise such a complicated issue in a simple flow chart, with so few facts known on the relevance for humans. He urged the discussion time to allow for correct, science-based decisions. For practical considerations, as an effort to try to avoid lengthy substance-by-substance discussions in a future **Specialised Experts** meeting, he could accept elaboration of a transparent scheme to guide the **Working Group** in drawing consistent classification conclusions. In relation to the chart concept, one **Expert** made the point that the approach advocated by IARC was noted but that it did not help the **Working Group** in making classification and labelling decisions about industrial chemicals and pesticides because for chemicals data from exposed humans would rarely be available. The IARC model seemed more appropriate for pharmaceuticals.

Further development of the structured 'considerations chart' was undertaken during the remaining time of the meeting. To the question whether knowledge of a threshold for TSH elevation in humans was needed for prediction of the human tumour hazard – as foreseen in the original draft chart, another **Expert** remarked that substance-dependent TPO inhibition could already and should be routinely tested in humans. Finally, the derivation of such a threshold concentration for TSH elevation in human tissue was disregarded as too difficult in practice for each compound in question. Similarly, consideration to test and use the results from species more suitable for extrapolation to humans, like the dog or pig, were disregarded for practical reasons. In the end, for some **Specialised Experts** the disregard of the insufficient knowledge on, but nevertheless inclusion of an assumed substance-dependent effect threshold in humans remained a crucial deficiency of the drafted scheme.

Chronic, steady-state exposure to the substance in question was defined as a starting assumption. Another principle suggested was to focus on mechanisms involving TSH elevation and, secondly, whether TSH elevation induced goitre and subsequently tumours in rodents, thereby including risk-associated assumptions of an effect threshold, which would be considerably higher in humans. Other **Specialised Experts** did not welcome this approach but preferred to start with the general presumption that rodent thyroid tumorigens 'may produce thyroid tumours in humans', in line with the IARC and US EPA approach, and then look at the mechanistic information.

As to the uncertainties in the science, one opinion was that TSH elevation alone, without knowledge of the thyroid hormone levels in relation to it, could not provide any information on the tumour initiating lesion. One **Specialised Expert** drew attention to results from recent research that enhancement of thyroid hormone clearance via induction of conjugating enzymes in the liver (e.g. UDPGT) hardly caused an increase in TSH levels. Thus no satisfactory mechanistic explanation for the thyroid tumours induced by liver enzyme inducing agents, such as phenobarbital, could be given and such substances were not covered in the flow chart. Another **Expert** reported on other recently detected inconsistencies, like the increase of TSH by propylene thiouracil, without being crucial for tumour formation. Direct local toxicity on the thyroid was suggested as possible causative factor for tumour formation by some substances. There was agreement that different mechanisms, identified and non-identified ones, could act concomitantly with one mechanism predominantly responsible for thyroid tumour formation. The possibility of local genotoxicity in the thyroid in the absence of genotoxicity demonstrated in *in vivo* standard tests could also never be excluded. Following these examples, the **Specialised Experts** emphasised that the chart in development could not provide advice for each individual substance, and depending on the available evidence decisions should be made on a case by case basis.

Finally, the **Specialised Experts** reached consensus on the list of biological events representative of initiating mechanisms leading to a disturbance of the thyroid-pituitary axis, which was included in the draft 'considerations chart'.

On the additional question, namely how to include the lower sensitivity of humans if one of these non-genotoxic mechanisms was operative, and how to express the varying degree of concern adequately, some **Specialised Experts** raised as a compromise the use of carcinogenic potency as developed and published in the 'Guidelines for setting specific concentration limits for carcinogens in Annex I of Dir. 67/548 - inclusion of potency considerations'. This was supported by the majority of **Specialised Experts**. The majority opinion, furthermore, was that non-genotoxic substances causing thyroid tumours in rodents by disturbance of the thyroid-pituitary axis with low and medium potency, if the mechanism was clearly established, should usually not be classified. If the dose-response indicated high potency, category 3 was most appropriate. Although there was a final agreement on this compromise, some **Experts** maintained that no classification even in case of high potency

was appropriate. Other **Specialised Experts** held the view that category 3 should always be considered, even for substances exhibiting low and medium potency, as the various mechanisms lead to different potential responses in humans. Some of these **Experts** felt additionally that high potency rodent-thyroid carcinogens in individual cases might merit classification with category 2. As a general point, the **Specialised Experts** stressed the importance of consideration of the total weight of evidence.

The **Specialised Experts** reached a consensus on the contents of a simplified chart (presented in Annex I), on the understanding that it was a compromise in order to provide transparency and that it did not reflect the wide range of different opinions voiced in this forum. Although they stressed that the scheme could not be strictly followed in all cases in question, the chart was still regarded to be useable as a general guidance and to assist in achieving consistency in classification decisions. The necessary flexibility and case-by-case considerations should not be precluded. It was also not meant to be comprehensive. Guidance in case of decisions on substances with unknown mechanism and insufficient data was included.
Finally, a few theoretical examples of evidence difficult to fit into the scheme and requiring expert judgement were discussed.

The importance of hazard classification was raised as an additional aspect with a bearing on the sequence of activities within the EU risk assessment programme. In case of no classification for a certain endpoint, there was the danger that no risk assessment would be performed for this endpoint in an appropriate way. No classification might imply a wrong message in this context. - On the other hand, it was remarked, such a decision tree could be very useful in the ESR (Existing Substances Regulation) programme to identify which additional information (test) is needed to improve the assessment of a risk.

Conclusion:

The **Specialised Experts** agreed that there is convincing scientific evidence that humans are considerably less sensitive than rodents (especially rats) regarding
(i) perturbation of thyroid hormone homeostasis induced by non-genotoxic xenobiotics
(ii) development of epithelial thyroid tumours after long-term exposure to such agents.
Furthermore, a majority of the **Specialised Experts** emphasised their view that, at least for point (i), these interspecies differences are of a quantitative rather than a qualitative nature. Some **Specialised Experts** underlined their opinion that the possibility of a qualitative difference rat-human still exists, since there is yet no convincing evidence that chemical-induced prolonged thyroid stimulation might cause thyroid-tumours in humans.

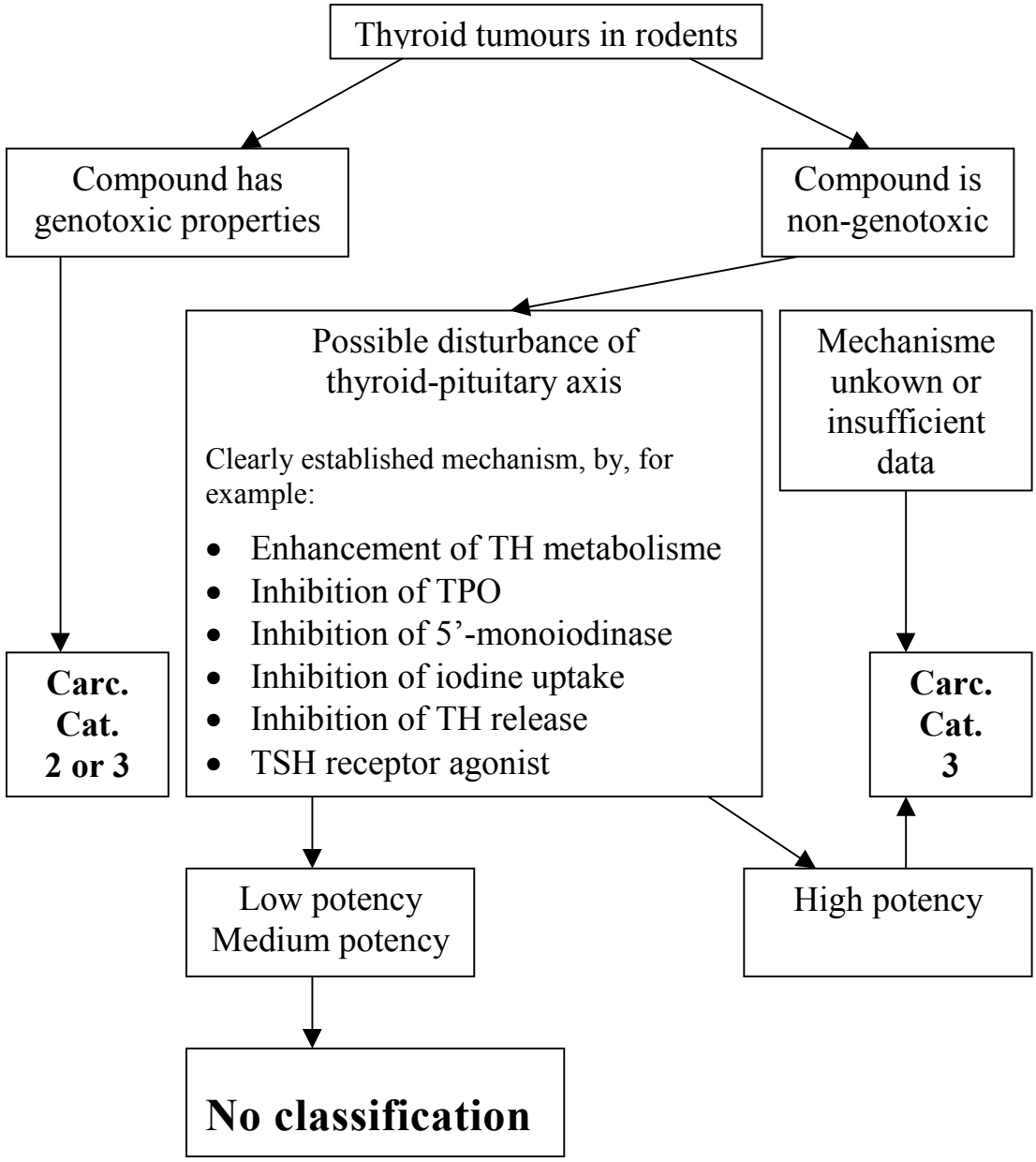
Consensus was reached by the **Specialised Experts** that classification in category 3 or even no classification at all may be appropriate, provided that detailed studies have shown, that an individual non-genotoxic substance causing tumours in rodent thyroid also induces persistent stimulation of the thyroid. For a distinction between classification in category 3 and no classification for carcinogenicity the tumourigenic/ carcinogenic potency of the agent should be considered. The human relevance of the specific mechanism(s) underlying the perturbation of the pituitary-thyroid hormone axis should also be taken into account. The **Specialised Experts** took note of results from recent research that enhancement of thyroid hormone clearance via induction of conjugating enzymes in the liver (e.g. UDPGT) may not give a satisfactory mechanistic explanation for the thyroid tumours induced by liver enzyme inducing agents such as phenobarbital.

Overall, the **Specialised Experts** reached a consensus on how to deal with rodent thyroid tumours in the classification for carcinogenicity. A diagram is attached (Annex I), which summarizes the views of the **Specialised Experts**: Essentially, it was agreed that non-genotoxic carcinogenic substances producing thyroid tumours in rodents with low or medium potency by a clearly established perturbation of the thyroid hormone axis, in general, do not need to be classified. Other rodent thyroid carcinogens merit classification in either category 2 or 3.

The **Specialised Experts** were aware that such a general recommendation may not be applicable to every substance in question and that decisions should be made on a case by case basis.

CLASSIFICATION OF SUBSTANCES CAUSING THYROID TUMOURS IN RODENTS

SPECIALISED EXPERTS recommendation



Abrreviations used:
TH thyroid hormone; TPO thyroid peroxidase; TSH thyroid stimulating hormone

3 Pesticide residue analysis in plant and animal products

Factsheet FSV-008/00 date 18 February 2002

Author:

T. van der Velde-Koerts

3.1 Introduction and aim

An analytical method is the series of procedures from receipt of a sample to the production of the final result. Validation is the process of verifying that a fully described analytical method is fit for the intended purpose.

At reviewing residue analytical methods the reviewer is often confronted with the problem that although certain validation requirements have to be evaluated according to EU or FAO documents, no criteria have been defined for these validation requirements or the validation requirements are amenable to more than one interpretation. This could lead to different views by alternative reviewers.
Furthermore there is no common opinion on the consequences when a certain residue analytical method is considered not valid.

National and EU assessments drafted in the scope of Council Directive 91/414/EEC are used for pesticide registration purposes and could have legal implications if the EU guidelines are not followed strictly. Therefore new requirements for national and EU assessments may not be introduced. On the other hand, assessments drafted for the Joint FAO/WHO Meeting on Pesticide Residues (JMPR), which is an ad hoc scientific advisory body to the Codex Committee on Pesticide Residues (CCPR), are meant to reflect the latest scientific views. Therefore in this case there is more freedom to deviate from the FAO Manual and new requirements for FAO assessments may be introduced.

The aim of this fact sheet is
- to define criteria for validation requirements, where these are not available (§2.2),
- to explain those requirements for which more than one interpretation is possible (§2.2),
- to introduce, to modify or to extend FAO requirements where necessary (§2.2),
- to indicate the consequences if a certain residue analytical method does not fulfil the criteria for validation requirements (§2.3 and §2.4).

3.2 Assessment of residue analytical methods

3.2.1 Introduction

Residue analytical methods can be divided in two groups:

Residue analytical methods for pre-registration purposes
These methods generate residue data which are used for setting maximum residue limits (MRLs), supervised trials median residues (STMRs), highest residues (HRs), processing factors and storage stability periods. These methods are laboratory specific and only the pesticide of interest (parent and/or metabolites) is determined in a limited set of plant and/or animal products.

Residue analytical methods for post-registration purposes (enforcement)
These methods are meant for enforcement purposes. These methods are meant to be used by more laboratories and ideally large numbers of pesticides should be determined simultaneously in a wide range of plant and/or animal products.

In appendix I an overview is given for validation requirements for residue analytical methods for pre-registration purposes (e.g. residue field trials) and in appendix II for enforcement purposes (post-registration) according to the EU-guidelines [1, 2] and the FAO/Codex-guidelines [3, 4]. The Codex requirements are focussed on post-registration methods. The CCPR (Codex Committee on Pesticide Residues) is preparing an update of the requirements for post-registration purposes for Codex [5] based on a joint expert consultation of AOAC/FAO/IAEA/IUPAC (Association of Official Analytical Chemists/Food and Agricultural Organisation of the United Nations/International Atomic Energy Agency/International Union of Pure and Applied Chemistry, [6]). The updated draft Codex guidance document [5] defines major changes in the assessment of analytical methods and an overview of the requirements is included in appendix II. An overview of terms and abbreviations used, is given in appendix III and IV.
Note: When the EU-guidelines and FAO/Codex guidelines are updated, the present factsheet has to be updated as well.

3.2.2 Assessment of pre- and post-registration analytical methods

Assessment of requirements for pre- and post-registration analytical methods is described below in the order in which the requirements appear in appendix I and II. In the evaluation of a residue analytical method, the EU-criteria [1, 2] are used for national and EU assessments and the FAO/Codex-criteria [3-6] are used for JMPR assessments. When a method is meant both for pre-registration and for post-registration, the method is evaluated twice: against both types of criteria. Requirements that are not very clear or that are amenable to more than one interpretation are explained (RIVM interpretation). For requirements for which no criteria are given, RIVM criteria are deduced. The final procedure per requirement for national and EU assessments is summarised as "RIVM procedure for EU", the final procedure per requirement for JMPR assessments is summarised as "RIVM procedure for FAO".

Requirement 1: Quality Assurance (QA) for method validation

Pre-registration:
EU [1]: not required.
FAO [3]: not addressed.
RIVM procedure for EU and FAO: the presence or absence of GLP or national QA is taken up in the summary, but is not taken into account for assessment of method validity. The EU-guidance document distinguishes between method validation (GLP not required) and the moment that the method is used to generate data (GLP required). Because method validation is assessed separately from data generation (e.g. residue trials) QA is not required. RIVM uses the EU interpretation also for FAO assessments.

Post-registration:
EU [2]: not required.
FAO [4]: not addressed.
Updated draft Codex [5]: not addressed.

RIVM procedure for EU and FAO: the presence or absence of GLP or national QA is taken up in the summary, but is not taken into account for assessment of method validity. The EU-guidance document distinguishes between method validation (GLP not required) and the moment that the method is used to generate data (GLP required). Because method validation is assessed separately from data generation (e.g. residue monitoring) QA is not required. RIVM uses the EU interpretation also for FAO assessments.

Requirement 2: Full description of the method
The meaning of a "full description of the method" is explained in the EU-guidance documents for pre-registration and post-registration [1, 2].

Pre-registration:
EU [1]: required.
FAO [3]: required.
RIVM procedure for EU: according to EU.
RIVM procedure for FAO: according to FAO.

Post-registration:
EU [2]: required.
FAO [3, 4]: published method required.
Updated draft Codex [5]: published method required.
RIVM procedure for EU: according to EU.
RIVM procedure for FAO: according to updated draft Codex.

Requirement 3: Commonly available analytical techniques
The meaning of "commonly available analytical techniques" (up to 31st December 2003) is explained in the EU-guidance document for post-registration [2].

Pre-registration:
EU [1]: not required.
FAO [3]: not required.
RIVM procedure for EU and FAO: not evaluated.

Post-registration:
EU [1]: required.
FAO [3, 4]: required.
Updated draft Codex [5]: required.
RIVM procedure for EU: according to EU.
RIVM procedure for FAO: according to updated draft Codex.

Requirement 4: Multi residue method (MRM)

Pre-registration:
EU [1]: not required.
FAO [3]: not required.
RIVM procedure for EU and FAO: not evaluated.

Post-registration:
EU [2]: MRM required, unless the applicant can show that the analyte cannot be implemented in an existing multi residue method [10-20].
FAO [3, 4]: preferred, but not required.

Updated draft Codex [5]: preferred, but not required.
RIVM procedure for EU: according to EU.
RIVM procedure for FAO: the presence or absence of an MRM is taken up in the summary, but is not taken into account for assessment of method validity. Because a multi residue method is preferred, but not required by the FAO, an analytical method cannot be rejected when a single or special residue method (SRM) is proposed.

Requirement 5: Validation report for all sample matrices
Pre-registration:
EU [1]: required. In the pre-registration context "all sample matrices" is interpreted by the EU as all matrices that are analysed (i.e. for which supervised residue trials, processing studies, storage stability studies or feeding studies are submitted). Reduced validation data for sample matrices within the same crop group (as defined for post-registration, table 3.1) are acceptable. In these cases data must include recovery at 2 fortification levels and associated precision (minimum 3 replicate samples instead of 5) together with an assessment of matrix interference (requirements 10, 11, 15, 18, 19/20). Processed crops are not addressed by the EU. For animal products a sample matrix is interpreted by the EU as: a. milk; b. liver, kidney and muscle (meat); c. fat; d. eggs. The animal species is not addressed.
FAO [3]: not addressed.
RIVM procedure for EU: according to EU. For raw agricultural commodities the EU-guidelines are used.
For processed crops, the RIVM considers the processing form as qualifying. Example: when residue data are generated for cooked crop A, than reduced validation data for cooked crop A are acceptable if crop A belongs to the same crop group as crop B and the analytical method for cooked crop B is considered valid.
For animal products the four EU groups are discriminated. Because in the updated draft Codex document (see post-registration) the animal species is considered not relevant, the animal species is considered not relevant for EU assessment either. Example: when residue data are generated for bovine muscle, for which no validation report is available, than residue data for bovine muscle are considered valid if the analytical method for chicken muscle is considered valid.
RIVM procedure for FAO: For raw agricultural commodities the analytical method is considered valid if the analytical method for at least one crop within a crop group is considered valid. In this case RIVM uses the crop group definitions from the Codex classification of foods and animal feeds [7]. Example: When residue data are generated for crop Y, for which no validation report is available, than residue data for crop Y are considered valid, if crop Y falls within the same crop group as crop X and the analytical method for crop X is considered valid.
For processed crops the processing form is qualifying. Example: When residue data are generated for cooked crop A, for which no validation report is available, than residue data for cooked crop A are considered valid, if crop A falls within the same crop group as crop B and the analytical method for cooked crop B is considered valid.
For animal products the following 5 groups are discriminated: eggs, milk, muscle (meat), fat and edible offal (kidney, liver). The animal species is considered not relevant (see post-registration). Example: when residue data are generated for bovine muscle, for which no validation report is available, than residue data for bovine muscle are considered valid if the analytical method for chicken muscle is considered valid.

Table 3.1 Crop groups for post-registration purposes (EU and FAO)

EU crop groups	EU representative sample matrices	FAO crop groups	FAO representative sample matrices
cereals and other dry crops	barley, wheat, rye, oats	dry materials	cereals, cereal products
fruits with high acid content	lemons, oranges, grapefruits, apples	high acid content	citrus fruits
commodities with high fat content	rape seed; linseed; olives, nuts, avocado	high oil or fat	oil seeds, nuts, avocado
commodities with high water content	lettuce, tomatoes, cherries, strawberries	high water and low or no chlorophyll content	pome fruits, stone fruits, berries, small fruits, fruiting vegetables, root vegetables
		high water and high chlorophyll content	leafy vegetables, brassica, legume vegetables
		high sugar content	raisins, dates

Post-registration:
EU [2]: Validation data are required for 4 crop groups and 6 animal product groups (table 3.1). Specific validation data are required for commodities which are difficult to analyse (hops, brassicas, bulb vegetables, herbs, tea). Validation is only required for crop groups for which an admission is asked for. Validation for animal products (muscle, eggs, milk, fat, liver, kidney) is only required if also a residue definition for animal products is proposed. Validation for liver and kidney is only required when an MRL is proposed for these tissues. Validation for fat is only required when
log $K_{o/w}$>3. The animal species is not addressed.
FAO [3, 4]: not addressed.
Updated draft Codex [5]: For multi residue methods, validation data for 6 crop groups (table 3.1) and 5 animal products are required. Specific validation data are required for commodities which are difficult to analyse (garlic, hops, tea, spices, cranberry). Validation is required for all crop groups, because enforcement methods are also used to detect pesticides in crops for which no admission is asked for. Validation for animal products (muscle, eggs, milk, edible offal (liver or kidney), fat) is only required if also a residue definition for animal products is proposed. The animal species is considered not relevant by FAO ("validation for chicken fat may apply to cattle fat"). For single or special residue methods (SRM), validation data for all matrices specified in the method are required.
RIVM procedure for EU: according to EU. Because in the updated draft Codex document the animal species is considered not relevant, the animal species is considered not relevant for EU assessment either.
RIVM procedure for FAO: according to the updated draft Codex.

Requirement 6: Validation report for all components of the residue definition
The residue analytical methods are evaluated separately for each component of the residue definition.
Explanatory EU note [1]: The term "separately for each component" implies that non-specific and common moiety methods are considered as "not valid". Non-specific methods (e.g. immuno assays, bioassays, colorimetric methods) and common moiety methods are only acceptable in exceptional circumstances where there is no other practical means of determining the target analyte, and in these cases, full justification by the applicant is required.

Explanatory RIVM note: A case where a common moiety method is acceptable (with full justification of the applicant) is, when the residue definition for compliance with MRL and the residue definition for risk assessment are identical and the residue definition comprises compounds which can easily be converted into one single compound. In this case the overall performance is improved.

Pre-registration:
Residue definition in this context means both the residue definition for compliance with MRL and the residue definition for risk assessment.
EU [1]: required.
FAO [3]: required.
RIVM procedure for EU: according to EU.
RIVM procedure for FAO: according to FAO.

Post-registration:
The residue definition in this context is the residue definition for compliance with MRL.
EU [2]: required.
FAO [3, 4]: required.
Updated draft Codex [5]: required.
RIVM procedure for EU: according to EU.
RIVM procedure for FAO: according to updated draft Codex.

Requirement 7: Confirmatory techniques
The meaning of "confirmatory techniques" is explained in the EU-guidance document for post-registration [2] and the updated draft Codex guidance document [5].

Pre-registration:
EU [1]: only required when a common moiety or non-specific method is used; no criteria given.
FAO [3]: not addressed.
RIVM procedure for EU: When a common moiety or non-specific method is used (see also requirement 6) the post-registration criterion is used: the mean $result_{confirmation\ technique}$/mean $result_{original\ technique}$ should be between 0.8 and 1.2 ($n \geq 1$).
RIVM procedure for FAO: not evaluated.

Post-registration:
EU [2]: required, no criteria given.
FAO [4]: required, criterion: when MRL is exceeded: within 20% of primary result. *Updated draft Codex [5]*: not addressed.
RIVM procedure for EU and FAO: ratio between mean results of the techniques used 0.8-1.2 ($n \geq 1$). The RIVM interpretation of the FAO criterion is that the same sample is analysed with two techniques. As no criteria are given, the RIVM assumes that each sample is analysed once with each technique. When samples are analysed in duplicate or triplicate, the mean result is taken. The mean $result_{confirmation\ technique}$/mean $result_{original\ technique}$ should be between 0.8 and 1.2.

Requirement 8: Independent laboratory validation
Pre-registration:
EU [1]: not required.
FAO [3]: not addressed.

RIVM procedure for EU and FAO: not evaluated.

Post-registration:
EU [3]: required for 2 plant groups (one of them with a high water content) and 2 animal products. The validation has to be conducted by two independent laboratories: the second laboratory has to verify that the measurement range specified by the first laboratory is valid by measuring 5 spiked samples at LOQ, 5 spiked samples at MRL or 10LOQ (whichever is higher) and 2 untreated control samples. For criteria see requirement 11, 15, 18, 19/20.
FAO [3]: preferred, but not required. Criteria: CV_R<77% (0.01 mg/kg), <45% (0.1 mg/kg), or <22% (1 mg/kg).
Updated draft Codex [5]: not addressed.
RIVM procedure for EU: according to EU.
RIVM procedure for FAO: the presence or absence of an independent laboratory validation is taken up in the summary, but is not taken into account for assessment of method validity. The FAO has set criteria for the between-laboratory reproducibility when the same sample is analysed by different laboratories. In practice independent laboratory validations are often not conducted on the same sample(s). In stead the method developed and validated in the first laboratory, is validated again in the second laboratory on spiked control samples from a different source. In this case the FAO criteria cannot be applied and therefore the EU procedure will be followed.

Requirement 9: Limit of detection
The most practicable definition for the limit of detection is: LD = 3s under reproducibility conditions [8, 9], where s is the standard deviation. The standard deviation is determined from mg/kg results (i.e. determination against the calibration model) for untreated control samples (when peaks appear in the chromatogram) or for untreated control samples spiked at the expected LD level (when no peaks appear in the chromatogram) under reproducibility conditions. Techniques which give no value or a zero value when no peaks appear in the chromatogram are GC-ECD, GC-NPD and HPLC-UV.

Pre-registration:
EU [1]: not addressed.
FAO [3]: not addressed.
RIVM procedure for EU and FAO: not evaluated.

Post-registration:
EU [2]: not addressed.
FAO [4]: required, but no criteria given.
Updated draft Codex [5]: required, but no criteria given.
RIVM procedure for EU: not evaluated.
RIVM procedure for FAO: the proposed limit of detection is taken up in the summary, but is not taken into account for assessment of method validity. Because criteria for the limit of quantification are set and the limit of quantification is the only relevant lower limit for quantitative methods, the limit of detection is not evaluated.

Requirement 10: Limit of quantification
The terms limit of detection (LD), limit of determination (LOD), limit of quantification and limit of quantitation (LOQ) are often used as synonyms [8, 9]. Because of the confusion between LOD being limit of detection or limit of determination, the abbreviation LD is used

for limit of detection and the abbreviation LOQ is used for limit of quantitation, limit of quantification or limit of determination.
RIVM interpretation: In study reports very often a LOQ is stated by the study author ($LOQ_{reported}$) without data to support this LOQ. The LOQ that is considered valid (LOQ_{valid}) by the RIVM is the lowest concentration level with acceptable matrix interferences, acceptable recovery, and acceptable precision (requirements 15, 18, 19/20).

Pre-registration:
EU [1]: required, but no criterion given.
FAO [3]: not addressed.
RIVM procedure for EU and FAO: <0.5x minimal expected residue concentration or <0.01 mg/kg, whichever is higher. The LOQ_{valid} should be low enough to be able to quantify the residues expected from metabolism studies conducted at GAP (good agricultural practice) conditions. In cases where residues are expected (from metabolism studies), the LOQ_{valid} should not exceed 0.5x the minimal expected residue concentration when the pesticide is used according to GAP. In cases where no or low residues are expected (from metabolism studies), the LOQ_{valid} should be as low as possible and should not exceed 0.01 mg/kg.
Special attention should be given to residue data that are generated by methods for which the LOQ_{valid} is considered higher than the $LOQ_{reported}$. When values reported in the range $<LOQ_{reported}$ to $<LOQ_{valid}$ have to be used in the calculation of an MRL, these values should be modified into $<LOQ_{valid}$ before MRL calculation.

Post-registration:
EU [2]: required, criterion: the LOQ_{valid} should be ≤0.1 mg/kg if MRL > 0.1 mg/kg, or ≤0.5MRL if MRL ≤0.1 mg/kg, except when the MRL is set at LOQ_{valid}.
RIVM explanatory note: The MRL is set at LOQ_{valid} for crops without registered uses for the compound in question or for crops where no residues are expected (confirmed with metabolism studies and supervised field trials).
FAO [4]: required, but no criterion given.
Updated draft Codex [5]: the LCL (lowest calibrated level) should be 0.5 mg/kg if MRL ≥ 5 mg/kg, or 0.1-0.5 mg/kg if MRL = 0.5-5 mg/kg, or 0.02-0.1 mg/kg if MRL = 0.05-0.5 mg/kg or 0.5MRL if MRL <0.05 mg/kg, except when the MRL is set at LOQ_{valid}.
RIVM procedure for EU: according to EU.
RIVM procedure for FAO: according to updated draft Codex, except that LCL = LOQ_{valid}.

Requirement 11:. Measurement range
The terms measurement range, working range, calibration range, analytical range and concentration range are synonyms.
Explanatory FAO note: The measurement range is defined as the range that is valid, i.e. the lowest and highest valid concentration level.
RIVM interpretation: The lowest valid concentration level (or LOQ_{valid}) is defined as the lowest concentration level with acceptable matrix interferences, acceptable recovery and acceptable precision (requirements 10, 15, 18, 19/20). The highest valid concentration level is defined as the highest concentration level with acceptable recovery and acceptable precision (requirements 18, 19/20).

Pre-registration:
EU [1]: required, criterion: LOQ-10LOQ or LOQ-likely residue levels (whichever is broader);
FAO [3]: not addressed.

RIVM procedure for EU: according to EU.
RIVM procedure for FAO: the measurement range should at least cover the concentration range that is relevant for the residue study in question.
Four cases can be discriminated:

1. When from metabolism studies no residue levels are expected in plants and/or in livestock, only the validation at the lowest quantifiable level or LOQ is relevant.
2. When from metabolism studies residue levels are expected in crops, only the validation at the concentration levels that are selected for setting the MRL, is relevant. The validated measurement range should at least cover the selected concentration levels (residues at the critical GAP).
3. When from metabolism studies residue levels are expected in livestock, the validated measurement range should cover the whole concentration range found in feeding studies.
4. All concentrations found in processing, stability or rotational crop studies are used for setting processing factors, stability periods and waiting periods, respectively. In this case the validated measurement range should cover the whole concentration range.

Post-registration:
EU [2]: required, criterion: LOQ-MRL or LOQ-10LOQ (whichever is broader).
FAO [4]: required, but no criterion given.
Updated draft Codex [5]: LCL (lowest calibrated level) to 2-3x AL (accepted level).
RIVM procedure for EU: according to EU.
RIVM procedure for FAO: according to updated draft Codex, except that LCL = LOQ_{valid} and AL = MRL. Concentrations above the MRL are expected from pesticide abuse (too short PHI intervals or too high pesticide dosages). Therefore a post-registration method has to be able to measure concentrations above MRL.
RIVM remark: Coupling validation to MRL levels has the practical problem that MRLs can be modified when new residue data become available. This implies that a previously accepted post-registration method or a new post-registration method valid for existing MRLs can be considered not-valid in a new evaluation round. Validation of a post-registration method for the broadest possible measurement range can circumvent these problems. All validated levels are therefore summarised.

Requirement 12: Calibration model
The calibration model is either a linear, a weighted linear or a polynomial function.

Pre-registration:
EU [1]: the calibration model should be deduced from at least 5 single concentration levels or from 3 duplicate concentration levels.
FAO [3]: not addressed.
RIVM procedure for EU: according to EU.
RIVM procedure for FAO: not evaluated.

Post-registration:
EU [2]: the calibration model should be deduced from at least 5 single concentration levels or 3 duplicate concentration levels.
FAO [4]: not addressed.
Updated draft Codex [5]: a linear calibration model should be deduced from at least 4 duplicate concentration levels (including blank) and a non-linear calibration model should be deduced from at least 7 triplicate concentration levels.
RIVM procedure for EU: according to EU.

RIVM procedure for FAO: not evaluated. Criteria for the updated draft Codex are not implemented by the RIVM, because lack of fit of the calibration model (requirement 13) is not evaluated.

Requirement 13: Lack of fit of the calibration model

Pre-registration:
EU [1]: correlation coefficient (r) must be reported and a calibration plot must be submitted; but criteria are not defined.
FAO [3]: not addressed.
RIVM procedure for EU: Linear and weighted linear calibration model: correlation coefficient (r) ≥0.99; polynomial calibration model: correlation coefficient (r) ≥0.98.
RIVM procedure for FAO: not evaluated. Lack of fit of the calibration model is not a very useful performance characteristic in itself, as it is only applicable to standard solutions and it does not give an indication of the trueness of concentration levels in samples. Lack of fit of the calibration model can give supporting evidence for the validity of the measurement range (because systematic errors at several points in the measurement range are available or can be calculated), but it cannot replace the validation of the measurement range. Because criteria for the measurement range are defined (requirement 11), lack of fit of the calibration model is not evaluated.

Post-registration:
EU [2]: the correlation coefficient (r) must be reported and the calibration plot must be submitted; but criteria are not defined.
FAO [4]: not addressed.
Updated draft Codex [5]: linear calibration model: correlation coefficient (r) ≥0.99 and standard deviation of the residuals ($S_{y/x}$) ≤0.1; the weighted linear model is not addressed; polynomial calibration model: correlation coefficient (r) ≥0.98.
RIVM procedure for EU: The criteria are deduced from the updated draft Codex document for post-registration.
Linear and weighted linear calibration model: correlation coefficient (r) ≥0.99; polynomial calibration model: correlation coefficient (r) ≥0.98.
RIVM procedure for FAO: not evaluated (see pre-registration).
The criteria for the updated draft Codex are not implemented by the RIVM, because lack of fit of the calibration model is not seen as an important performance characteristic (see pre-registration).

Requirement 14: Matrix effects

Explanatory RIVM note: Matrix effects are introduced by physical differences (e.g. polarity, volatility) between sample and calibration standards, thereby influencing chromatographic transmission or detection response. Matrix effects influence the slope of the calibration model (both positive and negative) and thereby introduce systematic errors in analytical determinations. Matrix effects in samples can be expected if the calibration model is deduced from standards made in pure extraction solvents. Matrix effects in samples are cancelled out when the calibration model is deduced from matrix matched standards or when proper internal standards are used. Matrix effects can be visualised when two calibration models are compared: one deduced from standards in pure solution and another one deduced from standards in extracts from (untreated) samples.

Pre-registration:
EU [1]: required, but no criterion given; matrix matched standards are preferred;

FAO [3]: not addressed.
RIVM procedure for EU: A matrix effect is confirmed if the difference between slopes (for standards in pure solution and standards in matrix) is significant at P=0.05. In order to do a significance test, slope and slope standard deviations have to be reported for both calibration models. When the matrix effect is significant, matrix matched standards have to be used.
RIVM procedure for FAO: not evaluated. Matrix effects are evident when too low or too high recoveries are found. Because criteria for recoveries are defined (requirement 18), matrix effects are evaluated implicitly in recovery.

Post-registration:
EU [2]: required, but no criterion given; matrix matched standards are preferred.
FAO [4]: not addressed.
Updated draft Codex [5]: matrix effects should be verified for representative analytes and representative matrices. A matrix effect is confirmed if the difference between slopes (for standards in pure solution and standards in matrix) is significant at P=0.05. When the matrix effect is significant, matrix matched standards have to be used.
RIVM procedure for EU: A matrix effect is confirmed if the difference between slopes (for standards in pure solution and standards in matrix) is significant at P=0.05. In order to do a significance test, slope and slope standard deviations have to be reported for both calibration models. When the matrix effect is significant, matrix matched standards have to be used.
RIVM procedure for FAO: not evaluated (see pre-registration). The updated draft Codex criteria are not implemented by the RIVM, because matrix effects are implicitly shown by recovery results.

Requirement 15:. Matrix interferences
Explanatory RIVM note: Matrix interferences are caused by (unknown or unexpected) compounds in the sample either through introduction of an interfering peak or through a raise in background signal. Matrix interferences can influence the intercept of the calibration model (positive effect) and thereby introduce systematic errors in analytical determinations. They may also affect the precision due to poor peak recognition of the chromatography data system or saturation of the detector (included in requirement 19/20).

Pre-registration:
EU [1]: at least two (untreated) control samples should be measured and the concentration of both samples should be ≤0.3*LOQ.
FAO [3]: not addressed.
RIVM procedure for EU and FAO: according to EU. When this criterion is not fulfilled, than the reported LOQ for the method is not realistic. More realistic is LOQ = $(1/0.3)*C_{\text{untreated control sample}}$, where C is the highest reported concentration in the untreated control samples. The method can be considered valid at this higher LOQ if requirement 10, 18, and 19/20 are fulfilled.

Post-registration:
EU [2]: at least two (untreated) control samples should be measured and the concentration of this sample should be ≤0.3*LOQ.
FAO [4]: not addressed.
Updated draft Codex [5]: at least 5 untreated control samples from different sources should be measured; a criterion is not given.
RIVM procedure for EU: according to EU.

RIVM procedure for FAO: ≤0.3*LOQ (n≥5). The number of samples is according to the draft Codex guidance document and the criterion is according to the EU. For further remarks see pre-registration.

Requirement 16: Pesticide interferences

Pre-registration:
EU [1]: not addressed.
FAO [3]: not addressed.
RIVM procedure for EU and FAO: not evaluated. Pesticide interference is not very likely as only the pesticide of interest is used in supervised residue trials, processing studies, storage stability studies and feeding studies. In cases where other pesticides are used during the field trial and the possibly existing pesticide interferences are not verified, a worst case situation arises, because the apparent residue level will be raised. The calculated MRL will be overestimated in this case.

Post-registration:
EU [2]: not addressed.
FAO [4]: not addressed.
Updated draft Codex [5]: the separation for all of the analytes in a multi residue method should be verified at AL (=MRL) level. The relative retention times, resolution and tailing factors of critical peaks should be reported. Peak maxima should be separated by at least one full peak width at 10% of the peak height.
RIVM procedure for EU and FAO: not evaluated. Pesticide interference evaluation is not seen as a very important performance characteristic as a confirmatory analysis is required when the MRL is exceeded. In practice the separation of 300 pesticides for a multi residue method is not perfect, but this is not seen as a practical problem as in general these pesticides will not be present at the same time at the same crop.

Requirement 17: Metabolite or isomer or analogue interferences

Pre-registration:
EU [1]: required, where this is necessary for carrying out risk assessment, but no criterion is given.
FAO [3]: not addressed.
RIVM procedure for EU: only evaluated when necessary for carrying out risk assessment. Criterion: The concentration of a solution (n≥1) without analyte but with normally expected concentrations of metabolites and/or isomers and/or analogues should be ≤0.3*LOQ_{valid} (see requirement 15). As an alternative, the selectivity of the extraction and/or the chromatographic separation may be shown by chromatographic spectra in a solution (n≥1) with mixed standards for parent compound, metabolites and/or isomers and/or analogues. In this case, peak maxima should be separated by at least one full peak width at 10% of the peak height.
RIVM procedure for FAO: not evaluated.

Post-registration:
EU [2]: not addressed.
FAO [4]: not addressed.
Updated draft Codex [5]: not addressed.
RIVM procedure for EU and FAO: not evaluated.
RIVM explanatory note. Metabolite, isomer or analogue interferences will increase the total concentration found, but for enforcement the most practical residue definition is chosen.

When parent and/or metabolites and/or isomers and/or analogues are difficult to separate, they are included in the residue definition.

Requirement 18: Mean recovery for analysis excluding sample processing
Explanatory RIVM note: Mean recovery is mean recovery per concentration level and per matrix. The spikes are added to samples that are already processed, but still need extraction, clean-up, derivatisation (if applicable) and measurement. The calibration is performed as described in the method (i.e. either standards in pure solution or in matrix).

Pre-registration:
EU [1]: required, recoveries should be determined at at least two concentration levels for at least 5 samples per level and per matrix and the mean recovery per concentration level and per matrix should be 70-110%.
FAO [3]: not addressed.
RIVM procedure for EU and FAO: according to EU. The selected concentration levels should meet the criteria for requirement 11 (measurement range).

Post-registration:
EU [2]: required, recoveries should be verified for at least two concentration levels (LOQ and MRL or 10LOQ) for at least 5 samples per level and per matrix and the mean recovery per concentration level and per matrix should be 70-110%.
FAO [4]: required, recovery ≥80%.
Updated draft Codex [5]: required, recoveries should be verified for at least three concentration levels for at least 5 samples per matrix at LCL (lowest calibrated level) and AL (accepted limit) and at least 3 samples per matrix at 2-3 AL for representative analyte matrix combinations. Criteria for the mean recovery are presented in table 3.2.
RIVM procedure for EU: according to EU.
RIVM procedure for FAO: according to updated draft Codex, except that LCL=LOQ_{valid} and AL = MRL. The selected concentration levels should meet the criteria for requirement 11 (measurement range).

Table 3.2 Validation criteria according to the draft Codex guidance document for repeatability, within-laboratory reproducibility and recovery of spikes added to samples after sample processing.

Concentration level	Repeatability ($CV_{A,r}$)	Within-laboratory reproducibility ($CV_{A,R}$)	Mean recovery
≤ 0.001 mg/kg	≤35%	≤53%	50%-120%
> 0.001 mg/kg to ≤0.01 mg/kg	≤30%	≤45%	60%-120%
> 0.01 mg/kg to ≤0.1 mg/kg	≤20%	≤32%	70%-120%
> 0.1 mg/kg to ≤1 mg/kg	≤15%	≤23%	70%-110%
> 1 mg/kg	≤10%	≤16%	70%-110%

RIVM remark: In some study reports the overall recovery per concentration level (for different matrices) or the overall recovery per matrix (for different concentration levels) is given. These performance characteristics are not very useful and are therefore not mentioned as a requirement. When the recovery per concentration level and per matrix lies within the criteria, than the overall recovery lies also within the criteria. But the opposite is not always the case. Overall recoveries can be used to juggle with data: a bad recovery at low concentrations is averaged out if recoveries at higher concentrations are good.

Requirement 19: Precision: repeatability for analysis excluding sample processing

Explanatory RIVM note: Repeatability is expressed as relative standard deviation (RSD_r or $CV_{A,r}$) per concentration level and per matrix. Repeatability conditions are assumed when the data are generated within the same day or run (whichever is longer) or when no measurement conditions are stated. The calibration is performed as described in the method (i.e. either standards in pure solution or in matrix). The spikes are added to samples that are already processed, but still need extraction, clean-up, derivatisation (if applicable) and measurement. When no repeatability data are given, the repeatability (as relative standard deviation) can be calculated from the reported recoveries per concentration level and per matrix.

Pre-registration:
EU [1]: ≤20% (n≥5), repeatability should be verified at at least two concentration levels for at least 5 samples per level and per matrix.
FAO [3]: ≤10%.
RIVM procedure for EU: according to EU. When repeatability data are not available but within-laboratory reproducibility data are, the within-laboratory reproducibility is evaluated against the repeatability criteria. When within-laboratory reproducibility fulfils the criteria for repeatability, the precision is considered valid.
RIVM procedure for FAO: ≤10% (n≥5), repeatability should be verified at at least two concentration levels for at least 5 samples per level and per matrix. When repeatability data are not available but within-laboratory reproducibility data are, the within-laboratory reproducibility is evaluated according to the repeatability criteria. When within-laboratory reproducibility fulfils the criteria for repeatability, the precision is considered valid.

Post-registration:
EU [2]: ≤20% (n≥5), repeatability should be verified at at least two concentration levels (LOQ and MRL or 10LOQ) for at least 5 samples per level and per matrix.
FAO [3]: ≤10%.
Updated draft Codex: the repeatability should be verified for at least three concentration levels for at least 5 samples per matrix at LCL and AL and at least 3 samples per matrix at 2-3 AL for representative analyte matrix combinations. Criteria for the repeatability are presented in table 3.2.
RIVM procedure for EU: according to EU. When repeatability data are not available but within-laboratory reproducibility data are, the within-laboratory reproducibility is evaluated against the repeatability criteria. When within-laboratory reproducibility fulfils the criteria for repeatability, the precision is considered valid.
RIVM procedure for FAO: according to the updated draft Codex.

Requirement 20: Precision: Within-laboratory reproducibility for analysis excluding sample processing

Explanatory RIVM note: Within-laboratory reproducibility is expressed as relative standard deviation (RSD_R or $CV_{A,R}$) per concentration level and per matrix. Within-laboratory reproducibility conditions are assumed when the data are generated on different days or runs (whichever is longer). The calibration is performed as described in the method (i.e. either standards in pure solution or in matrix). The spikes are added to samples that are already processed, but still need extraction, clean-up, derivatisation (if applicable) and measurement. When no reproducibility data are given, the reproducibility (as relative standard deviation) can be calculated from the reported recoveries per concentration level and per matrix.

Pre-registration:
EU [1]: not addressed.
FAO [3]: not addressed.
RIVM procedure for EU: not evaluated.
RIVM procedure for FAO: not evaluated.

Post-registration:
EU [2]: not addressed.
FAO [3, 4]: not addressed.
Updated draft Codex [5]: at least 5 samples per matrix at LCL and AL and at least 3 samples per matrix at 2-3 AL for representative analyte matrix combinations. Criteria for within-laboratory reproducibility are presented in 3. 2.
RIVM procedure for EU: not evaluated.
RIVM procedure for FAO: according to the updated draft Codex, except that LCL=LOQ_{valid} and AL=MRL.

Requirement 21: Individual recoveries
Pre-registration:
EU [1]: not addressed.
FAO [3]: not addressed.
RIVM procedure for EU and FAO: not evaluated.

Post-registration:
EU [2]: not addressed.
FAO [4]: 70-110%.
Updated draft Codex [5]: not addressed.
RIVM procedure for EU: not evaluated.
RIVM procedure for FAO: not evaluated. Reporting individual recoveries is another way of expressing the variation of recoveries. As variation of recoveries is already defined by within-laboratory reproducibility (requirement 20), individual recoveries are not evaluated.

Requirement 22/23: Mean recovery and precision for derivatisation
EU remark [1, 2]: The method is considered to remain specific to the analyte of interest if the derivatised species is specific to that analyte. However, where the derivative formed is common to two or more active substances or impurities or is classed as a different active substance, the method should be considered non-specific (see requirement 6).

Pre-registration:
EU [1]: required, no criteria given.
FAO [3]: not addressed.
RIVM procedure for EU and FAO: if applicable (see post-registration).

Post-registration:
EU [2]: required, mean recovery ≥70%.
FAO [3, 4]: not addressed.
Updated draft Codex [5]: not addressed.
RIVM procedure for EU and FAO: if applicable, mean recovery 70%-110% (n≥3) and RSD_r≤10% (n≥3; repeatability conditions).
Although it is possible to get a good method recovery if derivatisation efficiency is low (e.g. 10%), the method is probably not very robust and can be influenced by small changes in

conditions. Therefore also recovery and precision for derivatisation have to be verified. Precision data for repeatability conditions are sufficient.

Requirement 24/25: Mean recovery and precision for internal standards
Pre-registration:
EU [1]: required, no criteria given.
FAO [3]: not addressed.
RIVM procedure for EU and FAO: if applicable, mean recovery 70%-110% ($n \geq 3$) and $RSD_r \leq 10\%$ ($n \geq 3$; repeatability conditions). For isotopic internal standards (e.g. ^{13}C labelled analyte) the same criteria are used.
Although it is possible to get a good method recovery if the recovery for internal standards is low (e.g. 10%), the method is probably not very robust and can be influenced by small changes in conditions. Therefore also recovery and precision for internal standards has to be verified. Precision data for repeatability conditions are sufficient.

Post-registration:
EU [2]: required, criterion: similar as analyte.
FAO [3, 4]: not addressed.
Updated draft Codex [5]: not addressed.
RIVM procedure for EU and FAO: see pre-registration.

Requirement 26: Trueness
Pre-registration:
EU [1]: not addressed.
FAO [3]: not addressed.
RIVM procedure for EU: not evaluated.
RIVM procedure for FAO: the presence or absence of a trueness experiment is taken up in the summary, but is not taken into account for assessment of method validity.

Post-registration:
EU [2]: not addressed.
FAO [4]: preferred but not required: extraction of radiolabelled compounds or a comparative extraction with a method of known efficiency; criteria not given.
Updated draft Codex [5]: at least 5 replicate analytical portions of samples with incurred residues (at AL) or reference material with incurred residues (at AL) should be measured. The consensus value of reference material and the mean residue should not differ significantly at P=0.05, applying CV_A (table 3.2) in the calculation, or for samples with incurred residues, the mean results obtained with the reference procedure and the tested procedure should not differ significantly at P=0.05, applying CV_L (table 3.3) in the calculation.
RIVM procedure for EU: not evaluated.
RIVM procedure for FAO: the results of a trueness experiment are taken up in the summary, but are not taken into account for assessment of method validity. Trueness is an important test in method validation, because it is the only test where systematic errors can be detected. Although spike recovery also provides an estimate of systematic errors, not all systematic errors are covered [8]. For a trueness test, a homogeneous and stable sample with known amounts of *incurred* residues has to be available (certified reference material). These samples are not available for pesticides.
There are two possible alternatives:

1. When a reference method is available, systematic errors can be deduced from results of samples with incurred residues analysed by both methods (test method and reference method) and using either linear regression (at least 5 single concentration levels, expected slope=1; expected intercept = 0; P=0.05) or a paired t-test (at least 5 single concentration levels; P=0.05) or a t-test (at least one concentration level (at MRL) measured with 5 replicates by each method; P=0.05) to check for differences. When a significant difference is established, it is assumed that the test method contains systematic errors and the reference method does not contain systematic errors.
2. The extraction efficiency for the analytical method can be deduced from radiolabelled incurred residues in combination with radiometric separation methods (e.g. HPLC/LSC; TLC/autoradiography). The same criterion is used as for recovery (requirement 18).

But as a reference method is in most cases not available and radiolabelled methods are not available to routine laboratories, trueness can only be evaluated where applicable (when a study report on trueness is available). Because trueness experiments are in most cases not available, it would be unfair to reject a study with a bad trueness experiment, whereas a study without trueness experiment would be accepted. Therefore, a trueness experiment is taken up in the summary, but it cannot be a reason to reject a study.

Requirement 27: Homogeneity

Pre-registration:
EU [1]: not addressed.
FAO [3]: not addressed.
RIVM procedure for EU and FAO: not evaluated.

Post-registration:
EU [2]: not addressed.
FAO [3, 4]: not addressed.
Updated draft Codex [5]: the repeatability or within-laboratory reproducibility for at least 5 replicate analytical portions of one representative commodity per group (3. 1) should be determined. Use preferably commodities with incurred stable residues or treat the surface of a small portion of the natural unit (<20%) of a laboratory sample before cutting or chopping with a stable analyte. The criterion for relative standard deviation of sample processing is: $CV_{Sp} \leq 10\%$. As CV_{Sp} cannot be measured on its own, criteria for repeatability including sample processing ($CV_{L,r}$) or within-laboratory reproducibility including sample processing ($CV_{L,R}$) are calculated from: CV_L=square root ($CV_{Sp}^2 + CV_A^2$). The results are presented in table 3.3.
RIVM procedure for EU: not evaluated.
RIVM procedure for FAO: according to updated draft Codex. Either the repeatability or the within-laboratory reproducibility for spikes added to samples before sample processing, should be below the limits stated in table 3.3.

Table 3.3. Within-laboratory validation criteria according to the draft Codex guidance document for repeatability and reproducibility of spikes added to samples before sample processing, allowing up to 10% variability of sample processing.

Concentration level	Repeatability ($CV_{L,r}$)	Within-laboratory reproducibility ($CV_{L,R}$)
≤ 0.001 mg/kg	≤ 36%	≤ 54%
> 0.001 mg/kg ≤0.01 mg/kg	≤ 32%	≤ 46%
> 0.01 mg/kg ≤0.1 mg/kg	≤ 22%	≤ 34%
> 0.1 mg/kg ≤1 mg/kg	≤ 18%	≤ 25%
> 1 mg/kg	≤ 14%	≤ 19%

Requirement 28: Ruggedness or robustness
Ruggedness or robustness is expressed as the ratio between RSD_R/RSD_r.

Pre-registration:
EU [1]: not addressed.
FAO [3]: not addressed.
RIVM procedure for EU and FAO: not evaluated.

Post-registration:
EU [2]: not addressed.
FAO [3, 4]: not addressed.
Updated draft Codex [5]: not addressed
RIVM procedure for EU and FAO: not evaluated.

Requirement 29: Verification of the method during sample analysis
In most cases the validation of the method is conducted before the actual residue samples are analysed and the validation is presented in a separate validation report. But the validity of the method needs to be verified during actual sample analysis.

Pre-registration:
EU [1]: required, but no criterion given.
FAO [3]: not addressed.
RIVM procedure for EU and FAO: Experiments required to verify that the method is still valid at the actual sample analysis, are limited to one recovery experiment per sample batch at at least one concentration level within the valid range (requirement 18) and one matrix interference experiment (requirement 15) for at least one sample per sample batch.

Post-registration: not applicable.

Requirement 30: Quality Assurance during sample analysis
In most cases the validation of the method is conducted before the actual residue samples are analysed and the validation is presented in a separate validation report. At this point a method is not subject to GLP (or other QA systems). But when the method is used to generate data for safety purposes (for example residues data) data generation must be conducted to GLP (or other QA systems).

Pre-registration:
EU [1]: GLP: residue data generation must be conducted to GLP.
FAO [3]: GLP or national QA.
RIVM procedure for EU: GLP required. For national assessments very often old reports without QA-statements are submitted for assessment. Therefore a date is required from whereon QA-statements in documents are required. CTB (Dutch Board for the Authorisation of Pesticides) has defined 25th July 1991 from whereon GLP is required. Study reports without GLP are not acceptable, unless the study report is dated before 25th July 1991.
RIVM procedure for FAO: GLP or ISO 17025 required.

Post-registration: not applicable.

3.2.3 Evaluation of residue data for pre-registration purposes

From §2.2 it can be deduced that a pre-registration method is considered "valid for measurement range X-Y of analyte A in crop group C or animal product A" when all criteria for validation requirements 2, 5, 6, 7 (EU only), 10, 11, 12 (EU only), 13 (EU only), 14 (EU only), 15, 17 (EU only), 18, 19, 22/23 (if applicable), 24/25 (if applicable) are fulfilled (see appendix V and VI). For the validity of the method itself requirement 29 and 30 are not assessed, these requirements are connected to the different residue studies submitted. The results for a residue study are considered valid when a valid method is used *and* when requirement 29 and 30 are fulfilled. The reviewer can ask for missing data. The validation of the method can be done after sample analysis, but verification during sample analysis and GLP compliance during sample analysis cannot be conducted in a later stage.

Consequences
Conclusions about the validity of residue data generated by pre-registration methods have impact on the residue data that are selected for the setting of an MRL, STMR, HR, processing factor or stability periods.
When the pre-registration method is considered not valid and/or when the method is not verified during sample analysis, the results generated by this method may not be used for the setting of an MRL, STMR, HR, processing factor or stability period and the reviewer has to ask for new studies conducted according to the guidelines.
Special attention should be given to residue data that are generated by methods for which the LOQ_{valid} is considered higher than the $LOQ_{reported}$. When values in the range $<LOQ_{reported}$ to $<LOQ_{valid}$ have to be used in the calculation of an MRL, these values should be modified into $<LOQ_{valid}$ before MRL calculation.

3.2.4 Evaluation of residue analytical methods for post-registration purposes

From §2.2 it can be deduced that a post-registration method is considered "valid for measurement range X-Y of analyte A in crop group C or animal product A" when all criteria for validation requirements 2, 3, 4 (EU only), 5, 6, 7, 8 (EU only), 10, 11, 12 (EU only), 13 (EU only), 14 (EU only), 15, 18, 19, 20 (FAO only), 22/23 (if applicable), 24/25 (if applicable), 27 (FAO only) are fulfilled (see appendix V and VI). The reviewer can ask for missing data. For post-registration the timing of the validation is not relevant. The individual performance characteristics may be presented in separate reports from different periods of time.

Old pesticides
For *old pesticides* an enforcement method already exists [10-20]. For old pesticides less stringent criteria are applied, because enforcement methods are widely used and/or already existed before the criteria were defined. It is sufficient if the applicant refers to one of these existing enforcement methods and it is not necessary to ask for a full validation. It is only necessary to ask for an (extension of the) validation if the residue definition is changed, if the residue definition is extended to animal products or if the application is extended to other crop groups.
If the applicant suggests a new method is better than the current enforcement method, than the new method has to comply with all the criteria for post-registration (§2.2). Missing validation data can be asked for by the reviewer.

New pesticides
For *new pesticides* there is no existing enforcement method. The newly developed analytical method has to comply with all the criteria for post-registration.

Consequences for setting MRLs
The proposed MRL is based on residue data produced by valid pre-registration analytical methods. The setting of the MRL seems therefore not dependent on the validity of the enforcement method. But there are two cases where the setting of the MRL depends on the enforcement method.

1. When the proposed MRL from pre-registration studies turns out to be lower than the LOQ of the enforcement method, the MRL is set at the LOQ of the enforcement method. The actual LOQ of the pre-registration method should be recorded as a lower boundary in order to allow a reduction in MRLs when the enforcement method improves at a later stage.
2. When no enforcement method is available, the LOQ of the enforcement method is not known and the final MRL cannot be established. In cases where a valid enforcement method is not available, all proposed MRLs are provisional.

References

1. European Commission. Residues: Guidance for generating and reporting methods of analysis in support of pre-registration data requirements for Annex II (part A, Section 4) and Annex III (part A, Section 5) of Directive 91/414. Working document. SANCO/3029/99 rev. 4. 11-07-2000.
2. European Commission. Guidance document on residue analytical methods. SANCO/825/00 rev. 6. 20-06-2000.
3. FAO. FAO manual on the submission and evaluation of pesticide residues data for the estimation of maximum residue levels in food and feed. Rome. 1997. pp 9, 12, 13, 23, 72.
4. Codex Alimentarius. Volume II. 1993. Section 4.2 and 4.3.
5. Joint FAO/WHO food standards programme Codex Alimentarius Commission Twenty-fourth session. Geneva, 2-7 July 2001. Report of the thirty-third session of the Codex Committee on Pesticide Residues. ALINORM 01/24A. The Hague, 2-7 April 2001. Appendix VII and VIII. (="Updated draft Codex").
6. AOAC/FAO/IAEA/IUPAC Expert consultation. Guidelines for single-laboratory validation of analytical methods for trace-level concentrations of organic chemicals. In "Principles and practices of method validation" Eds. A. Fajgelj and A. Ambrus. Royal Society of Chemistry. Cambridge. UK. 2000. pp179-252.
7. Codex Alimentarius Commission. Guide to codex recommendations concerning pesticide residues. Part 4 Codex classification of foods and animal feeds. CAC/PR 4-1986.
8. IUPAC. Harmonised guidelines for single-laboratory validation of methods of analysis. Draft November 1999.
9. Nederlands Normalisatie Instituut. Ontwerp NEN 7777. Prestatiekenmerken van meetmethoden. Juli 2000.
10. Official Methods of AOAC INTERNATIONAL, 16th edition (1995)
    - (a) 970.52
    - (b) 976.23
    - (d) 977.18
    - (e) 975.40
    - (j) 978.16
    - (k) 977.19
    - (l) 960.43
    - (m) 963.24
    - (n) 983.21
    - (o) 984.21
    - (p) 985.22
    - (q) 985.23

11. Pesticide Analytical Manual, Food and Drug Administration, Washington, D.C., USA

		2nd edition	3rd edition
(a)	Vol. I,	Table 201-A and sections, 211.1, 212.1, 231.1, 232.1 and 252	non fat foods: Section 303 fatty foods: Section 304, E1-E5+C1-C4
(b)	Vol. I,	Table 201-D and section 221.1	Section 402
(c)	Vol. I,	Table 201-H and section 232.3	[method not in PAM I 3rd edition]
(d)	Vol. I,	Table 201-I and section 232.4	Section 302 E1-E4, no cleanup
(e)	Vol. II,	Method under compound name (when in this reference several methods have been given, they are generally listed in order of preference)	
(f)	Vol. I,	Table 651-A and sections 650 and 651	[not in PAM I 3rd edition]
(g)	Vol. I,	Table 242.2-1 and section 242.2	Section 401
(h)	Vol. I,	Section 242.3	Section 404

12. Manual on Analytical Methods for Pesticide Residues in Foods, Health Protection Branch, Health and Welfare Canada, Ottawa, Ont., Canada (1985) (available in English and French)
13. Methodensammlung zur Rückstandsanalytik von Pflanzenschutzmitteln, 1.- 11. Lieferung, VCH Verlagsgesellschaft, Weinheim, FRG (1991) (the numbers in parentheses refer to the numbers of the

methods in this manual; methods preceded by “S” are multi residue methods; the manual is also available in English, see ref. 16).

14. Laboratory Manual for Pesticide Residues Analysis in Agricultural Products, compiled by R.B. Maybury, Pesticide Laboratory, Food Production and Inspection Branch, Agriculture Canada, Ottawa, Ont., Canada (1984) (available in English and French).
15. Zweig, G. (edit.), Analytical Methods for Pesticides, Plant Growth Regulators , Academic Press, New York – San Francisco – London
    (a) Vol. VII (1974)
    (b) Vol. VIII (1976)
    (c) Vol. IX (1977)
    (d) Vol. X (1978)
    (e) Vol. XI (1980)
    (f) Vol. XII (1982) (Lawrence J.F. Editor)
    (g) Vol. XIII (1984) (Zweig, G. and Sherma, J. Editors)
    (h) Vol. XVI (1988) (Sherma, J. Editor)
16. Manual of Pesticide Residue Analysis, Deutsche Forschungsgemeinschaft, VCH Verlagsgesellschaft, Weinheim, FRG (1987) (English translation of ref. 13)
    (a) Vol. I, Section Clean-up Methods (the numbers in parentheses refer to the numbers of the clean-up methods in this volume)
    (b) Vol. I, Section Individual Pesticide Residue Analytical Methods
    (c) Vol. I, Section Multiple Pesticide Residue Analytical Methods (the numbers in parentheses refer to the numbers of the multi residue methods in this volume)
    (d) Vol. II (1992).
17. Chemistry Laboratory Guidebook, United States Department of Agriculture, Food Safety and Inspection Service, Science Program, Washington, D.C., USA
    (a) Section 5.001
    (b) Section 5.002
    (c) Section 5.003
    (d) Section 5.004
    (e) Section 5.006
    (f) Section 5.005
    (g) Section 5.050
18. Analytical Methods for Pesticides Residues in Foodstuffs, P. van Zoonen (Ed.), 6th edition, The Inspectorate for Health Protection, Ministry of Public Health, Welfare and Sport, Drukkerij T.O. Offset B.V, Maastricht, The Netherlands (1996).
    (a) Part I: Multi residue Methods (the numbers in parentheses refer to numbers of the multi residue methods in this volume)
    (b) Part II: Special Methods (methods given under compound name)
19. Materials and Methods Used for Pesticide Residues Monitoring in Sweden, Vår Föda, 38, Suppl.2, 79-136 (1986)
20. Comprehensive Analytical Profiles of Important Pesticides (Modern methods for pesticides analysis) Eds. J. Sherma & T Cairns 1992

Appendix I Overview of requirements for validation of pre-registration residue analytical methods

Requirements	EU-criteria for pre-registration [1]	FAO criteria for pre-registration [3]	RIVM criteria for pre-registration for EU	RIVM criteria for pre-registration for FAO
1. Quality Assurance for method validation	not required	-	Taken up in summary, but not taken into account	taken up in summary, but not taken into account
2. Full description of the method	Required	required	See EU	see FAO
3. Commonly available analytical techniques	not required	not required	not evaluated	not evaluated
4. Multi residue method (MRM)	not required	not required	not evaluated	not evaluated
5. Validation report for all sample matrices	Required	-	Required, see RIVM interpretation	required, see RIVM interpretation
6. Validation report for all components of the residue definition	Required for MRL-compliance and risk assessment	required for MRL-compliance and risk assessment	See EU	see FAO
7. Confirmatory techniques	only required for common moiety or non-specific methods	-	ratio of mean results = 0.8-1.2 ($n \geq 1$)	not evaluated
8. Independent laboratory validation (ILV)	not required	-	not evaluated	not evaluated
9. Limit of detection (LD)	-	-	not evaluated	not evaluated
10. Limit of quantification (LOQ)	Required; no criteria given	-	<0.5x min residue concentration or <0.01 mg/kg (whichever is higher)	<0.5x min residue concentration or <0.01 mg/kg (whichever is higher)
11. Measurement range	LOQ to 10LOQ or LOQ – likely residue levels (whichever is higher)	-	LOQ to 10LOQ or LOQ – likely residue levels (whichever is higher)	relevant residue concentrations should be covered
12. Calibration model	5 conc. levels or duplicates at 3 conc. levels;		see EU	not evaluated
13. Lack of fit of calibration model	Required; no criteria given	-	linear and weighted linear: $r \geq 0.99$ and $S_{y/x} \leq 0.1$; polynomial: $r \geq 0.98$	not evaluated
14. Matrix effects	Required; no criteria given; matrix matched standards are preferred	-	if the difference between slopes is significant (P=0.05) matrix matched standards are required	not evaluated

15. Matrix interferences	≤0.3LOQ (n≥2)	-	see EU	see EU
16. Pesticide interferences	-	-	not evaluated	not evaluated
17. Metabolite, isomer or analogue interferences	Required, when necessary for carrying out risk assessment; no criteria given	-	≤0.3LOQ (n≥1) or peak maxima should be separated by at least one full width at 10% of the peak height.	not evaluated
18. Mean recovery (Q_{mean}) for analysis excluding sample processing	spike at 2 levels, 70-110% (n≥5) for each level	-	see EU	see EU
19.Repeatability (RSD_r or $CV_{A,r}$) for analysis excluding sample processing	spike at 2 levels, ≤20%; (n≥5) for each level	≤10%	see EU	spike at 2 levels, ≤10%; (n≥5) for each level
20. Within-laboratory reproducibility for analysis excluding sample processing, (RSD_R; $CV_{A,R}$)	-	-	not evaluated	not evaluated
21. Individual recoveries for analysis excluding sample processing	-	-	not evaluated	not evaluated
22. Mean recovery for derivatisation (if applicable)	data required; no criteria given	-	70-110% (n≥3)	70-110% (n≥3)
23. Precision for derivatisation (if applicable)	data required; no criteria given	-	≤10% (n≥3); repeatability conditions	≤10% (n≥3); repeatability conditions
24. Recovery for internal standard (if applicable)	-	-	70-110% (n≥3)	70-110% (n≥3)
25. Precision for internal standard (if applicable)	-	-	≤10% (n≥3); repeatability conditions	≤10% (n≥3); repeatability conditions
26. Trueness (recovery of incurred residues)	-	-	not evaluated	taken up in summary, but not taken into account
27. Homogeneity (CV_{Sp})	-	-	not evaluated	not evaluated
28. Ruggedness or robustness (RSD_R/RSD_r)	-	-	not evaluated	not evaluated
29. Verification of the method during sample analysis (required for all submitted residue studies, but not in method validation report)	Required, no criterion given	-	for each sample batch: requirement 18 for n=1 and requirement 15 for n=1	for each sample batch: requirement 18 for n=1 and requirement 15 for n=1
30. Quality Assurance during sample analysis (required for all submitted residue studies but not in method validation report)	GLP	GLP or national QA	GLP	GLP or ISO 17025

- not addressed in the guidance document;

Appendix II Overview of requirements for validation of post-registration residue analytical methods

Requirements	EU-criteria post-registration [2]	FAO/Codex criteria; post-registration [3, 4]	Updated draft Codex criteria [5]	RIVM criteria for post-registration for EU	RIVM criteria for post-registration for FAO
1. Quality Assurance for method validation	not required	-	-	taken up in summary, but not taken into account	taken up in summary, but not taken into account
2. Full description of the method	Required	published method required	published method required	see EU	see updated draft Codex
3. Commonly available analytical techniques	Required	Required	required	see EU	see updated draft Codex
4. Multi residue method (MRM)	required, unless MRM not possible	Preferred	preferred	required, unless MRM not possible	taken up in summary, but not taken into account
5. Validation report for all sample matrices	4 crop groups and 6 animal product groups	-	MRM: 6 crop groups and 5 animal product groups; SRM: all matrices specified	see EU	see updated draft Codex
6. Validation report for all components of the residue definition	required for MRL-compliance	required for MRL-compliance	required for MRL-compliance	see EU	see updated draft Codex
7. Confirmatory techniques	required; no criteria given	when MRL is exceeded; within 20% of primary result	-	ratio of mean results = 0.8-1.2 ($n \geq 1$)	ratio of mean results = 0.8-1.2 ($n \geq 1$)
8. Independent laboratory validation (ILV)	2 plant groups and 2 animal products; criteria see requirement 11, 15, 18, 19	preferred; $CV_R < 77\%$ (0.01 mg/kg); $CV_R < 45\%$ (0.1 mg/kg); $CV_R < 22\%$ (1 mg/kg)	-	2 plant groups and 2 animal products; criteria see requirement 11, 15, 18, 19/20	taken up in summary, but not taken into account
9. Limit of detection (LD)	-	required, no criteria given	required, no criteria given	not evaluated	taken up in summary, but not taken into account

Requirements	EU-criteria post-registration [2]	FAO/Codex criteria; post-registration [3, 4]	Updated draft Codex criteria [5]	RIVM criteria for post-registration for EU	RIVM criteria for post-registration for FAO
10. Limit of quantification (LOQ)	≤0.1 mg/kg if MRL > 0.1 mg/kg or ≤0.5MRL if MRL≤0.1 mg/kg, except when MRL=LOQ	-	LCL = 0.5 if MRL ≥5 mg/kg; LCL =0.1-0.5 if MRL ≥0.5-<5 mg/kg LCL = 0.02-0.1 if MRL = ≥0.05-<0.5 mg/kg; LCL = 0.5MRL if MRL <0.05 mg/kg, except when MRL = LCL	see EU	see updated draft Codex, except that LCL=LOQ.
11. Measurement range	LOQ-MRL or LOQ-10LOQ (whichever is broader)	-	LCL – 2-3 AL	see EU	see updated draft Codex, except that LCL=LOQ; AL=MRL
12. Calibration model	5 conc. levels or duplicates at 3 conc. levels;	-	linear: duplicates at 4 conc. levels; polynomial: triplicates at 7 concentrations;	see EU	not evaluated
13. Lack of fit of calibration model	required; no criteria given	-	linear: r≥0.99 and $S_{y/x}$≤0.1; polynomial: r≥0.98	linear and weighted linear: r≥0.99 and $S_{y/x}$≤0.1; polynomial: r≥0.98	not evaluated
14. Matrix effects	data required; no criteria given; matrix matched standards are preferred	-	If the difference between slopes is significant (P=0.05) matrix matched standards are required	If the difference between slopes is significant (P=0.05) matrix matched standards are required	not evaluated
15. Matrix interferences	≤0.3LOQ (n≥2)	-	required; no criteria given, except n≥5	see EU	≤0.3*LOQ (n≥5)
16. Pesticide interferences	-	-	at AL peak maxima should be separated by at least one full width at 10% of the peak height	not evaluated	not evaluated

Requirements	EU-criteria post-registration [2]	FAO/Codex criteria; post-registration [3, 4]	Updated draft Codex criteria [5]	RIVM criteria for post-registration for EU	RIVM criteria for post-registration for FAO
17. Metabolite, isomer or analogue interferences	required if necessary for risk assessment, no criteria given	-	-	≤0.3LOQ (n≥1) for solution without analyte. At MRL peak maxima should be separated by at least one full width at 10% of the peak height for solution with analyte	not evaluated
18. Mean recovery (Q_{mean}) for analysis excluding sample processing	spike at 2 levels, 70-110% (n≥5) for each level	≥80%	spike at LCL, AL (n≥5) and 2-3 AL (n≥3); 50-120% (≤0.001 mg/kg) 60-120% (>0.001-≤0.01 mg/kg); 70-120% (>0.01-≤0.1 mg/kg); 70-110% (>0.1 mg/kg)	see EU	see updated draft Codex, except LCL=LOQ and AL=MRL
19.Precision: repeatability (RSD_r or $CV_{A,r}$) for analysis excluding sample processing	spike at 2 levels, ≤20%; (n≥5) for each level	≤10%	spike at LCL, AL (n≥5) and 2-3 AL (n≥3) ≤35% (≤0.001 mg/kg) ≤30% (>0.001-≤0.01 mg/kg) ≤20% (>0.01-≤0.1 mg/kg); ≤15% (>0.1-≤1 mg/kg); ≤10% (> 1 mg/kg)	see EU	see updated draft Codex, except LCL=LOQ and AL=MRL

Requirements	EU-criteria post-registration [2]	FAO/Codex criteria; post-registration [3, 4]	Updated draft Codex criteria [5]	RIVM criteria for post-registration for EU	RIVM criteria for post-registration for FAO
20. Precision: within-laboratory reproducibility for analysis excluding sample processing, (RSD_R; $CV_{A,R}$)	-	-	spike at LCL, AL (n≥5) and 2-3 AL (n≥3) <53% (≤0.001 mg/kg) ≤45% (>0.001-≤0.01 mg/kg) ≤32% (>0.01-≤0.1 mg/kg); ≤23% (>0.1-≤1 mg/kg); ≤16% (> 1 mg/kg)	not evaluated	see updated draft Codex, except LCL=LOQ and AL=MRL
21. Individual recoveries for analysis excluding sample processing	-	70%-110%	-	not evaluated	not evaluated
22. Mean recovery for derivatisation (if applicable)	≥70%	-	-	70-110% (n≥3)	70-110% (n≥3)
23. Precision for derivatisation (if applicable)	data required; no criteria given	-	-	≤10% (n≥3); repeatability conditions	≤10% (n≥3); repeatability conditions
24. Recovery for internal standard (if applicable)	similar as analyte	-	-	70-110% (n≥3)	70-110% (n≥3)
25. Precision for internal standard (if applicable)	similar as analyte	-	-	≤10% (n≥3); repeatability conditions	≤10% (n≥3); repeatability conditions
26. Trueness (recovery of incurred residues)	-	preferred; no criteria given	mean residue (n≥5) is not significantly different (P=0.05) from consensus value	taken up in summary, but not taken into account	taken up in summary, but not taken into account
27. Homogeneity (CV_{Sp})	-	-	≤10% (n≥5); otherwise increase analytical sample portions	not evaluated	(n≥5) repeatability or within laboratory reproducibility including sample processing below levels from 3.3.
28. Ruggedness or robustness (RSD_R/RSD_r)	-	-	-	not evaluated	not evaluated

- not addressed in the guidance document

Appendix III Glossary of terms

Analyte [6]	The chemical substance sought or determined in a sample
Limit of detection [6] (aantoonbaarheidsgrens); LD	Smallest concentration of the analyte that can be identified. Commonly defined as the minimum concentration of analyte in the test sample that can be measured with a stated probability that the analyte is present at a concentration above that in a blank sample. IUPAC and ISO have recommended the abbreviation LD.
Limit of quantification [6] (bepalingsgrens); LOQ	Smallest concentration of the analyte that can be quantified. Commonly defined as the minimum concentration of analyte in a test sample that can be determined with acceptable accuracy (trueness and precision) under the stated conditions of the test.
Measurement range; (meetbereik)	Concentration range in a specified matrix for which a method is valid.
Calibration model (kalibratiemodel)	Mathematical relation between the concentration of the analyte and the detector response. The calibration model is a either a linear, a weighted linear or a polynomial function.
Interferences (interferenties)	Measure of the degree to which the analyte is likely to be distinguished from other sample components, either by separation (e.g. chromatography) or by the relative response of the detection system. Interferences influence the intercept of the calibration model.
Matrix effects (matrixeffecten)	Measure of the degree to which the analyte concentration is likely to be influenced by other sample components, either at separation (e.g. chromatography) or at the relative response of the detection system. Matrix effects influence the slope of the calibration model.
Precision [1] (precisie)	Closeness of agreement between independent test results obtained under prescribed conditions. A measure of random errors, which may be expressed as repeatability, within-lab reproducibility or between-lab reproducibility.
Repeatability [6] (herhaalbaarheid)	Precision (=random error) under repeatability conditions, i.e., conditions where independent test results are obtained with the same method on replicate analytical portions in the same laboratory by the same operator using the same equipment within short intervals of time (ISO 3534-1).

Within-laboratory reproducibility [6] (binnen-laboratorium-reproduceerbaarheid)	Precision (=random error) under within-laboratory reproducibility conditions, i.e., conditions where independent test results are obtained with the same method on replicate analytical portions in the same laboratory with different operators (if present) and using different equipment (if available) covering a long period of time.
Between-laboratory reproducibility [6] (tussen-laboratorium-reproduceerbaarheid)	When the above mentioned test is conducted in different laboratories, the between-laboratory reproducibility is obtained.
Recovery [6] (terugvinding)	Fraction or percentage of an analyte recovered following extraction and analysis of a blank sample to which an analyte has been added at a known concentration (spiked sample).
Trueness [6] (juistheid)	Closeness of agreement between the average value obtained from a large series of test results and an accepted reference value (=systematic error).
Homogeneity [6] (homogeniteit)	Uniformity of dispersion of the analyte in matrix.
Robustness or ruggedness [6] (robuustheid)	Ability of a chemical measurement process to resist changes in test results when subjected to minor changes in environmental and method procedural variables, laboratories, personnel, etc.

Appendix IV Abbreviations

Abbreviation	Name
AL	Accepted limit (in this context equivalent to MRL)
AOAC	Association of Official Analytical Chemists
CTB	Dutch Board for the Authorisation of Pesticides
CV_R	Coefficient of Variation under between-laboratory reproducibility conditions (CV=RSD)
$CV_{A,r}$	Coefficient of Variation for analysis excluding sample processing under repeatability conditions (CV=RSD)
$CV_{A,R}$	Coefficient of Variation for analysis excluding sample processing under within-laboratory reproducibility conditions (CV=RSD)
$CV_{L,r}$	Coefficient of Variation for analysis including sample processing under repeatability conditions (CV=RSD)
$CV_{L,R}$	Coefficient of Variation for analysis including sample processing under within-laboratory reproducibility conditions (CV=RSD)
CV_{Sp}	Coefficient of Variation for sample processing
EU	European Union
FAO	Food and Agricultural Organisation of the United Nations
GAP	Good Agricultural Practice
GLP	Good Laboratory Practice
HR	Highest Residue
IAEA	International Atomic Energy Agency
ILV	Independent Laboratory Validation
ISO	International Standard Organisation
IUPAC	International Union for Pure and Applied Chemistry
LCL	Lowest calibrated level (in this context equivalent to LOQ)
LD	Limit of detection
LOQ	Limit of quantification
MRL	Maximum Residue Level or Maximum Residue Limit
MRM	Multi Residue Method
n	Number of samples
P	Probability
Q	Recovery of an individual spike
Q_{mean}	Mean recovery of spikes
QA	Quality Assurance
r	Correlation coefficient
RIVM	National Institute of Public Health and the Environment
RSD_r	Relative Standard Deviation under repeatability conditions
RSD_R	Relative Standard Deviation under reproducibility conditions
s	Standard deviation
$S_{y/x}$	Standard deviation of residuals
SRM	Single or Special Residue Method
STMR	Supervised trial median residue

Appendix V Checklist for national and EU assessments

Requirement (see §2.2, 2.3)	pre-registration	post-registration
1. Quality assurance for method validation		
2. Full description of the method		
3. Commonly available analytical methods		
4. Multi residue method		
5. Validation report for all sample matrices		
6. Validation report for all components of the residue definition		
7. Confirmatory techniques		
8. Independent laboratory validation		
9. Limit of detection		
10. Limit of quantification		
11. Measurement range		
12. Calibration model		
13. Lack of fit of the calibration model		
14. Matrix effects		
15. Matrix interferences		
16. Pesticide interferences		
17. Metabolite, isomer, analogue interferences		
18. Mean recovery for analysis excluding sample processing		
19. Precision: repeatability for analysis excluding sample processing		
20. Precision: within-laboratory reproducibility for analysis excluding sample processing		
21. Individual recoveries		
22. Mean recovery for derivatisation		
23. Precision for derivatisation		
24. Mean recovery for internal standards		
25. Precision for internal standards		
26. Trueness		
27. Homogeneity		
28. Ruggedness		
29. Verification of the method during analysis		
30. Quality assurance during sample analysis		

grey areas = not required

Appendix VI Checklist for FAO assessments

Requirement (see §2.2, 2.3)	pre-registration	post-registration
1. Quality assurance for method validation		
2. Full description of the method		
3. Commonly available analytical methods		
4. Multi residue method		
5. Validation report for all sample matrices		
6. Validation report for all components of the residue definition		
7. Confirmatory techniques		
8. Independent laboratory validation		
9. Limit of detection		
10. Limit of quantification		
11. Measurement range		
12. Calibration model		
13. Lack of fit of the calibration model		
14. Matrix effects		
15. Matrix interferences		
16. Pesticide interferences		
17. Metabolite, isomer, analogue interferences		
18. Mean recovery for analysis excluding sample processing		
19. Precision: repeatability for analysis excluding sample processing		
20. Precision: within-laboratory reproducibility for analysis excluding sample processing		
21. Individual recoveries		
22. Mean recovery for derivatisation		
23. Precision for derivatisation		
24. Mean recovery for internal standards		
25. Precision for internal standards		
26. Trueness		
27. Homogeneity		
28. Ruggedness		
29. Verification of the method during analysis		
30. Quality assurance during sample analysis		

grey areas: not required

4 Sediment risk assessment for pesticides

Factsheet FSM-004/00 date 20-02-2002

Authors:

S. Ciarelli and M.H.M.M. Montforts

4.1 Introduction

4.1.1 Background and purposes of this factsheet

Within the regulatory framework of pesticide registration, an ecological risk assessment for aquatic, terrestrial and for soil micro-organisms is performed. The risk evaluation is based on the combination of results obtained from studies about the fate and behaviour of the compounds of concern on one hand, and toxicity results obtained from the different bioassays on the other hand. For each of the taxonomic groups official and specific test methods (e.g. OECD/EPA guidelines) are available. In these routine procedures, however, toxicity tests with sediment-dwelling organisms are not required and risk assessment for these organisms is therefore not taken into account excepted for special cases (e.g. “higher tier” risk assessments). In these cases only one test involving a water-sediment system with spiked water using the midge larvae, *Chironomus sp.* is encouraged by EU-guidances. This test is, however, not unanimously accepted yet within the European Community as it poses some problems in the interpretation and in the reliability of results. In addition, the mode of exposure of *Chironomus sp.* is another issue of debate as this is not representative for the majority of sediment-dwelling organisms (these items are discussed in the factsheet).
Presently, official Dutch or European guidelines describing test methods that can measure toxic effects of pesticides to sediment-dwelling organisms or possible alternatives to the *Chironomus* test, are not available.
The purposes of this factsheet are multiple:

- to document which (standardised) laboratory sediment spiked-test methods are available and can appropriately measure toxic effects of pesticides to sediment-dwelling organisms;
- to document a possible decision procedure that identifies pesticides for which sediment toxicity tests would be required in the regulatory framework of pesticide registration (91/414/EEC);
- to inform assessors performing assessment at RIVM on behalf of the Bord for the Authorisation of Pesticides in the regulatory framework of pesticide registration on possible alternatives concerning sediment toxicity evaluations;
- to produce a document which can serve as an opinion document in the revision process of the ECCO/EU guidances on sediment risk assessments issues.

4.1.2 Sediment Quality Criteria: problem definition

Sediments are final repositories for many persistent chemicals that are introduced in the aquatic environments due to their strong association with sediment particles. Sediments may also act as a source of contaminants to the water column due to physical processes such as diffusion, fugacity and resuspension and biological processes (i.e. bioturbation) produced by sediment organisms burrowing activities (Burton, 1991; Ciarelli et al., 1999; 2000). Although many chemicals are highly sorbed to sediment particles, they may still be directly available to sediment organisms through pore and/or overlying water uptake or sediment ingestion, or indirectly, through accumulation in the aquatic food chain. Assessments of sediment quality have been based in the past on chemical analyses on one hand, and on benthic community surveys on the other hand. Both methods, however, did not provide sufficient insight into the bioavailability of sediment contaminants and could not predict appropriately potential adverse effects caused to benthic organisms (USEPA, 2000).
In sediment risk assessments, the impact of pesticides to sediment-dwelling organisms has, until recently, not been considered. Clear guidance on the risk assessment for pesticides in

agricultural use is lacking. Within the framework of pesticide registration sediment should be considered as a compartment of concern. It was assumed that pesticides in the water column were available and taken up by aquatic organisms only and that the bioavailability and toxicity of bound pesticides to suspended or bottom sediments were reduced to benthic organisms (ECOFRAM, 1999). Increasing evidence that this assumption is not universally correct led to the development of standard sediment laboratory tests methods able to measure directly toxic effect to sediment organisms (Burton, 1991; USEPA, 2000; EPA, 1998). In pesticides risk assessments methodology of the Netherlands, environmental quality standards are based on Maximum Permissible Concentrations (MPCs) and Negligible concentrations (NCs) which are directly derived from ecotoxicological studies, when available. If sediment toxicity data are lacking, sediment MPCs and NCs are derived indirectly from those from water by applying the equilibrium partitioning method (EP-method). However, due to the uncertainties and discussions underlying this method, the use of sediment toxicity tests data is preferable to derive (more) reliable environmental risk limits (Crommentuijn et al., 2000). Besides benthic (macro) organisms also micro-organisms play an important role in sediments because anaerobic micro-organisms live within the sediment whereas benthic animals depend on oxygen from the overlying water. Various studies have shown that micro-organisms are important in the degradation of natural and xenobiotic substances and in the cycling of nutrients and elements. Moreover the anaerobic processes in sediments can be much more sensitive for chlorinated compounds or metals than fish or aquatic invertebrates (Van Vlaardingen and Van Beelen, 1992; Van Beelen and Van Vlaardingen, 1994). In anaerobic freshwater sediments, methanogenesis is a vital process, which prevents the acidification of the aquatic ecosystem by organic acids (Hines et al., 1994). Recently, a draft standard test ISO (International Standard Organisation) for the determination of inhibition of activity of anaerobic bacteria was developed to measure the sensitivity of the methanogenesis for pollutants (ISO, 2001). Since the sensitivity of microbial processes and benthic animals for a specific compound can differ many orders of magnitude, both type of organisms tests should be incorporated in pesticides sediment risk assessments.

4.1.3 Sediment-spiked tests objectives

The main objective of sediment spiked tests in general is to measure possible biological effects of one or more contaminants which occur in field sediments or in sediments which have been spiked in the laboratory at a range of concentrations. Sediment tests can be used for several purposes: (1) to establish toxic effects and bioavailability of chemicals, (2) to evaluate dredge material and to rank areas for clean-up, (3) to determine spatial and temporal distribution of contamination, (4) to compare sensitivity of different organisms. Specific goals of a sediment-spiked test are: (1) to establish cause-response relationship between a specific chemical and adverse biological response and (2) to predict the risk for sediment-dwelling organisms of single substances (e.g. pesticides) which are expected to reach sediments. A variety of standard sediment test methods have been published for amphipods, midges, polychaetes, oligochaetes and mayflies and several endpoints were suggested such as mortality, growth, reproduction or behaviour by Burton (1992), Hill et al. (1993), USEPA (1994 a; b; 2000) and ASTM (1993, 1999a) guidelines. In the OECD's work on Environment Health and Safety, two new guidelines: "Sediment-water Chironomid Toxicity Test Using Spiked Sediment" (218) and "Sediment-Water Chironomid Toxicity Test Using Spiked Water" (219) have been proposed.

4.2 Risk assessment for sediment

4.2.1 Exposure assessment: risk characterisation for sediments

Risks characterisation for sediments can be performed in the analogous way as it is done for water and soils, i.e. by expressing the risk quotient PEC/PNEC (= Predicted Environmental Concentration/Predicted No Effect Concentration) ratios, provided that the method for calculating the PEC is known and reliable and that PNEC values are derived from ecotoxicity data with sediment organisms. Where the predicted no effect concentration is exceeded, i.e. PEC/PNEC > 1, a risk might be present.

4.2.1.1 PEC calculation

The currently method used by USES 3.0 (RIVM, 1999) for calculating PEC in sediments is based on the assumption of equilibrium partitioning between the sediment solids and the water. The PEC sediment for a certain time of exposure (T), ($PEC_{sediment,}$ T) is calculated in the following way:

$$PEC_{sediment,\ T} = (PEC_{water,\ T} * K_{sediment\text{-}water}.* 1000)/(RHOsediment)$$

where:
the $PEC_{water,\ T}$ is the Predicted Environmental Concentration in water which can be calculated in different ways depending on the model used (Slootbox, Toxswa etc.), the number of applications and the period of time interval (T) in which the compound is applied.

The sediment is further defined as having a volume fraction water $Fwater_{sediment}$ ($m^3.m^{-3}$) of 0.8, a volume fraction solids $Fsolid_{sediment}$ ($m^3.m^{-3}$) of 0.2, and a density RHOsediment ($kg_{wwt}.m^{-3}$) of 1300 and considering an organic carbon content of the solids fraction of 5% ($Foc_{sediment}$). The bulk density of the solids $RHOsolid_{sediment}$ is 2500 ($kg.m^{-3}$).

$K_{sediment\text{-}water} = Fwater_{sediment} + Fsolid_{sediment} * Kp_{sediment} * RHOsolid_{sediment}$

$Kp_{sediment} = Koc*Foc_{sediment}$ ($dm^3.kg^{-1}$)
1000 = is used to convert the water density to the bulk density of the sediment

For highly lypofilic compounds with log K_{ow} greater than 5 the PEC/PNEC in sediment is increased by a factor of 10 to account for uptake via ingestion of sediment (USES 3.0, 1999). Whether the factor of 10 is enough protective or overestimates the risk for sediment organisms is questionable as this depends on factors such as the potential of the compounds to be biotransformed, on the feeding behaviour of the organisms and on the different factors that control bioavailability (Sijm et al., 2001).

4.2.2 Effect assessment

4.2.2.1 Sediment-spiked versus water-spiked test methods

The sediment spiked test is a widely accepted approach to predict toxicity on sediment-dwelling organisms. In a sediment-spiked test, the toxic substance is directly added to the sediment and the test organisms are placed in the sediment-water system after equilibrium time between sediment and (pore) water is established. Toxicity results (EC_{50}, LC_{50} and NOEC) obtained from sediment-spiked tests are expressed as mg/kg dwt. or mg/kg oc. A

large amount of literature is available on sediment-spiked tests and on different issues related to the sediment spiking (see section 3). Standard sediment-spiked tests are described by ASTM (1999), USEPA (2000) and OECD (2000, draft) guidelines.
Alternatively, in the 'water-spiked test' method proposed by the OECD (guideline 219, draft) for the chronic test with the midge larvae, *Chironomus tentans*/*riparius*, calculated volumes of a chemical are applied below the surface of the water with a pipette after the addition of the test organisms in the sediment-water system. This method is "intended to simulate a pesticide spray drift event and covers the initial peak of concentrations in pore water". The method pretends to provide "a more realistic exposure regime". However, some doubts are expressed on the effectivity of this method and on the interpretation of the test results. The water-spiked method might underestimate the exposure of the sediment organisms to the pesticides as it has the following shortcomings:

- addition of the chemical through the overlying water in the presence of the organisms might disturb the sediment-water system and has the disadvantage that mixing of the test substance will not occur homogeneously in vertically and horizontally direction; this might cause inaccurate sampling and unequal exposure of the sediment organisms;
- partitioning of the compound between sediment and water will require a longer time and steady state conditions might not be reached within the duration of the test; consequently, higher concentrations of the compound in sediment can be expected at a time later than the end of the exposure time;
- possible losses of the test compound due to glass adsorption (especially for highly hydrophobic compounds), volatilisation or/and degradation processes can be higher than in the case that the substance is directly added to the wet sediment;

In addition, the results (i.e. NOEC or EC_{50}/EC_{10}) of the water-spiked test expressed as mg/L and related to the amount of active ingredient spiked to the water, suggest erroneously that the exposure occurs through the overlying water only. Finally, the mode of exposure of the midge larvae (i.e. trough sediment water-interface) is not representative for the majority of sediment-dwelling organisms which can be exposed through the sediment directly or through the sediment pore water.
To avoid the shortcomings mentioned above and the uncertainties around the route of exposure, the sediment-spiked test approach is preferred and recommended for the risk assessment of pesticides in sediments.

4.2.2.2 Equilibrium partitioning method

The Equilibrium Partitioning approach (EP) is a widely accepted approach used to derive Environmental Risk Limits from aquatic data by predicting (pore) water concentration and normalised sediment concentrations (DiToro et al., 1991; OECD, 1992; Crommentuijn et al., 2000). NOECs for sediments can be derived from aquatic NOECs by predicting interstitial water concentrations normalised to sediment concentrations using the following formula:

$$NOEC\ (_{sed}) = K_{p(sed)} * NOEC\ (_{water})$$

where $NOEC_{sed}$ is the No effect Concentration for sediment (mg/kg dry weight), K_p is the solids-water partitioning coefficient (L/kg) and $NOEC_{water}$ the No observed effect Concentration for water (mg/L). The assumptions made by this model are: (a) the chemical concentration in sediment, interstitial water and benthic organisms are in thermodynamic equilibrium and concentration of the chemical in any of theses phases can be predicted using appropriate partition coefficients (K_p); (b) sediment-dwelling organisms and water column organisms are equally sensitive to the chemical; (c) bioavailability, bioaccumulation and

toxicity are closely related to the pore water concentrations. This method proved to be a valuable tool to predict toxicity of chemicals by calculating the concentration in the pore water for different substances among which the pesticides, DDT, endrin (Nebeker et al., 1989), chlorpyryfos (Ankley et al., 1994), dieldrin (Hoke et al., 1995), and for several others hydrophobic compounds (DiToro et al., 1991). However, several studies showed that the assumptions underlying this theory are not always correct and that this approach might underestimate toxicity for sediment-dwelling organisms. One of the arguments is that the pore water is not the only route of exposure but ingestion of sediment particles might be an additional source of exposure and for uptake (Belfroid et al., 1995; Meador et al., 1995). The biology of the sediment organisms and the feeding preferences are also important factors that can increase the exposure (Harkey et al., 1994; Kukkonen and Landrum, 1995; Driscoll & Landrum, 1997). Moreover, equilibrium might not be reached in the case of pesticides that are readily biodegradable in the environment (Belfroid et al., 1996). Other studies have shown that bioavailability or toxicity of compounds decrease with contact time (ageing) and that partition coefficients based on K_{ow} may lead to overestimate the risk (Landrum et al., 1992; Cornelissen et al., 1997; Kraaij et al., 2000).
However, despite the discussions on the assumptions underlying the EP method (Crommentuijn et al, 2000) and the awareness that a modified risk assessment method which takes into account the chemical and biological processes of chemicals bioavailability is needed (Peijnenburg et al., 2001), this method is still used in case that ecotoxicological data for sediment-dwelling organisms are lacking.

4.2.3 Criteria for sediment toxicity testing requirements

The actual pesticide risk assessment in The Netherlands is based on the quotient PEC/PNEC approach. Different quotients are calculated for the different ecosystems: the aquatic, the terrestrial, the predators (or secondary poisoning) and one for micro-organisms in the Sewage Treament Plant (STP). A risk assessment for sediment dwelling organisms is lacking. The need to develop a risk assessment approach which includes also sediment toxicity tests originates from the following considerations: (a) persistent pesticides can partition and accumulate into the sediment layer and cause directly or indirectly significant effects to sediment-dwelling organisms, (b) sediment-dwelling-organisms play an important role in the aquatic food chain and in the cycle of elements such as carbon (C) and nitrogen (N); (c) if the whole aquatic ecosystem need to be protected from the risk of the use of pesticides, effects assessments for sediment organisms should be incorporated. However, as not every chemical that enters aquatic ecosystems will accumulate in the sediment or be sufficiently toxic to affect the organisms inhabiting there, criteria need to be established when and which chemical or pesticide will require sediment toxicity testing.

4.2.3.1 State of the art according to the current literature on sediment risk assessment

Concerning criteria for sediment toxicity testing, the EU-guidance document on aquatic ecotoxicology - data requirements (Directive 91/414/EEC, Annex II, 8.2.7), for example, states that expert judgement should be used to decide whether an acute or chronic sediment toxicity test is required when an active substance is likely to partition to and persist in aquatic sediments; such expert judgement should take into account whether effects on sediment-dwelling invertebrates are likely by comparing the aquatic invertebrate toxicity EC_{50} data (acute and chronic) with the predicted levels of the active substances in sediment. Specific approaches or triggers for sediment toxicity testing are further not specified in this document.

In the EU-8075/VI/97 (draft) working document on the guidance document on aquatic ecotoxicology, it is stated that, a test on sediment-dwelling organisms should be required if the following two criteria are met:

1. In a sediment-water study, the distribution of total applied radioactivity indicates significant partitioning to sediment and this residue persists such that 10 % of the total applied radioactivity is measured in the sediment after day 14;
2. The NOEC in the chronic *Daphnia* test (or in a comparable study with insects) must be lower than 0.1 mg/L. Although some reservations are expressed on the analytical measurements in the *Chironomus* test, both the EU and the CTB (Board for the Authorisation of Pesticides- Aanvraagformulier B (Instructie onderdeel H.2.3.1-versie 0.1, 1999)) recommend the use of the sediment spiked (OECD 218) and the water spiked (OECD 219) tests with the midge larvae, *Chironomus sp.*

The EPA-OPPTS (Office of Prevention, Pesticides, and Toxic Substances, 1998) for example, will require sediment toxicity testing if any of the following conditions are met: (a) solubility ≤ 0.1 mg/L, K_{oc} > 50,000, $K_d \geq 1000$, $DT_{50} \geq 10$ days, or the concentration in the (pore) water is equivalent to concentrations known to be toxic in the water column.

On the basis of an evaluation study on a wide variety (140) of pesticides (including 39 insecticides, 64 herbicides and 37 fungicides), Maund et al. (1997), proposed the use of three trigger values for sediment toxicity testing of pesticides with benthic macroinvertebrates: partitioning (K_{oc}), persistency (DT_{50}) and potential to cause toxicity to *Daphnia*. Sediment testing should be required if all of the following conditions are met:

a) $K_{oc} \geq 1000$,

b) $DT_{50} \geq 30$ days for soil aerobic degradation and

c) *Daphnia* 48-hr acute toxicity value < 1 mg/L or 21-day NOEC < 0.1 mg/L.

This latter criterion derives from the assumption made by the EP- method that pelagic organisms have comparable sensitivity to benthic organisms and that toxicity to sediment-dwelling organisms is predictable on the basis of toxicity to water organisms. The partitioning trigger (K_{oc}) derives from the observation that for compounds with a K_{oc} lower than 1000, exposure via the sediment (organic content ranging from 1 to 3 %) may not be greatly different from that for pelagic organisms because less than 50 % of the compound will be present in the sediment phase. The persistence trigger ($DT_{50} \geq 30$ days) should be used as a guide as it has quite a high level of uncertainty due to problems of extrapolation from laboratory study conducted under aerobic conditions to the natural aquatic sediments. Degradation of chemicals could be either slower or faster under anaerobic conditions depending on the degradation route. Results from this study showed that within the chemical types, 44 % of all insecticides, 27 % of fungicides and 6 % of herbicides triggered sediment testing. This was explained by the fact that benthic macro-invertebrates are likely more susceptible to insecticides than to herbicides or fungicides.

4.2.3.2 Proposal of a decision tree to be used for sediment risk assessments

Based on the literature information mentioned above (section 2.3.1) a decision tree consisting of criteria which should be established when pesticides will require sediment toxicity testing, is performed. According to the EU- 8075/VI/97 (draft) working document on the guidance document on aquatic ecotoxicology, a similar approach on sediment testing requirements can be adopted, i.e., the first criterion which should be fulfilled is that pesticides (parent compounds or/and the respective metabolites) that reach the sediment through spray drift or through surface run-off, are present in the sediment in ≥ 10 % of applied radioactivity after 14 days. If this is not the case, no risk for sediment dwelling organisms is expected. If this is the case the following question is the time of exposure. If the exposure is continuous or frequent and repetitive, chronic sediment tests are required. If this is not the case, the

following questions concerning adsorption and persistency as proposed by Maund et al. (1997) should be asked i.e., $logK_{oc} \geq 3$ and $DT_{50} \geq 90$ days. If this is the case, chronic sediment tests are required. If this is not the case, acute sediment tests only, are required. To assess the risks of pesticides in sediments, the quotient PEC/L(E)C_{50} or PEC /NOEC should be calculated; the acute and chronic tests should be compared with the $PEC_{sed} \geq 10$ days and the $PEC_{sed} \geq 28$ days, respectively which can be calculated with USES 3.0 (1999). The assessment factors for the acute and chronic exposure are established in the same way as for aquatic organisms, i.e. 100 and 10, respectively. If the risk quotients exceed the regulatory level of concern there will be a risk for sediment-dwelling organisms; if they do not exceed, no risk will be present (see scheme in ANNEX A).
The toxicity trigger of *Daphnia* 48-h acute toxicity value < 1 mg/L or 21-day NOEC < 0.1 mg/L, as proposed by the EU- 8075/VI/97 (draft), the CTB and by Maund et al. (1997) is not included in this proposal. The reasons for this are explained here below. The assumption that pelagic organisms have similar sensitivity as benthic organisms is considered not correct. As mentioned also in section 2.2.2 there is evidence that for compounds with $logK_{ow}$ greater than 5 exposure can be underestimated. Different factors such as exposure route (sediment vs. water), organisms behaviour (e.g. sediment avoidance) and variation in bioavailability of the compounds due to different organic carbon quantity/quality and particle grain size of sediments, might account for the differences in sensitivity. Finally, toxicity data on several pesticides have also demonstrated that differences in sensitivity between aquatic and benthic organisms can occur also for compounds with $logK_{oc} < 5$. For the insecticide thiacloprid ($logK_{ow} = 1.26$) for example, differences in sensitivity between water organisms and benthic organisms were a factor 500 to 3000 for chronic and acute exposure, respectively (Smit et al., 2001).

4.3 Laboratory spiked-sediment tests

4.3.1 Sediment manipulation, characterisation and spiking procedures

Spiked sediments are usually collected from relatively clean locations and are spiked in the laboratory with a chemical in known concentrations. Before spiking, sediments should be handled, stored and characterised following accurate procedures which are recommended and described in different guidelines (Hill et al, 1993; ASTM, 1999b; USEPA, 2000). Several studies on storage time and storage temperature have been performed to investigate the effects of storage on physical-chemical characteristics alterations of the sediment and on the consequent bioavailability and toxicity of the compound of concern (Moore et al., 1996; Sijm et al., 1997; DeFoe and Ankley, 1994). Based on the conclusions and recommendations of these studies and, depending on the type and class of compound, long periods of storage (for examplemore than 8 weeks), should be avoided. Also homogenisation of sediment samples (by mixing or/and sieving) should occur with care to avoid alterations of sediment characteristics (ASTM, 1999b). Prior to spike, sediments should be characterised and parameters such as dry weight, organic and inorganic carbon content, grain size, pH, presence of ammonia and sulphides in pore water should be measured (ASTM, 1999a; USEPA, 2000). Standard methods for sediment spiking do not exist. Various methods for introducing the test chemical to the sediment, however, are available depending mainly on the type of chemical added and the test objectives (Ditsworth et al., 1990; Hill et al, 1993; USEPA, 2000; ASTM, 1999b). Spiking methods available are: rolling mill, food mixer, hand mixing; with or without the use of an organic solvent or directly wet mixing to the sediment or through application to the walls of the mixing vessels etc. Prior to spike the following information should be known on the test compound: solubility in water and in organic solvents, $logK_{ow}$, persistence,

volatility, biological and chemical degradation rate, photolysis. There are a number of items for which particularly care should be addressed during the spiking procedure used as these can influence the partitioning and the results on toxicity:

a. volume of organic solvent used as carrier (should be minimised);
b. time and temperature of mixing;
c. equilibration time;
d. storing conditions before testing.

It is advised to normalise nonpolar organic compound concentrations to dry weight (mg/kg dry wt.) and organic carbon content (mg/kg oc) (ASTM, 1993, 1999b; USEPA, 2000).
Although standard procedures for spiking are lacking, general information on the methods available are described in Hill et al. (1993) and in the USEPA (2000) guideline. Additional useful information on the above mentioned items can be find in Landrum et al, 1992; Landrum and Faust, 1992; DeFoe and Ankley, 1994).
Finally, recently new results have been published concerning the relationship between sediment manipulation and organic carbon content. Sediments can be easily manipulated such that, on a short time scale of 1 day, sorption may be increased by one or two orders of magnitude, without affecting the organic carbon content (Van Noort, RIZA, personal communication). Consequently, effect concentrations for less sorptive sediments may be underestimated by orders of magnitude. For the worst case situation, the magnitude of such a "hidden" increased sorptive power can, however, in approximation, be easily detected by a Tenax extraction of the loaded sediment for 6 h (Cornelissen et al., 2001). It is, therefore, recommended to perform Tenax extractions on loaded sediments in addition to total content determinations by solvent extraction. If the total content exceeds the Tenax-extractable content by more than a factor of 5, two times the Tenax-extractable content should be taken to approximate the effective concentration on the sorption properties of organic compounds.

4.3.2 Test types and recommended test species

A range of freshwater test species and methods are commonly used to test the toxicity of whole sediments. On the basis of the following main criteria: ecological relevance, contact with sediment, short generation time, sediment tolerance of a wide range of particle size, sensitivity in toxicity assessments, the most recommended species for toxicity tests are the amphipod *Hyalella azteca* (Arthropoda Crustacea), the midge larvae *Chironomus riparius* or *tentans* (Arthropoda Insecta) and among the worms, *Lumbriculus variegatus* (Anellida Oligochaeta). These standard test organisms can be used to assess acute (10-d) and chronic (28-d) toxicity and to determine survival, growth and reproduction or emergence (in the case of the midge larvae) and are recommended test species by Hill et al. (1993), ASTM (1999a), Ingersoll et al. (1995) and USEPA (2000). These organisms have also the advantage to be easy to handle and to culture in the laboratory. Although test methods described for *L. variegatus* are as far advanced as those described for *H. azteca* and for *Chironomus sp.*, methods for *L. variegatus* have not been validated yet through interlaboratory studies (USEPA, 2000). A short description of test methods with the three species is described here below.

4.3.2.1 Hyalella azteca: survival and growth test

4.3.2.1.1 Introduction

H. azteca has been successfully used in several studies to evaluate the toxicity of freshwater sediments (Ingersoll and Nelson, 1990; Hoke et al., 1995; Burton et al., 1996; Kemble et al., 1998). *H. azteca* is an epibenthic detritivore that burrows in the sediment surface and inhabits mostly warm (20 to 30°C) permanent lakes, ponds and streams throughout the entire

American continent. *H. azteca* has a wide tolerance of sediment grain size (sediment particles ranging from > 90% silt and clay-size to 100% sand-size particles). The life cycle of this amphipod can be divided into three stages: (1) an immature stage consisting of 5 instars; (2) a juvenile stage including instars 6 and 7; (3) an adult stage, the 8th instar and older. Culturing methods are described in Ingersoll and Nelson (1990), Ankley et al. (1994a) and USEPA (1994a).

4.3.2.1.2 Summary of test methods

The ASTM (1999a) guideline recommends either a static or a flow-through system using 1-L beakers containing about 200 mL (2 cm) of sediment and 800 mL of overlying water. A minimum of 4 replicates with 20 amphipods per replicate are used and the temperature of exposure can be ranged from 20 to 25°C. A photoperiod 16 to 8-h ratio of light to darkness and a feeding regime based on rabbit pellets 2 to 3 times a week is proposed. The test duration can range from a < 10 day short-term test to a long-term test > 10 days continuing up to 30 days. Endpoints suggested are survival and growth (through body-length or wet and dry weight measurements) after 10 days, reproduction (number of eggs or/and juveniles per female) in long-term tests lasting 30 days (Annex B).
The USEPA (2000) guideline recommends a 10-d sediment toxicity test at 23°C using test chambers of 300 mL beakers containing 100 mL of sediment and 175 mL of overlying water. Ten 7-to 14-d-old juveniles are used in each chamber. The number of replicates recommended are 8 for routine testing but might be more or less depending on the objectives of the test. Overlying water should be renewed during the exposure period. The organisms are daily fed with YCT food (see USEPA 1994a for details) and the photoperiod proposed is 16 to 8-h ratio of light to darkness. Test endpoints determined after 10-d of exposure are survival and dry weight (60 to 90°C) or body length as a measure for growth. 10-d LC_{50} values for survival and NOEC values for growth are expressed in mg/kg dry weight sediment. The USEPA (2000) guideline proposes also a chronic test lasting 42-days and is based on comparable test methods as described for the acute test (Annex B).

4.3.2.2 Chironomus riparius/tentans: survival and growth test

4.3.2.2.1 Introduction

Chironomid species (*C. tentans* and *C. riparius*) midge larvae have been successfully used in evaluating the toxicity of freshwater sediments (Ankley et al., 1994; Hoke et al., 1995; Burton et al., 1996). Both species are important in the diet of young and adult fish and surface feeding ducks. Both species are in direct contact with the sediment by burrowing into the sediment to build a case.
C. tentans has a holartic distribution and is locally common in the midcontinental areas of North America and throughout the Mediterranean, alpine and Balkan regions up to the whole western Europe (Janssens de Bisthoven, 1995). The larval stages of *C. tentans* often inhabit eutrophic lakes and ponds and occur mostly in fine sediment (sediment particles ranging from 0.15 to 2 mm) and penetrate in the upper 10 cm of substrate.
C. riparius has a world wide distribution; the midge larvae inhabit eutrophic lakes and ponds and occur mostly in mud-bottom littoral habitats characterised by silty and sandy sediments to depths up to 1 m. The life cycle of both species can be divided into three stages: (1) larval stage consisting of 4 instars; (2) a pupal stage and (3) an adult stage. Midge egg masses hatch in 2 or 3 days after deposition in water at 19 to 22°C. Larval growth occurs in 4 instars of about 1 week each and under optimal conditions larvae will pupate and emerge as adults at 20°C after 15 to 21 days and 24 to 28 days for *C. riparius* and *C. tentans*, respectively. Both

species are easily reared in the laboratory. Culturing procedures for *C. tentans* and *C. riparius* can be found in the ASTM (1999a) and in USEPA (2000) guidelines.

Chironomus riparius/tentans: survival and growth test Summary of test methods

4.3.2.2.2 Summary of test methods

The ASTM (1999a) describes different methods of exposure for second-instar larvae of *C. tentans*: either static or flow-through system. Test systems consist of either 3-L aquaria containing 100g of sediment and 2L of overlying water using 25 larvae per test chamber (Adams et al., 1985) or 20L aquaria containing 2 to 3 cm sediment with 15 cm of overlying water in a static system using 100 larvae (second instar) per test chamber (Nebeker et al., 1984). In both test systems exposures occurred at a temperature ranging from 20 to 23°C and the larvae were fed during the test. The test duration suggested by the authors can range from less than 10 days up to 25 days and, survival, growth or adult emergence can be monitored as biological endpoints. Larval survival, growth or adult emergence can be monitored as biological endpoints. 10-d LC_{50} values for survival and NOEC values for growth are expressed in mg/kg dry weight sediment (Annex C).

The USEPA (1999) recommends a 10-day survival and growth and a life cycle sediment toxicity test (50-65 days) with *C. tentans*. Both tests are carried out at 23℃ (± 1) using test chambers or 300 mL beakers containing 100 mL of sediment and 175 mL of overlying water with 10 to 12 second to third-instar midges (about 10-days old) in each. The number of replicates ranges from 8 to 16 in the short and long-term test, respectively. Overlying water should be renewed during the exposure period. The larvae should be fed daily with Tetrafin goldfish. Test endpoints determined after 10-d of exposure are survival and ash-free dry weight (AFDW) as a measure for growth. The endpoints for the life-cycle test are: survival and growth after 20-d and, survival, emergence and reproduction at the end of the test. 10-d and 50-65-d LC_{50} values for survival and NOEC values for growth and reproduction are expressed in mg/kg dry weight sediment (Annex C).

Two OECD guidelines are available for conducting sediment-water tests with chironomids using either spiked sediment (218) or spiked water (219). Both guidelines are designed to assess the effects of prolonged or chronic exposure of pesticides and industrial chemicals to the sediment-dwelling larvae of *Chironomus sp.* Test systems described in both guidelines (i.e. number of concentrations, test vessels and organisms, temperature, feeding regimes and test duration and test endpoints) are practically comparable. However, for the reasons explained above in section 2.2.1 only the sediment spiked method is recommended. Larval midges are exposed in static sediment-water systems at a temperature of 20-23℃ (± 2) in 600-mL beakers containing a depth of 1.5 cm of sediment and 6 cm of overlying water. The number of organisms per test beaker is 5 and the number of replicates ranges from 3 to 4. Renewal of overlying water is not necessary unless the water quality criteria are not met or the test substance is not stable. Continuous aeration of overlying water is provided through a glass Pasteur pipette fixed 2-3 cm above the sediment layer. The larvae are fed preferably daily or at least three times per week with Tetra-Min or Tetra-Phyll.

The two guidelines support the use of formulated sediments (i.e. artificial or synthetic sediment which consists of a mixture of materials used to mimic the physical components of a natural sediment). The use of formulated sediments has the advantage to eliminate interference caused by the possible presence of indigenous organisms and the variation in physicochemical characteristics of the sediments which can make standardisation difficult (Naylor and Rodrigues, 1995). The methods described are suitable for 3 species of chironomids: *C. riparius*, *tentans* and *yoshimatsui* but the use of *C. riparius* is preferable because of its shorter exposure time (20-28 days) compared to *C. tentans* (28-65 days) and is

easier to handle in the laboratory. In both guidelines the endpoints suggested are survival, emergence and growth and the concentration that would cause x % reduction in emergence or larval survival or growth can be calculated and expressed as EC_{15} or EC_{50} etc.). In addition, and as far as possible, lethal and sublethal effects values can be compared with control values to determine the lowest observed effect concentration (LOEC) or/and the no observed effect concentration (NOEC) (Annex C).

4.3.2.3 Lumbriculus variegatus: survival, dry weight and reproduction test

4.3.2.3.1 Introduction

Many investigators have successfully used the oligochaete, *L. variegatus* to evaluate the toxicity of freshwater sediments and to examine bioaccumulation of chemicals from sediments (Nebeker et al., 1989; Ankley et al., 1992; Phipps et al., 1993; Kukkonen and Landrum, 1994). *L. variegatus* is a sediment-dwelling organisms that burrows into the sediment with the anterior part of its body while the posterior portion undulates in the overlying water to achieve respiratory exchanges. The worms feed on organic material of the aerobic zone and do not penetrate in the anaerobic zone. They inhabit a variety of sediment types of lakes, rivers, ponds throughout the United States and Europe (Chekanovskaya, 1962). This species reproduces mostly asexually through architomy in which new individuals are produced by the anterior end of the parent and subsequently replaced with 8 new segments. *L. variegatus* is easy to culture and maintain in the laboratory. Culturing procedures are described in Phipps et al. (1993) and USEPA (1994a).

4.3.2.3.2 Summary of test methods

A recommended sediment toxicity test (10 to 28 d) with *Lumbriculus variegatus* is described in Phipps et al. (1993). Test chambers used are 300 mL glass beakers containing 100 mL of sediment and 100 to 150 mL of overlying water. The testchambers are placed in tanks where clean water is allowed to flow slowly into the beakers. Ten worms are placed in each beaker and tests are conducted at 25°C ± 2°C. The number of replicates is not specified as it depends on the objectives of the test. The worms are fed every 3 days with salmon starter per 100 mL of sediment. At the end of the test, worms are removed by sieving (5 mm sieve), counted, dried into the oven at 100°C for 4 to 24 h. Average individual weight, survival and, in exposures lasting > 10 days, reproduction. This latter endpoint consists in recording total number of organisms as, given the mode of asexual reproduction, it is impossible to distinguish between young and adult organisms (Annex D).

4.4 Conclusions

This factsheet has the aim to develop a decision tree based on suitable trigger values in identifying pesticides for which sediment toxicity testing may be necessary. The second aim is to propose short and long-term standard sediment test methods that can measure toxic effects of pesticides to sediment-dwelling organisms to be used in the framework of Pesticide registration.

Criteria that trigger sediment toxicity testing are (a): parent compounds and metabolites that reach the sediment phase via drift or run-off and persist such that >10% of the total applied radioactivity is measured in the sediment after 14 days; (b) the period of exposure (prolonged or/and repetitive); (c) logKoc ≥ 3 and sediment DT_{50} ≥ 90 days.

Sediment chronic tests (≥ 28 days) are recommended when criteria (a) and (b) and (a), (b) and (c) occur together. Sediment acute tests are recommended when criteria (a) occurs. The PEC_{sed} ≥ 10 days and the PEC_{sed} ≥ 28 days for the risk characterisation are calculated

according to USES 3.0. Risk quotients PEC_{acute}/L(E)C50 and $PEC_{chronic}$/NOEC that meet the standards 0.01 and 0.1, respectively, indicate acceptable risk. Proposed laboratory sediment tests for measuring acute and chronic toxic effects are: the test with (a) the midge larvae *Chironomus riparius/tentans*, (b) the amphipod *Hyalella azteca* (c), the oligochaete, *Lumbriculus variegatus.* As endpoints, survival, growth and emergence are measured after 10-days of exposure. In chronic exposures reproduction is measured as an additional endpoint after 28 days.

References

Adams W.J., Kimerle R.A., & Mosher R.G. (1985). An approach for assessing the environmental safety of chemicals sorbed to sediments. In Aquatic Toxicology and Hazard Evaluation:7th Symposium, eds. Cardwell R.D., Purdy R., Bahner R.C., 429-453. ASTM STP 854, Philadelphia, PA.

Ankley G.T., Cook P.M., Carlson A.R., Call D.J., Swenson J.A., Corcoran, H.F., Hoke R.A. (1992). Bioaccumulation of PCBs from sediments by oligochaetes and fishes: comparison of laboratory and field studies. Can.J.Fish.Aquat.Sci.49:2080-2085.

Ankley G.T., Call D.J., Cox J.S., Kahl M.D., Hoke R.A, Kosian P.A. (1994). Organic carbon partitioning as a basis for predicting the toxicity of chlorpyrifos in sediments. Environ.Toxicol.Chem.13 (4): 621-626.

ASTM (1999a). Standard test methods for measuring the toxicity of sediment-associated-contaminants with freshwater invertebrates. E 1706-95b. In Annual Book of ASTM Standards. Vol. 11.05, Philadelphia, PA.

ASTM (1999b). Standard guide for collection, storage, characterization, and manipulation of sediments for toxicological testing. E-1391-94. In Annual Book of ASTM Standards. Vol. 11.05, Philadelphia, PA.

BBA (1995). Long-term toxicity test with *Chironomus riparius*: Development and validation of a new test system. Edited by M. Streloke and H.Köpp. Berlin 1995.

Belfroid A., Seinen W., van de Berg M., Hermens J. & van Gestel K. (1995). Uptake, bioavailability and elimination of hydrophobic compounds in earthworms (*Eisenia andrei*) in field contaminated soil. Environ.Toxicol.Chem. 14:605-612.

Belfroid A.C., Sijm D.T.H.M., and van Gestel C. A.M. (1996). Bioavailability and toxicokinetics of hydrophobic aromatic compounds in benthic and terrestrial invertebrates. Environ Rev 4: 276-299.

Burton G.A., Jr. (1991). Assessing freshwater sediment toxicity. Environ.Toxicol.Chem. 10: 1585-1627.

Burton G.A., Nelson M.K., & Ingersoll C.G. (1992). Freshwater benthic toxicity tests. In: Sediment Toxicity Assessment, ed. Burton G.A., 213-240. Lewis Publishers, Chelsea, Mi.

Burton G.A. A., Norberg-King T.J., Ingersoll C.G., Ankley G.T., Winger P.V., Kubitz J., Lazorchak J.M., Smith M.E., Greer I.E., Dwyer F.J., Call D.J., Day K.E., Kennedy P., & Stinson M. (1996). Interlaboratory study of precision: *Hyalella azteca* and *Chironomus tentans* freshwater sediment toxicity assays. Environ.Toxicol.Chem.15(8):1335-1343.

Chekanovskaya, O.V. (1962). Aquatic oligochaeta from the USSR.Akademiya Nauk SSR. Moskow, USSR.

Ciarelli S., van Straalen N.M., Klap V.A. & Van Wezel A.P. (1999). Effects of sediment bioturbation by the estuarine amphipod *Corophium volutator* on fluoranthene resuspension and transfer into the mussel, *Mytilus edulis* L. Environ.Toxicol.Chem. 18(2): 318-328

Ciarelli S., Kater B.J. & van Straalen N.M. (2000). The influence of bioturbation by the amphipod *Corophium volutator* on fluoranthene uptake in the marine polychaete, *Nereis virens*. Environ.Toxicol.Chem. 19(6): 1575-1581.

Cornelissen G., van Noort P. & Govers H .(1997). Desorption kinetics of chlorobenzenens, polycylic aromatic hydrocarbons, and polychlorinated biphenyls: sediment extraction with tenax and effects of contact time and solute hydrophobicity. Environ. Toxicol. Chem. 16: 1351-1357.

Cornelissen G., ten Hulscher Th.E.M., Rigterink H., Vrind B.A., & van Noort P.C.M. (2001). A simple Tenax method to determine the chemical availability of sediment-sorbed organic compounds. Environ. Toxicol. Chem. 20: 706-711.

Crommentuijn-T., Sijm-D., de Bruijn J., van Leeuwen K., & van de Plassche E. (2000). Maximum permissible and negligible concentrations for some organic substances and pesticides. Journal-of-Environmental-Management: 58 (4) 297-312.

CTB (2001). Aanvraagformulier B – Instructieonderdeel H.2.3.Toxiciteitsonderzoek sedimentorganismen-Onderzoeksvraag H.2.3a. Board for the authorisation of pesticides (CTB - Wageningen).

DeFoe D.L. & Ankley G.T.(1994). Effects of storage time on the toxicity of sediments to freshwater benthic invertebrates. Abstract book, 15th annual meeting, SETAC, Denver, CO.

DiToro D.M., Zarba C.S., Hansen D.J., Berry W.J., Swartz R.C., Cowan C.E., Pavlou S.P., Allen H.E., Thomas N.A.& Paquin P.R. 1991. Technical basis for examining sediment quality criteria for non-ionic organic chemicals using equilibrium partitioning. Environ.Toxicol.Chem.13:1163-1175.

Ditsworth G.R. and Schults W. (1990). Preparation of benthic substrates for sediment toxicity testing. Environ.Toxicol.Chem.9:1523-1529.

Driscoll S.K., Landrum P.F., Tigue E., (1997). Accumulation and toxicokinetics of fluoranthene in water-only exposures with freshwater amphipods. Environ.Toxicol.Chem.16:754-761.

ECOFRAM (Ecological Committee on Federal, Insecticide, fungicide and Rodenticide Act, Risk Assessment) Aquatic Draft Report (1999).

Environment Canada (1997). Test for Growth and Survival in Sediment using Larvae of Freshwater Midges (*Chironomus tentans* or *Chironomus riparius*). Biological Test Method. Report SPE 1/RM/32. December 1997.

EPA (1998). Guidelines for ecological risk assessment. EPA 630/R-95/002F.

Harkey G.A., Landrum P.F., & Klaine S.J.(1994). Preliminary study on the effect of feeding during whole sediment bioassays using *Chironomus riparius* larvae. Chemosphere 28: 597-606.

Hill I.R., Matthiessen P., & Heimbach F. eds.(1993). Guidance document on sediment toxicity tests and bioassays for freshwater and marine environments. From: WOSTA Workshop On Sediment Toxicity Assessment. Society of Environmetal Toxicology and Chemistry- Europe, Brussels, Belgium.

Hines M.E., Banta G.T., Giblin A.E., Hobbie J.E. & Tugel J.B. (1994). Acetate concentrations and oxidation in salt-marhs sediments. Limnol. Oceanogr. 39(1):140-148.

Hoke R.A ., Kosian P.A., Ankley G.A., Cotter A.M., Vandenneiden F.M., Phipps G.L., Durhan E.J. (1995). Check studies with *Hyalella azteca* and *Chironomus tentans* in support of the development of a sediment quality criterion for dieldrin. Environ.Toxicol. Chem.14(3):435-443.

Ingersoll C.G. and Nelson M.K.(1990). Testing sediment toxicity with H*yalella azteca* (Amphipoda) and *Chironomus riparius* (Diptera). In: Aquatic Toxicology and Risk Assessment, 13th volume, eds. Landis and van der Schalie, 93-109.ASTM STP 1096, Philadelphia, PA.

Ingersoll C.G., Ankley G.T., Benoit D.A., Burton G.A., Dwyer F.J., Greer I.E., Norberg-King T.J., & Winger P.V. (1995). Toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates: a review of methods and applications. Environ.Toxicol.Chem.14:1885-1894.

ISO/DIS 13641-1 (draft guideline 2001) Water quality- Determination of inhibition of activity of anaerobic bacteria –Part 1: Inhibition of anaerobic digestion.

Janssens de Bisthoven, L. (1995). Morphological deformities in *Chironomus* gr. *Thummi* (Diptera, chironomidae) as bioindicators for micropollutants in sediments of Belgian lowland rivers. Ph. D. thesis, University of Leuven, Belgium.

Kemble N.E., Brunson E.L., Canfield T.J., Dwyer F.J., & Ingersoll C.G. (1998). Assessing sediment toxicity from navigational pools of the Mississipi River using a 28-d *Hyalella azteca* test. Arch.Environ.Contam.Toxicol. 35: 181-190.

Kraaij R., Ciarelli S., Tolls J., Kater B. & Belfroid A. (2001). Bioavailability of lab-contaminated and native PAHs to the amphipod *Corophium volutator* relates to chemical desorption. Environ.Toxicol. Chem. 20:1716-1724.

Kukkonen J. & Landrum P.F.(1994). Toxicokinetics and toxicity of sediment-associated pyrene to *Lumbriculus variegatus* (Oligochaeta). Environ.Toxicol.Chem.13: 1457-1468.

Kukkonen J. & Landrum P.F.(1995). Effects of sediment-bound polydimethylsiloxane on the bioavailability and distribution of benzo(a)pyrene in lake sediment to *Lumbriculus variegatus*. Environ.Toxicol.Chem.14:523-531

Landrum P., Eadie B., & Faust W. (1992). Variation in the bioavailability of polycyclic aromatic hydrcarbons to the amphipod *Diporeia* spp. with sediment aging. Environ.Toxicol.Chem.11:1197-1208.

Landrum P. & Faust W. (1992). Bioavailability and toxicokinetics of polycyclic aromatic hydrcarbons sorbed to sediments for the amphipod *Pontoporeia hoyi*. Environ.Sci.Technol.23:588-595.

Maund S., Barber I., Dulka J., Gonzalez-Valero J., Hamer M., Heimbach F., Marshall M., McCahon P., Staudenmaier H. & Wustner D. (1997). Development and evaluation of triggers for sediment toxicity testing of pesticides with benthic macroinvertebrates. Environ.Toxicol. Chem. 16 (12):2590-2596.

Meador J., Stein J., Reichert W., & Varanasi U. (1995). Bioaccumulation of polycyclic aromatic hydrocarbons by marine organisms. Rev. Environ. Contam. Toxicol. 143:80-164.

Moore D.W., Dillon T.M., & Gamble E.W.(1996). Long-term storage of sediments: implications for sediment toxicity testing. Environ. Pollut. 89:341-342.

Naylor C. and C. Rodrigues. (1995). Development of a test method for *Chironomus riparius* using a formulated sediment. Chemosphere 31:3291-3303.

Nebeker A.V., Cairns M.A, Gakstatter J.H., Malueg K.W., Schuytema G.S. & Krawczyk D.F. (1984). Biological methods for determining toxicity of contaminated freshwater sediments to invertebrates. Environ.Toxicol. Chem. 3:617-630.

Nebeker A.V., Cairns M.A, & Wise C.M., (1984). Relative sensitivity of *Chironomus tenstans* life stages to copper. Environ.Toxicol. Chem. 3: 151-158.

OECD (1992). Report of the OECD Workshop on Effects Assessment of Chemicals in Sediment. OECD Environment Monograph No. 60. Organisation for Economic Cooperation and Development, Paris.

Peijnenburg W., Jager D.T., Posthuma L., Sijm D.T.H.M., 2001. Huidige mogelijkheden en inzichten voor implementatie van metaalbiobeschikbaarheid in de risicobeoordeling van landbodems. RIVM report no. 607220004.

Phipps G.L., Ankley G.T., Benoit D.A., & Mattson V.R. (1993). Use of the aquatic oligochaete *Lumbriculus variegatus* for assessing the toxicity and bioaccumulation of sediment-associated contaminants. Environ.Toxicol. Chem.12:269-279.

RIVM (1999). Uniform System for the Evaluation of Substances (USES), version 3.0. National Institute of Public Health and the Environment (RIVM), Ministry of Housing, Spatial Planning and the Environment (VROM), and Ministery of Welfare, Health and Sports (WVS). RIVM Report 679102044.

Sijm R.T.H., Haller M.& Schrap S.M. (1997). Influence of storage on sediment characteristics and of drying sediment on sorption coefficients of organic contaminants. Bull. Environ. Contam. Toxicol:58:961-968.

Sijm, D., Lijzen J., Peijnenburg W., Traas T. & Verbruggen E. (2002): "Biobeschikbaarheid in beleid … wat er aan voorafging … wat nog komt". Workshop on bioavailability held in Voorburg (The Netherlands) on 13 september 2001. Draft document.

Smit C.E., Rila J.P. & Posthuma C. (2001). RIVM/CSR Adviesrapport 08301a00. Bilthoven, The Netherlands.

USEPA (1994). Methods for measuring the toxicity and bioaccumulation of sediment associated contaminants with freshwater invertebrates. First edition. EPA 600/4-91/002, Cincinnati, OH.

USEPA (2000). Methods for measuring the toxicity and bioaccumulation of sediment associated contaminants with freshwater invertebrates. Second edition. EPA 600/R-99/064. Draft Revision to the first edition dated June 1994**.**

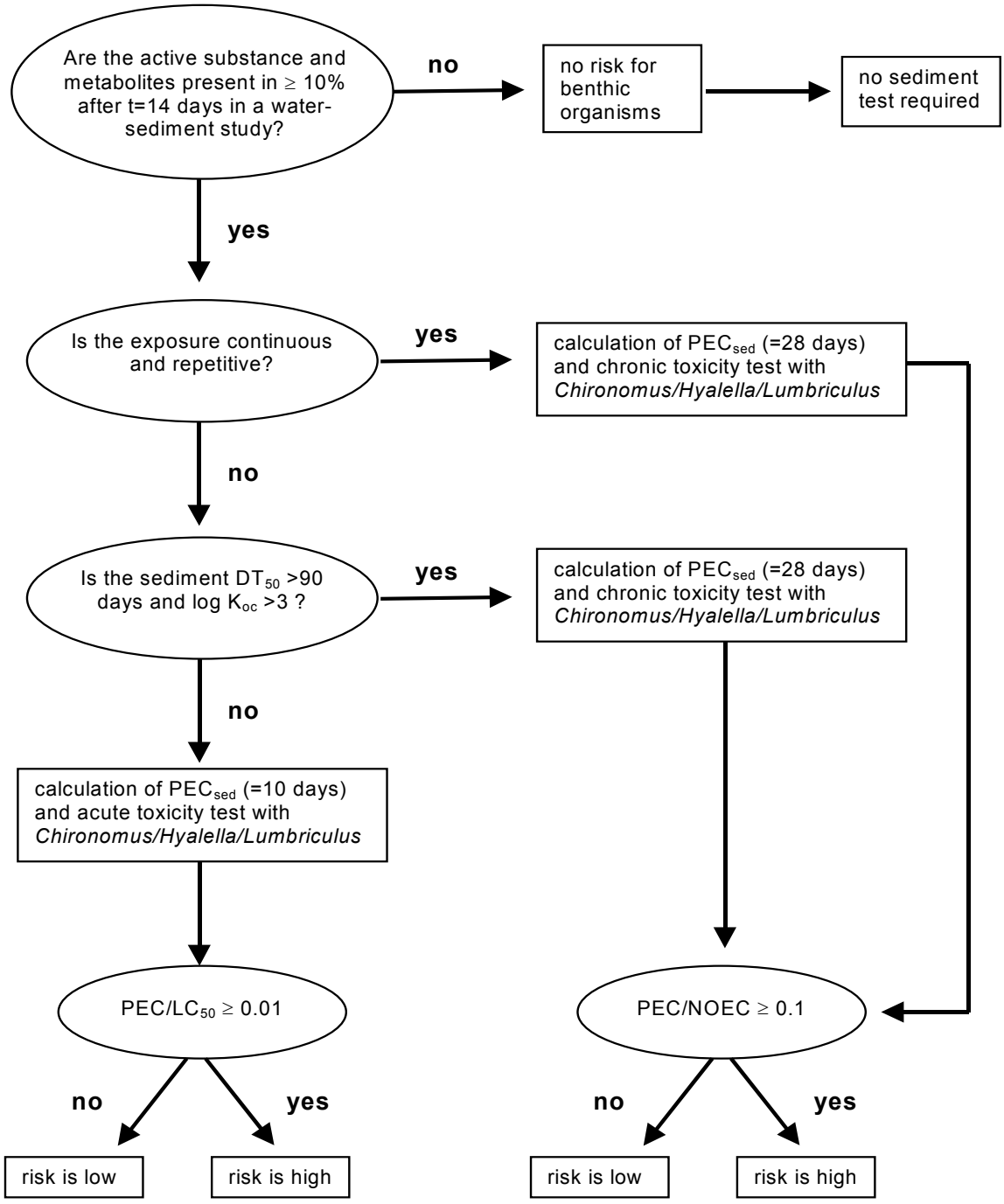
US-EPA/OPPTS 850.1735. (1996): Whole Sediment Acute Toxicity Invertebrates.

RIVM, VROM, VWS (1999). Uniform System for the Evaluation of Substances 3.0 (USES 3.0). National Institute of Public Health and the Environment (RIVM), Ministry of Housing, Spatial Planning and the Environment (VROM), Ministry of Health, Welfare and Sport (VWS), The Netherlands. RIVM Report 601450004.

Van Beelen P. & Van Vlaardingen P.L.A. (1994). A method for the ecotoxicological risk analysis of polluted sediments by the measurement of microbial activities In: Ecotoxicology of Soil Organisms, eds. Donker M., Eijsackers H., Heimbach F, p.105-112 (Chap. 6. SETAC Special Publications Series.

Van Vlaardingen P.L.A. & Van Beelen P. (1992). Toxic effects of pollutants on methane production in sediments of the river Rhine. Bull.Environ.Contam.Toxicol. 49:789-786.

Annex A. Risk for sediment dwelling organisms



Annex B. Recommended test conditions for *H. Azteca*

Parameters	10-d test	28-d test/42-d test*	References
Test type	- flow through - renewal of overlying water	- flow through - renewal of overlying water	ASTM, 1999 USEPA, 2000
Temperature	- 20 ± 1 - 23 ± 1	- 20 – 25 - 23 ± 1	ASTM, 1999 USEPA, 2000
Photoperiod	- 16L:8D - 16L:8D	- 16L:8D - 16L:8D	ASTM, 1999 USEPA, 2000
Test chamber	- 1-L beakerglass - 300-mL high form beakerglass	- 1-L beakerglass - 300-mL high form beakerglass	ASTM, 1999 USEPA, 2000
Sediment-water ratio	- 1:4 - - 1:1.75	- 1:4 - 1:1.75	ASTM, 1999 USEPA, 2000
Renewal of overlying water	- continuously - continuously: 2 volumes addition/d (or intermittent: 1 volume/12hr)	- continuously - continuously: 2 volumes addition/d (or intermittent: 1 volume/12hr)	ASTM, 1999 USEPA, 2000
Age of organisms	- juveniles (2 – 3mm length) - juveniles (7-14- old)	- juveniles (2 – 3mm length) - juveniles (7-8-d old)	ASTM, 1999 USEPA, 2000
Number of organisms/chamber	- 20 - 10	- 20 - 10	ASTM, 1999 USEPA, 2000
Number of replicates/treatment	- 4 - 8	- 4 - 12 (4 for 28-d survival and growth, 8 for 35- and 42-d survival, growth, reproduction)	ASTM, 1999 USEPA, 2000
Feeding	- 6 – 14 mg rabbit pellets (3 x week) - YCT (Yeast,Cerophyll, Trout) – 1 mL daily	- 6 – 14 mg rabbit pellets (3 x week) - YCT (Yeast,Cerophyll, Trout) – 1 mL daily	ASTM, 1999 USEPA, 2000
Aeration	- Yes - None	- yes - none	ASTM, 1999 USEPA, 2000
Overlying water parameters	- pH, dissolved oxygen, hardness (beginning and end) - pH, dissolved oxygen, hardness, ammonia (daily)	- not specified - hardness, ammonia (beginning and end); T, daily, pH, dissolved oxygen 3 x week	ASTM, 1999 USEPA, 2000
Endpoints	- survival, growth - survival, growth	- survival, growth, reproduction - survival, growth, reproduction	ASTM, 1999 USEPA, 2000

* 28-d and 42-d tests are described in ASTM (1999a) and USEPA (2000), respectively.

Annex C. Recommended test conditions for *C. tentans/C. riparius*

Parameters	10-d test*	28-d/ 28-65-d test**	References
Test type	- static or flow-through - static with renewal of overlying water	- static or flow-through - static with renewal of overlying water - static	ASTM, 1999 USEPA, 2000 OECD, 2000
Temperature	- 20 – 23 °C - 23 ± 1°C	- 20 – 23 °C - 23 ± 1°C - 20 – 23 ° ± 2°C	ASTM, 1999 USEPA, 2000 OECD, 2000
Photoperiod	- 16L:8D - 16L: 8D	- 16L:8D - 16L: 8D - not specified	ASTM, 1999 USEPA, 2000 OECD, 2000
Test chamber	- 3 – 20-L aquaria - 300-mL high-form beakers	- 3 – 20-L aquaria - 300-mL high-form beakers - 600- mL high-form beakers	ASTM, 1999 USEPA, 2000 OECD, 2000
Sediment-water ratio	- 1: 5 (±) - 1:1.75	- ± 1-20/1-5 - 1:1.75 - 1:4	ASTM, 1999 USEPA, 2000 OECD, 2000
Renewal of overlying water	- none or continuously - continuously: 2 volumes addition/d (or intermittent: 1 volume/12hr)	- none or continuously - continuously: 2 volumes addition/d (or intermittent: 1 volume/12hr) - preferably not	ASTM, 1999 USEPA, 2000 OECD, 2000
Age of organisms	- 2nd- instar larvae - 2nd-3th instar larvae (± 10-d old larvae)	- 2nd- instar larvae - < 24 h-old larvae - 1st instar larvae	ASTM, 1999 USEPA, 2000 OECD, 2000
Number of organisms/chamber	- 25- 100 - 10	- 25- 100 - 12 - 5	ASTM, 1999 USEPA, 2000 OECD, 2000
Number of replicates/treatment	- not specified - 8 for routine testing	- not specified - 16 - 3-4	ASTM, 1999 USEPA, 2000 OECD, 2000
Feeding	- ground cereal leaves + fish food flakes - Tetrafin goldfish food (1.5 mL daily to each test chamber)	- ground cereal leaves + fish food flakes - Tetrafin goldfish food (1.5 mL daily to each test chamber) - Tetra Min, Tetra Phyll fish food (5 ml per vessel/day)	ASTM, 1999 USEPA, 2000 OECD, 2000
Aeration	- Yes - none, unless oxygen drops below 2.5 mg/L	- yes - none, unless oxygen drops below 2.5 mg/L - yes, gentle	ASTM, 1999 USEPA, 2000 OECD, 2000
Overlying water parameters	- not specified - temperature and dissolved oxygen daily; hardness, pH, conductivity and ammonia at the beginning and end of test	- not specified - temperature daily; dissolved oxygen and pH 3x/week; hardness, conductivity and ammonia at the beginning, on day 20 and end of test - dissolved oxygen daily, pH, temperature, hardness and ammonia at start and end	ASTM, 1999 USEPA, 2000 OECD, 2000
Endpoints	- survival, growth, adult emergence - survival, growth (AFDW)	- survival, growth, adult emergence,reproduction	ASTM, 1999, USEPA, 2000, OECD, 2000

*Can last longer than 10-days and up to 25 and 30-days for *C. tentans* and *C. riparius*, respectively (according to the ASTM guideline).

**28-d and 28-65-d are referred to C. riparius and to *C. tentans*, respectively (according to OECD guideline)

Annex D. Recommended test conditions for *L. variegatus*

Parameters	10 to 28-d toxicity test	References
Test type	flow-through	Phipps et al., 1993
Temperature	25 °C ± 2	Phipps et al., 1993
Photoperiod	16L:8D	Phipps et al., 1993
Test chamber	300- mL	Phipps et al., 1993
Sediment-water ratio	1:1/1.5	Phipps et al., 1993
Renewal of overlying water	2 to 6 volumes addition/d	Phipps et al., 1993
Age of organisms	Adults	Phipps et al., 1993
Number (loading) of organisms/chamber	10	Phipps et al., 1993
Number of replicates/treatment	not specified	Phipps et al., 1993
Feeding	20 mg salmon starter every 3 days	Phipps et al., 1993
Aeration	None	Phipps et al., 1993
Overlying water parameters	not specified	Phipps et al., 1993
Endpoints	survival, dry weight, reproduction	Phipps et al., 1993

5 How to evaluate and use ecotoxicological field tests for regulatory purposes

Factsheet FSM-006/00 date 13-02-2002

Authors:

B.J.W.G. Mensink, C.E. Smit and F.M.W. de Jong

Preface

The present document has been prepared to provide risk assessors at the Centre for Substances and Risk assessment (CSR) of the Dutch National Institute for Public Health and the Environment (RIVM) with guidance for evaluating and use of ecotoxicological field tests. In this field more initiatives are being taken, both on a national and an international level. In October 2001, a workshop was held at the RIVM on the subject of higher-tier studies. One of the recommendations of this workshop was to commit to paper the methods that are currently used for performing and assessing higher-tier studies. The present factsheet, although initiated before the workshop was held, can be seen as part of the execution of this recommendation.
This factsheet reflects current insights and gives an outline of the views and practices that are adopted at CSR at this moment. Specific guidance for the evaluation of aquatic field tests is included in the document. The factsheet is not a static guideline and will be updated regularly in the future. Extension of the document with specific guidance on field tests with earthworms and vertebrates is foreseen for the next version.

5.1 Introduction

5.1.1 Higher-tier risk assessment

In Annex VI to EU directive 91/414/EC, the uniform principles for the evaluation and authorisation of plant protection products have been established [1]. Annex VI specifies the criteria that have to be met with respect to the effects on human and environmental health. Regarding effects on non-target organisms, it is stated that a product will not be authorised when those criteria are not met, unless an appropriate risk assessment shows that under field conditions no unacceptable effects occur. In practice, this means that further testing may be considered by the applicant when the *first-tier* risk assessment on the basis of laboratory toxicity data and model calculations indicates that the use of a pesticide may impose a risk to the aquatic or terrestrial ecosystem. The directive, however, does not give explicit guidance as to what type of information should be submitted for this *higher-tier* risk assessment.
In the HARAP-guidance document on higher-tier aquatic risk assessment for pesticides [2], a number of approaches and techniques are discussed to further characterise the potential risks identified in the preliminary risk assessment. Partly based on this guidance, the following studies can be distinguished:

Table 5.1. Higher- tier studies

Indoor	additional tests with single species to study sensitivity distribution
	single species tests to study modified exposure, *e.g.* tests with sediment or plants present
	tests to study population level effects, *e.g.* tests with all life stages
	tests with multiple species, *e.g.* simple interaction tests or semi-realistic microcosms
Outdoor	tests with single (standard) species, *e.g.* bioassays
	tests with multiple species, *e.g.* micro- or mesocosms or enclosures

This overview of possible higher-tier studies indicates that the phrase “under field conditions” as used in the directive, is not interpreted strictly as “field studies”. As the term “field conditions” refers to a number of abiotic and biotic conditions that involve climate, physico-chemical characteristics, species composition, biological interactions and a number of other aspects, all studies that improve the insight into any of these aspects are considered valuable.

5.1.2 Aim and scope of the factsheet

5.1.2.1 Guidance for evaluation

Of the above mentioned types of studies, indoor and outdoor tests with multiple species are the most complex to evaluate. Methodological and interpretative difficulties have made that there is a high need for the development of evaluation tools for these studies [2-6]. Moreover, the most important recommendation of a recent workshop about higher-tier studies is the need to report present knowledge and practice concerning among other assessment of higher-tier studies [7]. There is a strong need for systematic guidance in field data evaluation based on the state-of-the-art, as it may increase (a) the consistency and (b) the transparency of evaluations and (c) the appropriateness of evaluations to be discussed by all parties involved. The aim of this factsheet therefore is:

1. To provide evaluators and risk assessors with an overview of the existing guidance on the performance and evaluation of field studies and related literature, and,
2. To give specific technical guidance on the evaluation of field tests in the process of pesticide registration. This second point involves the following topics:
   a. *The evaluation of the scientific quality of a field test*. The minimum package of requirements to which a field test should comply with is discussed, the test items that contribute to the overall scientific reliability are determined, and evaluation criteria are listed. The statistical substantiation of field test results, the interpretation of field test results and the derivation of a suitable endpoint also fall under this caption.
   b. *The evaluation of the usefulness of a field test for risk assessment*. The factors that determine the usefulness of a field test for risk assessment will be discussed.
   c. *The use of an endpoint from a field test for risk assessment.* This part of the factsheet will especially focus on the use of endpoints for pesticide registration.

5.1.2.2 Scope of the document

The factsheet is focused on field tests that are performed for risk assessment within the registration process. A related topic such as the use of field tests for deriving or validating environmental quality criteria will not be dealt with in the present factsheet, but is subject of a related factsheet [8]. Within the registration procedure, three types of field studies can be distinguished [9]:

1. Tests to validate the starting points and models that are used in the registration procedure, for example tests to establish the basic drift values that are used to predict the concentrations in surface water.
2. Pre-registration tests that are performed to prove that a product is harmless when a first-tier assessment indicates a potential risk.
3. Post-registration monitoring tests that are performed after a product has been allowed.

This document primarily aims at the second type of field tests, in the situation where tests are submitted as part of a dossier and have to be evaluated for a higher-tier assessment. As the performance of field tests is laborious and costly, it is expected that in the future applicants will consult the regulating authorities more often beforehand to achieve a test design that gives the requested information while being as cost effective as possible. This factsheet may therefore also serve as a tool to decide on the acceptance of a study proposal for a given goal. For the purpose of this document, field tests are defined as *Any ecotoxicological test that provides information on the effects of a chemical under (semi)field conditions*. This means that tests are considered that are performed under conditions that reflect the biological complexity and/or the abiotic conditions of the field situation. The tests are usually performed outdoors with multiple species, but indoor microcosms or field tests with only bees are also within the scope of this document. Depending on the scale, the outcomes are assumed to

reflect a complex range of interactions within and between species and may include food web relations. Ideally, the field test thus reflects the effects of a toxicant on all levels of biological organisation from sub-individual processes via communities and populations to ecosystems. Tests that are solely designed to give insight into the fate and behaviour of chemicals under field conditions, such as long term accumulation studies or field lysimeter experiments, are not considered in this factsheet.
In view of the complexity of many field tests and the serious gaps in the understanding of the ecotoxicological or ecological cause-effect chains on all levels of biological organisation, the experience of specialised scientists should be taken into account whenever necessary. By establishing criteria for data evaluation, it is tempted to make the evaluation more transparent in addition to the expert's view and to help risk assessors who are not primarily specialised in performing and interpreting particular types of field tests, to identify when to consult a specialised scientist.

5.1.3 Reader's guide

Chapter 5.2 gives an overview of the existing guidance on field tests. Chapter 3 goes into the assessment of the scientific quality and the usefulness of field tests. The minimum data requirement package that is to be addressed is discussed and Summary Tables as checklists for the evaluation of the scientific reliability of field test data are presented (§ 5.3.2), the selection of an appropriate endpoint is discussed (§5.3.3) and criteria to assess the usefulness of an ecotoxicological field test for risk assessment are presented (§ 5.3.4). Chapter 4 gives guidance on how the $NOEC_{FIELD}$ can be used for risk assessment. Chapter 5 contains the conclusions, recommendations and the working agreements that have been made on the basis of this factsheet.
A glossary of terms and abbreviations is given in Annex 1. Annex 2 contains a summarised overview of available guidance and literature concerning field trials, and an extended Summary Table for the evaluation of aquatic field tests is given in Annex 3.

5.2 Existing guidance on field tests

5.2.1 International activities in field test development

As authorisation of pesticides is increasingly performed within the framework of European legislation, it is logical that most efforts are on the European rather than the national level. There are a number of (inter)national working groups and discussion platforms active in the development of field tests and their use in higher-tier risk assessment, and the present factsheet elaborates further on their guidance.
A survey of recent activities on field tests is listed inTable 5.2. Some of the groups are mainly involved in the development of harmonised field test protocols, which is a way to facilitate the performance and evaluation of field tests and to increase their (international) acceptance. Other activities are more specifically focused on the use of field tests for risk assessment.

The activities of working groups and discussion platforms are organised according to various groups of non-target organisms. There are no activities that aim at the development of field tests in general, irrespective of a specific non-target group. However, the harmonisation of test protocols for a particular group of organisms is not the only answer to the problems associated with the interpretation of field tests. Attempts to improve the general scientific aspects, *e.g.* the statistical backbone, are as much important. In view of this, one could profit

Table 5.2. Recent activities in developing and harmonising field tests in the EU. EPPO activities on field tests are part of formalising environmental risk assessment decision trees.

NON-TARGET GROUPS	DISCUSSION PLATFORMS	NOTES
❑ Aquatic organisms	❑ HARAP, CLASSIC, EPPO, SAVE (TNO, RIZA, Alterra, RIVM)	❑ SAVE considers mesocosms as intermediate between laboratory and field tests. See references for HARAP [2] and CLASSIC [10]
❑ Earthworms (and other soil organisms)	❑ EPPO	❑ A draft version on guidance has been published: [11]
❑ Non-target arthropods I (bees)	❑ EPPO	❑ A draft version on guidance has been published: EPPO (2000)
❑ Non-target arthropods II (other than bees)	❑ EPPO, SETAC/ESCORT2, IOBC/WPRS, BART	❑ A draft decision making scheme and guidance has been published [12-15]
❑ Terrestrial vertebrates	❑ EU	❑ The EU-expert group birds and mammals, that involves the refined risk assessment for birds and mammals, also considers the role of higher-tier studies and field tests
❑ Plants	❑ not active	❑

from the activities of all above mentioned groups when seeking guidance for a specific type of field test.

5.2.2 Guidance in the Dutch Pesticide Act

The Dutch Pesticide Act has been updated in 1997 with guidance on when and how to perform ecotoxicological field tests for registration purposes [16]. As was noted in the previous paragraph, this guidance also refers to specific non-target groups, rather than to a general view on field tests, their set-up and their value for regulatory purposes. The guidance in the Pesticide Act has been summarised inTable 5.3.

Table 5.3. Guidance in the Dutch Pesticide Act on ecotoxicity field tests [16].

NON-TARGET GROUPS	GUIDANCE IN PESTICIDE ACT
❑ ***Aquatic organisms***	❑ Micro or mesocosm studies should generate sufficient data to evaluate important effects under field conditions. At least the highest level of exposure that can really occur by direct use, drift, drainage or run-off should be tested. A (model) ecosystem should be representative and include a sediment. Further guidance in [4,17]
❑ ***Earthworms***	❑ Field tests should generate sufficient data to evaluate effects under field conditions. A reasonable number of earthworms should be present in the test plots. The highest recommended application rate should be tested. A toxic reference product should be co-tested.
❑ ***Non-target arthropods I (bees)***	❑ Field tests should generate sufficient data to evaluate potential risks to bees, respecting their behaviour, mortality, and development as a beehive. Field tests should be performed with healthy populations of comparable natural strength. When bees are tested with varroacides, field testing should not occur within four weeks thereafter. Test conditions should be representative for the recommended type of application. Special effects on larvae, rest effects after a long period or disorientating effects may be tested in a follow-up. Field tests should be performed in accordance with EPPO Guideline 170 [18].
❑ ***Non-target arthropods II (other than bees)***	❑ (Semi)field tests should generate sufficient data to evaluate risks to non-target arthropods, other than bees. Field tests should be performed under representative agricultural conditions and in accordance with the recommendations for use, in such a way, that the realistic worst-case conditions are tested. Further guidance in [15]
❑ ***Terrestrial vertebrates***	❑ Field tests should generate sufficient data to evaluate the type and extent of risks, when the product is used in practice. Field tests should address foraging behavior, repellent product properties, feed alternatives, actual residues in feed, product persistence in vegetation, product degradation — *e.g.* on the treated crop, the extent of predation, the acceptation of bait, granules or seed, and the possibility of bioconcentration.
❑ ***Plants***	❑ No guidance on field tests with non-target plants

The requirements listed above are a mix of scientific reliability and usefulness of tests and the notion that there should be sufficient data for evaluation. There are no statements on the required extent of similarity between the field test conditions and the proposed conditions of use, *i.e.* those conditions that are recommended for GAP.

5.2.3 Available guidance per non-target group

In the following paragraphs, the available guidance for field tests is presented per non-target group (aquatic organisms, earthworms, bees, other non-target arthropods, terrestrial vertebrates and plants). The reader will be referred to the most relevant sources for further reading or guidance. A number of the cited references is summarised in Annex 2.

5.2.3.1 Aquatic organisms

Guidance on higher-tier testing for the aquatic environment can be found in the proceedings of the HARAP and CLASSIC-workshops [2,10]. Specific guidance on static mesocosm and fresh water field tests can be found in earlier publications [4,17] (See also Annex 1).
In Annex VI to Directive 91/414 [1], the endpoints in field tests are restricted to the "viability of exposed species, direct and indirect - via predators". Various endpoints, however, may be studied in a field test with aquatic organisms. Apart from the conventional endpoints in single species laboratory tests - mortality, growth and reproduction - a wide variety of additional endpoints may be studied that generally can be divided into within-organism, whole-organism, population and community level endpoints (see [19,20]). Functional endpoints can comprise all levels of organisation.
Within organism endpoints can be e.g. EROD activity, DNA damage, histopathology (gill, kidney, liver). Whole-organism endpoints can be e.g. survival and growth, while reproduction is a population level endpoint. Community endpoints can be e.g. diversity, density and taxa abundance. Functional endpoints can be e.g. oxygen production or decomposition. It is recommended to identify organisms to the lowest practical taxonomic level [19]. Guidance with respect to what is the lowest practical level is given in the HARAP document [2]. For arthropods, this usually means species or genus (although immature insects can sometimes only be identified at the level of family), for algae identification to class is sometimes the best achievable level. Ecological evaluations are usually easier if at least the main taxa of concern in the study are identified to species. Identification to species may also permit the application of more powerful multivariate analyses of community structure. The appropriate level is therefore partly determined by the objectives of the study.

The technical guidance that is given in Chapter 3 focuses specifically on aquatic field tests.

5.2.3.2 Earthworms

Discussions on the advantages and limitations of field tests with earthworms, effect criteria, endpoints, test design, sampling methods and the results comparison between laboratory and field tests have been compiled in a book on earthworm ecotoxicology [21]. A recent EPPO proposal for earthworm field tests recommends ca. 10 endpoints, among which: species diversity, total biomass of at least the two most abundant species, and the biomass of all earthworms and the biomass of the tanylobous and epilobous individuals (juveniles and adults separately) [11]. Further guidance for field tests with earthworms can be found in the ISO Guideline 11268-3 [22]. Specific technical guidance for the evaluation of earthworm field tests, including Summary Tables, will be part of the next update of this factsheet.

Some comments related to the statistical evaluation of earthworm field tests are given in § 3.2.2.2 (end) and § 3.3.1

5.2.3.3 Non-target arthropods I (bees)

The EU [1] restricts the endpoints in field tests with bees to the viability of larvae, the behaviour of the honeybee, the viability of the colony, and the development of the colony. Further guidance on evaluating field tests with honeybees can be found in EPPO publications [18,23].

5.2.3.4 Non-target arthropods II (other)

The evaluation for non-target arthropods has been separated into an in-field and an off-field evaluation in an EPPO proposal [12]. These evaluations deal with the land intended for agricultural production activities (incl. horticulture and forestry), and with all areas that surround the field, natural and semi-natural habitats, in particular hedgerows and woodland. This EPPO proposal for field test formalising aims at testing in accordance with GAP. It also aims at testing at the highest recommended dose, and not at different doses to indicate dose-effect relations. Further guidance for field tests with non-target arthropods can be found in several documents [12,13,15,24].

5.2.3.5 Other non-target groups

There is no generally accepted guidance available with respect to other non-target groups such as terrestrial vertebrates and plants. Further guidance on field tests with birds and mammals can be found in AEDG [25]. Field testing with these groups is probably more *ad hoc* oriented than for the aforementioned non-target groups. For vertebrates, this is partly due to the fact that testing is kept to a minimum where possible in view of ethical considerations. This also implies that any field test with birds or mammals is very much specified by the underlying question. There are, however, examples of tests available that can give some insight into the possible set-up. An example of such a field test was one in the US on the impact of azinphos-methyl on the grey-tailed vole *Microtus canicaudus* [26]. Various physiological endpoints were measured in this field test with 24 enclosures of 45 by 45 m. Five increasing dosages were tested with four replicates per dosage. The remaining four plots were used to increase the replicates of the middle and the top-dose. Various endpoints that were sensitive to both the pesticide treatment and the geographical location were detected using multivariate statistics, and biochemical methods were used to verify exposure of voles in the treated plots.
Specific technical guidance for the evaluation of field tests with birds, including Summary Tables, will be part of the next update of this factsheet.

5.3 Technical guidance for the evaluation of field tests

5.3.1 Introduction

The quality of any test may be defined in terms of its scientific reliability and its usefulness for a particular purpose (in conformity with USES 3.0 [27]). Field tests can be very diverse in their problem definition and methodology. Not all field tests are therefore scientifically reliable and/or useful for environmental risk assessment or registration. However, there are general criteria to which any field test should comply. In principle, all tests with non-target groups should meet these general criteria (seeFigure 5.1).

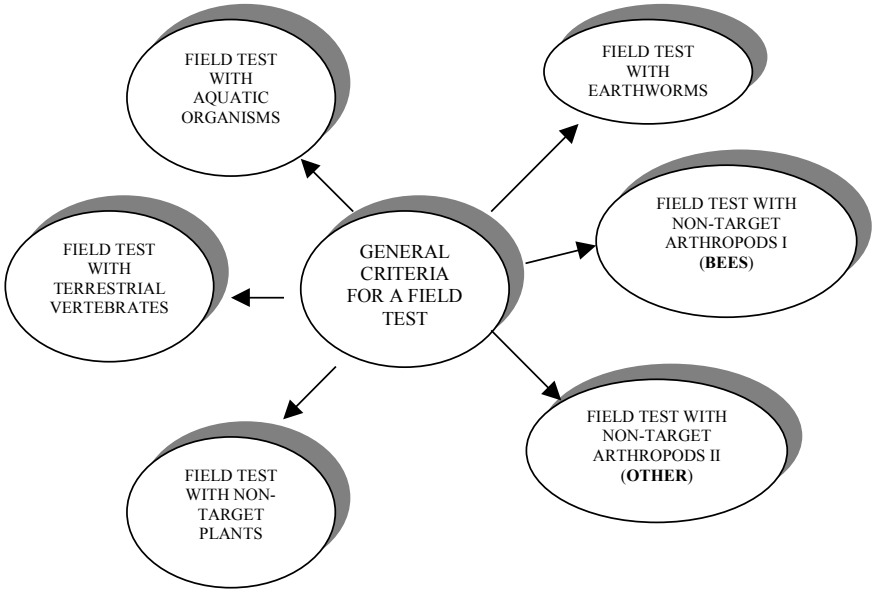


Figure 5.1. The general criteria of a field test and related field tests with non-target organisms

Section 5.3.2 goes into the specific requirements for field tests in terms of scientific reliability. A Summary Table is presented that can serve as checklist for evaluating the scientific reliability of field tests with aquatic organisms in particular. Quality criteria for the evaluation will be presented by listing the relevant items on: (1) methodology and field test description, (2) results, and (3) additional remarks. An extensive Summary Table has been worked out for aquatic organisms only, as most experience with the methodology and interpretation of field data seems to have been built up for this non-target group (Annex 3). Many of these test items, however, also refer to field tests with other non-target groups. The reader is referred to the HARAP and CLASSIC-document and additional literature for more specific guidance [2,4,10].
Section 5.3.3 describes when effects are deemed ecological relevant, and gives some guidance how to handle recovery. The aspect of usefulness for environmental risk assessment is dealt with in Section 5.3.3, and Section 5.4 discusses the risk assessment on the basis of the selected endpoint.

5.3.2 Evaluation of scientific reliability

5.3.2.1 Minimum data package and Summary Tables

The scientific reliability of a field test is determined by a 'minimum' data package that should be addressed properly in any test (seeTable 5.4).

If a field test does not comply with one of these test items, the test is unreliable and therefore not useful for risk assessment (except perhaps as circumstantial evidence, depending on the quality of specific elements of the test). These items must be dealt with to perform a proper effects analysis: it should be clear whether there is any cause-effect relationship between the administration of a pesticide and the observed effects.

The effects analysis in a field test should be linked with the exposure analysis: it should be clear to what extent the (preferably most susceptible) organisms are exposed.

Table 5.4. Minimum data package that should be addressed properly in any field test

TEST ITEM & REQUIREMENT	EXPLANATORY NOTES
1. Concentration(s) or dosage(s) must be reported properly	1. Concentration(s) or dosage(s) should be seen in relation to exposure: are the test organisms actually exposed as intended? See also the use of positive controls below (point 5). It is absolutely necessary that in a field test actual concentrations are measured (to verify the actual exposure).
2. Duration (of exposure) and observation time must be reported properly	2. Exposure should be seen in relation to time and the extent of the effects (e.g. acute versus short-term or acute versus chronic).
3. Application	3. The type of application (*e.g.* spraying or direct mixing) should be reported.
4. Test design should allow proper statistical analysis to determine a valid endpoint (*e.g.* an EC_{10}, EC_{50} and/or NOEC).	4. The number of replicates per treatment may be decisive for identifying differences in the system response. Identification of small but significant differences require more replicates (see e.g. [28]).
5. Control(s) should be included in the test design and their results should be reported properly	5. Negative controls are essential; positive controls may improve test reliability, as they may confirm whether exposure has actually occurred. However, for reasons of animal welfare, positive controls may not be preferred. In Germany *e.g.* bee field tests with positive controls are not allowed.
6. Endpoints(s) should be properly defined and reported in relation to the time points of observation.	6. Field tests may include more endpoints than the traditional first-tier endpoints of mortality, growth or reproduction.
7. Biological interpretation should be plausible.	7. Statistically significant differences — that may be due to the pesticide — should be eco(toxico)logically interpreted, where possible.

Relevant guidance for pesticide evaluation has been condensed to Summary Tables. These are used for pesticide evaluation by RIVM/CSR [29]. A Summary Table is helpful in determining the reliability of experimental tests. Summary Tables comprise a wide array of test items that may influence the test quality, they function as a checklist and they refer to scientific guidelines, when necessary.
The reliability is assessed by assigning a Reliability Index (Ri) to a particular (field) test: Ri1 stands for a reliable test, Ri2 for a less reliable, and Ri3 for an unreliable test.

Ri3 tests are not used for any purpose. Both Ri1 and Ri2 tests can be used for risk assessment, but it depends on the overall data availability, whether only Ri1 tests should be used, or whether Ri2 tests can be used as well.

An overview of all relevant test items of a Summary Table has been listed in Table 5.5 with a few key words. Table 5.5 has been further worked-out extensively for aquatic organisms in Annex 3 (Table A3.1). Table A3.1 has been provided with a user's instruction.

Most of the test items and notes handled in Table 5.5 are valid for terrestrial higher-tier studies as well. An important difference could be the way of exposure. Terrestrial non-target organism could be exposed directly by spraying of the pesticide, or indirectly by food or contact to contaminated surfaces or soil particles. In this case, it is of special importance that it is made plausible that actual exposure did occur and that the applied dose and exposure concentrations are measured. The existence of refuge areas should be kept to minimum, unless of course avoidance is explicitly included as an effect parameter. Where in aquatic field studies the regular analysis of concentrations during the test is necessary for the interpretation of the results, dose-verification at application may be sufficient for other types of studies, such as field tests with earthworms or non-target arthropods.
The test item on statistical design and interpretation will be dealt with specifically in paragraph 5.3.2.2.

Table 5.5. The test items determining the quality of a field test in general (further explanation for field tests with freshwater organisms in Annex 3, Table A3.1)

TEST ITEM	NOTES
1. substance	
1.1 concentration	
1.2 purity	[identity of impurities? how much?]
1.3 formulation	[identity? how much?]
1.4 vehicle	[identity? how much?]
2. duration, observation time	[if not reported ⇨ Ri 3]
3. application	
3.1 type I	[if not reported ⇨ Ri 3]
3.2 rate(s)	[dosage, *e.g.* kg.ha-1]
3.3 type II	[additional technical data]
3.4 date, year, location	[properly reported?]
3.5 (micro) climate	[reported?]
4. test design	
4.1 statistics	[proper for NOEC and interpretation of other results? if not ⇨ Ri3]
4.2 type & size	[properly reported?]
4.3 dose-response	[properly indicated and reported?]
4.4 test system I	[how stable? how natural?]
4.5 test system II	[properties during test]
4.6 pre-treatment	[proper equilibration?]
5. test organisms I	[representative? sensitive?]
6. test organisms II	
6.1 general features	[monitored?]
7. control	[if invalid ⇨ Ri3]
8. extraction/analysis	[appropriate?]
9. endpoint	
9.1 type	[relevant?]
9.2 value	[if not properly reported ⇨ Ri3]
10. effect class	[properly derivable? If not ⇨ Ri3]
11. biological meaning of statistically significant differences	[sufficiently explained?]
12. verification of endpoint	[possible?]
13. actual concentrations/dose	[measured? if not ⇨ Ri3]

5.3.2.2 Reliability and statistics

5.3.2.2.1 NOEC and ECx-design

The ultimate goal of a field test is to prove that the use of a pesticide in practice will not result in unacceptable effects. Several ways to achieve this aim exist.
One way is to derive the concentration at which no effects on the structure or functioning of the ecosystem occur. This No Effect Concentration (NEC) lies somewhere between the highest tested concentration without effects (No Observed Effect Concentration, NOEC) and the lowest tested concentration with effects (Lowest Observed Effect Concentration, LOEC). The NOEC can thus be regarded as a safe estimate of the NEC.
A NOEC can be derived by applying an ANOVA followed by a (multiple) comparison of means. Several parametric and non-parametric tests can be used for this purpose. A set of the most appropriate tests is included in the TOXSTAT® computer program [30]. By nature of these types of tests, the NOEC decreases when the number of replicates per concentration is increased, because the variance within the tested groups is be decreased in this way and differences are detected as significant at a lower level. However, by increasing the number of replicates, the number of test concentrations has to be equally lowered to keep a manageable experiment. As the NOEC is by definition determined by the chosen test concentrations, lowering the number of test concentrations and thus increasing the space between them, will result in a NOEC value that can be lower or higher than the actual concentration at which no effect occurs. In several OECD and ISO guidelines, a geometric series of concentrations is advised, and concentrations should not differ by more than a factor of 1.8 ($^4\sqrt{10}$) to 3.2.

For a proper NOEC estimation, a geometric series of concentrations should be used and as a rule of thumb, concentrations should not differ by more than a factor of 2 to 3.

In some cases, a test is performed with a control and only one dose or concentration in a large number of replicates. A NOEC from such an experiment does not give an answer as to what concentration is “safe”: if there is no adverse effect, the actual NEC may be much higher and

if there is an effect, the NOEC is somewhere between the test concentration and the control. Strictly speaking, a test with one test concentration cannot yield a NOEC, neither can a test without an effect despite the number of concentrations tested. In this case, the NOEC is always greater than or equal to the highest test concentration. If such a test is performed with a concentration that is higher than the Predicted Environmental Concentration (PEC), such an experiment could be convincing for the absence of effects in practice. In such case however, the inclusion of a "positive" control, by means of applying a very high concentration at which clear effects are found or using a toxic reference compound, is needed. Only then one can be sure that no other circumstances caused the absence of effects.

A test with one concentration which does not result in an effect, can only be used for risk assessment when a 'positive' control is included by application of a very high concentration of the test compound, or by using a toxic reference compound with known activity. The test can, however, not result in a real NOEC and the result is expressed as NOEC ≥ test concentration.

An alternative approach is the concentration-response design that aims at revealing the relationship between concentrations and effects. This relationship is obtained by fitting the effect data according to a response model. The most frequently applied model is the sigmoïd dose-response model in which the effect increases with the concentration according to an S-shaped curve. From this curve it is possible to fit an EC_x, the concentration at which x% effect occurs (see [31,32] for formulas for EC_{50} and EC_{10}). It must be noted that other response models might be more appropriate. An overview of different models can be found in the Dutch Leerboek Oecotoxicologie [33] and references cited therein.

5.3.2.2.2 Statistical power

Regardless of the model fitted, the number of concentrations has to be high enough for a reliable estimate of the concentration-response curve, whereas the number of replicates has to be sufficient to gain insight into the natural variability. As a rule of thumb, at least five concentrations are required. This implies that for a manageable experiment the number of replicates may be restricted. With a limited number of replicates in a field test it may, however, appear that the observed or measured effects are due to coincidence rather than to the mere action of the pesticide. There is no easy answer whether a proper and sophisticated statistical design can tackle this problem. Besides, due to the large natural variation in untreated plots or groups, it may be very difficult to discern between this variation and the effects actually due to the treatment, *e.g.* the application of a pesticide. The control plots in a field test with earthworms may show large differences: control plots with no earthworms and control plots with 100 individual earthworms per m^2 are no exception. A proper statistical design and a proper analysis of the test results are then essential. The statistical analysis is primarily dependent on the field test protocol: after all, the number of replicates per treatment and control is part of the experimental set-up.
There are three questions with respect to the power of the test. In an ideal case, the aforementioned three questions were discussed in the test protocol prior to carrying out the field test, so that a test is performed only when sufficient power is ensured.

1. using an experimental protocol, how large is the probability P (=1-β) that a difference of size δ will be detected,
2. how many replicates are needed to be able to observe a difference of δ with probability P,
3. when is a difference of size δ ecologically relevant [3,34]?

Question (1) and (2) make heavy demands on the experimental protocol. These questions refer to the power of a statistical design *c.q.* test: the power of a test is 1-β, where β is the

type-II error. β is the probability that a factual difference of size δ remains unnoticed, so that the null hypothesis of no significant difference is falsely accepted (false negative). To improve the power of a given test (or decrease β) while keeping α constant for a stated null hypothesis, the sample size should be increased. The significance level α is referred to as the type-I error, that is the probability of incorrectly rejecting the null hypothesis (false positive). It is noted that if α is 0.05, one out of twenty tests will indicate an effect that is not present. If for instance a lot of parameters are measured, it is to be expected and accepted by the statistical method chosen, that in 5% of the cases an effect will be indicated.
Since the differences are expressed as a fraction of the standard deviation, all efforts to decrease the standard deviation will increase the sensitivity of the trial. An important possibility is to obtain insight in the variation in relevant parameters *before* actually performing the experiment. It could be decided, for instance to exclude outliers, or to choose a randomised block design. On forehand, it can be calculated what differences could be traced with the number of replicates used. In the aforementioned case of earthworms, it will be clear that only very severe effects can be traced. An investigation into the number of replicas needed for the desired power in a field experiment with four treatments and a control [28] showed that the staggering amount of 65 replicates would be necessary to identify significant effects of a difference of 0.528 times the standard deviation σ (power 0.8, α is 0.05).
It is recommended to take this kind of options into account prior to the actual field test performance and to weigh the expected scientific relevance of the data against the costs**.**
There are several possibilities to increase the sensitivity of the test. It is for instance possible to increase only the number of replicates in the untreated control. In the above mentioned example, this would mean that 45 replicates would be needed per treatment, and 90 in the untreated, to achieve the same conditions. Another option can be to compare the same test items before and after treatment. In the case of an experiment with natural vegetation, the increase or decrease per plot was compared, leaving a part of the variations between plots with the same treatments partly out of scope; the sensitivity of the test in this way increased considerably (see [35]).
Question (3) may be very difficult to answer, as it requires sufficient knowledge on all eco(toxico)logical aspects of e.g. a Terrestrial Model Ecosystem (TME), whereas the experimental experience with such systems is still in an early stage (*c.f.* [36]).
A first indication of the power of a test can be obtained by comparing a comparison of the NOEC with the EC_x. both endpoints gives a first indication of the power of the test. When for a certain parameter the NOEC is higher than the EC_{50}, this means that a 50% effect cannot be detected as significant. In this case, a possible effect on this parameter cannot be detected with sufficient reliability. NOECs lying between the EC_{10} and EC_{40} should be judged carefully. It should then be considered whether the test provides sufficient evidence to prove that there are “no unacceptable effects under field conditions”.

Where possible, the data should be used to establish a concentration-response curve. Tests in which the NOEC is higher than the EC_{50} should not be used for risk assessment. NOECs lying between the EC_{10} and EC_{40} should be judged carefully. A decision should be made as to whether the test provides sufficient evidence that there are “no unacceptable effects under field conditions” and the test is thus suitable to be used for risk assessment.

This rule of thumb may impose a problem with respect to the earthworm field tests, as the minimal effect that can be significantly detected these tests is ca. 50% [11]. A Student t-test on the logarithmically transformed earthworm counts of each species can be a useful tool to determine whether the statistical power is sufficient to obtain a confidence level of 95%.

The logarithmic transformation is needed to convert the Poisson distribution of the earthworm counts into a normal distribution. A difference of log (0.5) = - 0.301030 between the logarithmically transformed control counts and the logarithmically transformed treated counts indicates a 50% reduction in numbers. Next, a two sample Students t-test assuming equal variances with a hypothesised mean difference of 0.30103 is performed after a logarithmic transformation of the data for the control plots (98, 51, 554, 140) and for the treated plots (212, 70, 66, 395). The difference between the log transformed average values is only 0.000215 but the pooled variance is 0.1676. This gives a t-value of -1.0390, which corresponds to a one-tail percentage of 16.94. Therefore the statistical power of the above test is not sufficient to show that the numbers of worms are reduced by less than 50%. When more data are added, for example: 75, 661, 193, 73, 63, 441 for the control and 88, 54, 331, 99, 411, 301 for the treated plots, the statistical power increases and a t-value of -1.745 is obtained. This corresponds to one tail percentage of 4.9, which is sufficient to show that the numbers of worms are reduced by less than 50%. When the data are less variable, it is much easier to show that less than 50% reduction has occurred. The difference between the control counts of 98, 90, 111, 140 and the treated counts of 84, 70, 66, 60 is relatively large because the difference between the average of the log-transformed data amounts to 0.1946. The pooled variance of these log-transformed data is only 0.00537 and the t-test shows that there is only a 4.06% chance that more than 50% of the worms are inhibited.

5.3.2.2.3 Multivariate statistics

Most field studies are performed in a system where the toxicant is only one of several factors that determine the development of the system. To disentangle the effects from the test substance from those of other factors, an adequate experimental layout and the use of advanced statistical techniques are required, especially when the response is varying with time. Multivariate statistics such as Principle Component Analysis (PCA), Nonmetric Clustering and Association Analysis (NCAA) or Principle Response Curves (PRC) seem to be promising tools to derive a NOEC or EC_x in this respect [26,37].
In the aforementioned field test with rodents (paragraph 2.2.5, [26]), the use of a one-way ANOVA was compared with a NCAA and it was demonstrated that a conventional method of statistical analysis did not reveal particular pesticide-related effects, in contrast with the use of the more advanced statistical technique. It was concluded that a high within-treatment variance might obscure subtler but treatment-related differences when using ANOVA. The more robust NCAA approach may detect associations among responses, as the individual response is the basic unit for NCAA, whereas the plot is the basic unit for ANOVA. In the analysis of zinc effects on nematodes in outdoor soil mesocosms, it was also observed that PRC yielded a lower NOEC than more “classical” endpoints, such as number of taxa or species diversity indices [38]. Part of the strength of PRC is that all available information on species numbers per time and concentration is analysed simultaneously. For the other endpoints, information and/or statistical sensitivity is lost in the procedure by integrating underlying information into one single value or by a step-wise analysis of data.

If multivariate analysis is performed by the applicant, a specialist should be consulted to evaluate the reliability of the methods and results. If no multivariate analysis is performed, a specialist should be consulted to judge whether the data are suitable for such an analysis. In this case, the regulatory authority should be advised to have such an analysis performed, either by the applicant or by an independent specialist.

5.3.3 Selection of a relevant endpoint

5.3.3.1 Ecological relevance

In most cases, several scientifically reliable NOECs can be derived from one field experiment: NOECs for single species, for community endpoints such as species diversity or for ecosystem functions such as decomposition. Furthermore, different NOECs for different time points can be derived. Choosing the lowest of all available values as the $NOEC_{FIELD}$ is a too straightforward way to handle this. Defining ecological relevance by statistical difference only is incorrect: first, the ecological relevance of an effect should be determined and second, the experiment should be designed to test for the existence of such an ecologically significant effect, and not in the reversed order [26]. The Dutch Health Council [3] points at the fact that significance is not necessarily related to ecological meaning and that statistically insignificant effects can have large implications in an ecological sense. However, as long as the concept of ecological relevance lacks a concrete interpretation, it is impossible to state to what extent an effect is acceptable. The Council points at the possibility to relate observed effects to the natural fluctuations that are observed in an undisturbed control area. This indicates that time plays a crucial role in the decision as to whether an effect is acceptable: changes that recover within a reasonable time span are considered acceptable.

The fact that the extent and duration of an effect determine the relevance, is also recognised by the EPPO-working group on soil organisms [11]. Based on the practical experience that the minimal effect that can be significantly detected in a field test with earthworms is ca. 50% and the fact that "the potential for recovery [..] needs to be considered", the group proposes the following classification of results for an earthworm field test:

- no risk if the effects ≤50%;
- medium risk, if the effects >50% during the test, but recovery has been observed within one year;
- high risk, if the effects are still >50% one year after application.

It is clear that this classification is still largely based on the statistically detectable differences, and that the test is not suited to trace effect percentages below 50%. The fact that it is generally not possible to detect differences up to 50% as significant is felt as a serious drawback of this type of test. As was mentioned above in § 3.2.2.2 on statistical power, tests in which the NOEC is higher than the EC_{50} should not be accepted for risk assessment, and NOECs between the EC_{10} and EC_{40} should be judged carefully. As the test was performed to prove that there are "no unacceptable effects under field conditions", care should be taken that this point is adequately addressed with the test.

HARAP [2] and CLASSIC [10] propose to use an Ecologically Acceptable Concentration (EAC) for the higher-tier risk assessment of pesticides in aquatic ecosystems, and this view is also present in the Guidance Document on Aquatic Ecotoxicology that was prepared in the framework of EU Directive 91/414/EC [39]. The wording makes clear that to a certain extent effects are accepted and are assumed not to cause damage to the ecosystem on the long-term. The Guidance document states that "EACs from reliable static mesocosm studies should be regarded as generally representative or possibly conservative for surface waters in most agricultural landscapes". This statement, however, does not consider the similarity of the test system with the situation to be assessed (see § 5.3.4). Besides, there is no consensus about the qualitative interpretation of the EAC, the wording "ecologically acceptable" is not substantiated with an effect percentage. EACs should therefore not automatically be regarded as NOECs, the classification of effects as used in the STOWA-reports (see below) is preferred.

EACs from aquatic field studies are not automatically accepted as NOECs for risk assessment.

In the STOWA-reviews on aquatic field studies [40,41], a classification was used that is based on a qualitative description of effects, in combination with the duration, the number of times the effect is observed and the sensitivity of the endpoints considered:

- Class 1: *no observed effects*: (a) due to the treatment, and (b) no clear causal relations between the observed differences in treated groups and controls;
- Class 2: *slight effects:* (a) effects reported as: slight and transient, and (b) short-term and/or quantitatively limited response of sensitive endpoints, and (c) effects only found on individual time-points;
- Class 3: *large, acute/short-term effects:* (a) a very clear response of sensitive endpoints, but a total recovery within 8 weeks after the last application, and (b) effects are reported as "temporary effects on more sensitive species", or as "temporary elimination of sensitive species", or as "temporary effects on less sensitive species or endpoints", and (c) effects observed on some consecutive time-points;
- Class 4: *large effects in short-term test:* very clear effects (*e.g.* large reductions of functional endpoints and elimination of sensitive species) during the whole test, though the test duration is too short to demonstrate a complete recovery within 8 weeks after the (last) application of the pesticide;
- Class 5: *large, long-term effects:* (a) very clear response of sensitive species and recovery of species after >8 weeks after the last application, and (b) effects are reported as "long-term effects on more sensitive species", or as "elimination of sensitive species", or as "effects on less sensitive species or endpoints", and (c) effects observed on various consecutive time-points.

From this classification it follows that ecological relevance is not only related to the extent of the effect relative to the control performance and its duration, but also to the number of times the effect is observed and to whether species or endpoints that were considered insensitive are yet affected.

Based on the available laboratory data, the use category (herbicide, insecticide etc.) and/or the mode of action of the substance, it may be expected that some species or endpoints show a (temporary) effect. Unexpected effects on species or functional endpoints that were initially not identified as sensitive are regarded as a weighty argument for classification in a higher effect class. The occurrence of an effect on consecutive time points is likely to be more related to substantial damage to the ecosystem than an effect that is observed once or with some intervals. The potential of a system to recover from the application is an important item. In the above system, a duration of 8 weeks is chosen because in most experiments the sampling frequency was biweekly or monthly. Eight weeks thus means that no effects are observed at 2 to 4 consecutive sampling occasions.

A translation of this classification in terms of NOECs and LOECs can be made as follows. The lowest concentration belonging to class 1 is considered as the NOEC, the lowest concentration with effects in classes 2 to 5 is the LOEC. For herbicides, concentrations with class 2-effects that relate to physiological endpoints such as oxygen production, conductivity and primary production, are sometimes regarded as a NOEC. For insecticides, class 2 effects relate to structural effects and are only used as LOECs. Where possible, this classification should be used to derive NOECs for aquatic field studies.

Because the potential for recovery is recognised as an important issue for the determination of an ecologically relevant test result, the next section discusses some considerations with respect to this item.

5.3.3.2 Recovery

Recovery is a general term that indicates that the endpoint under consideration has returned to normal values. The following theoretical prerequisites for recovery were mentioned in the STOWA-review on herbicides [40]:

- the toxic substance disappears and/or the bioavailability decreases to concentrations below the level where effects occur
- other relevant environmental conditions such as food availability, nutrient status, temperature remain at or return to values that are sufficient for the affected population
- the generation time of the affected species is shorter than the duration of the study
- re-colonisation is possible in case populations are diminished completely.

Brock and co-authors [42] regard a population as recovered when the abundance of individuals, after having been significantly reduced or increased, is within the scatter of the control during a longer time period. Populations start to increase when surviving individuals are reproducing again, or when the system is re-colonised from nearby undisturbed areas. Relating recovery of a treated system to the control performance implies that control and treatment systems were comparable prior to the test. In most aquatic studies, experimental micro- or mesocosms are used and reasonable similarity between the systems can be achieved. This also holds true for controlled outdoor bioassay experiments with selected test species [43]. In other cases, for instance in earthworm studies or studies into the effects on a natural vegetation, experiments are performed on existing field sites where similarity between the control site and the treatments cannot always be achieved. In this case, pre-treatment characterisation of the plots is necessary to gain insight into the development over time.
In a field study aimed at the effects of glufosinate-ammonium on off-field vegetation [35], both short-term (weeks after the sprayings in May-June) and longer-term (month) effects were studied. In this case short-term phytotoxic effects were severe, and were statistical significant even at the lowest dosages applied (2-4% of the actual field rate of 800 g a.i./ha). A comparison of the effect parameters after ca. two months showed only a small significant decrease in species number and cover in the highest treatment (64% of 800 g a.i./ha) compared to the control. When the *increase* or *the decrease* between spring and autumn of species number and cover was compared between treatments, effects were found from 4% of the highest field dose upwards.
Recovery of a functional endpoint is not necessarily related to a recovery of species numbers: different species can have taken over a process (functional redundancy). In theory, a species diversity index can be unchanged while all species have been replaced by others. In case individuals are not identified to the species level, the species composition within a taxon can be markedly changed while abundance on the taxon level has recovered. The Health Council states that recovery means that “a new situation arises that resembles the original or an undisturbed control situation more, as the disturbance has been less drastic and temporary”. The discussion as to what changes and what recovery periods are acceptable are only starting yet. However, when recovery is included as variable in an experiment, a worst-case approach should be followed.

Taking a worst-case scenario into account, a field test that is designed to include measurements of the recovery potential of populations or communities, should comprise species with a limited recovery potential, *e.g.* due to a slow life-cycle, or a limited mobility [37].

5.3.4 Usefulness of an ecotoxicological field test for risk assessment

In the previous Sections, the evaluation of a field test for the purpose of risk assessment has been addressed, with emphasis on scientific quality. Scientific reliability is a prerequisite for a field test to be used for registration purposes. This Section focuses on the next step: how to use a valid $NOEC_{FIELD}$ from a field test (*i.e.* reliable or less reliable, though not unreliable, see also § 5.3.2) in environmental risk assessment. Two main questions can be identified:

1. Does the test answer the question?
2. Is the field test representative for the situation that is to be assessed?

The first question may seem obvious, but too often tests are supplied that are highly reliable in a scientific way, but do not assess those aspects that were reason of concern in the first-tier assessment. It should be clear that no definite criteria can be given for this aspect. Nevertheless this aspect of problem definition should ideally be handle during the phase of designing the test protocol. Aspects that can give direction to the effects that should be studied, are for instance compound properties and proposed use. These data, available from the first-tier risk assessment, should be taken into account while developing and assessing a field trial.

Information on the working mechanism and proposed use, and the available first-tier toxicity data give information on the sensitive species groups and should be considered when evaluating the usefulness of a field test for higher-tier risk assessment.

The second question relates to whether the conditions of the field test resemble the situation under consideration. Some aspects that determine the similarity of a test system to the system to be assessed are given in Figure 5.2. The items will be discussed below.

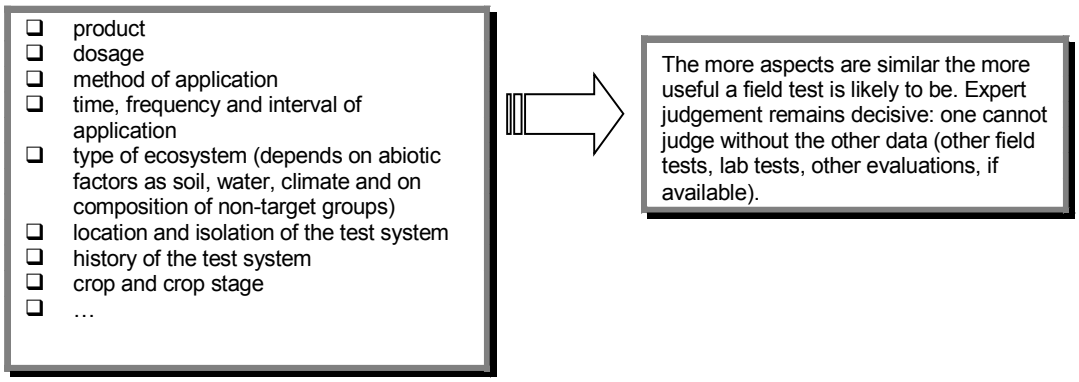


Figure 5.2. The similarity aspects that determine the usefulness of a field test

As a rule of thumb, it can be expected that the more of these test aspects are similar, the more useful the field test is expected to be for environmental risk assessment.

5.3.4.1 Product and dosage

The test should be preferably carried out with the product under consideration. Other products may be used provided that the dosage in terms of the active ingredient is the same, and the effect of the blank formulation is similar. With respect to the latter, inert formulations such as water dispersible granulates or wetting powders can generally not be used to assess the effects of suspensible or emulsifiable concentrates that contain surfactants which can enhance effects. Spray solutions cannot be used to assess the risks of granules or pellets to terrestrial organisms.

5.3.4.2 Method of application and exposure

The method of application is one of the factors that determine exposure. In a field test where e.g. a pond with floating vegetation has been sprayed from above, the exposure of aquatic organisms may be delayed as the intercepted pesticide may be washed off the leaves by the first rain. Experiments in which the product is diluted in the water column may be adequate to assess the risks for species living in the water, but the contact toxicity of herbicides to emerged plants or floating plants such as water lily (*Nymphaea alba)* is likely not to be covered by this exposure method.
In principle, the product should be applied to the test system in a way that simulates the real situation. However, simulating drift in an aquatic experiment by spraying from a certain distance of the systems would lead to uncontrolled exposure. Therefore, spraying the systems with a fraction of the intended field dose can simulate the exposure after drift.

5.3.4.3 Time, frequency and interval of application

In general, the time, frequency and interval of application in the field test should follow the recommended field practice. This means that in principle a test with a single application cannot be used to assess the effects of a product that is applied multiple times. For non-target arthropods, however, the first SETAC/ESCORT workshop [15] recommended to test a product at 2x the recommended field rate when its is used two or three times per crop season. It should be noted that a single application at the double dosage is not the same as a double application at a lower rate. A single application could be sufficient when the test gives no effects or shows clear recovery and the product under consideration is rapidly disappearing and applied a large time intervals. Recent developments in non-target arthropod testing point into the direction of a dose-response test design [12,13].
Considering the time of application, the CLASSIC-workshop on the interpretation of aquatic micro- and mesocosm studies [10] recommended to perform field tests in spring or summer, because the level of biological activity is higher and the chance that effects are observed is thus more likely. The moment of administration of the compound in practice should be considered as well. A field trial performed in autumn is for instance less useful to assess the effect of an application in spring. A general requirement is that the test should include the most sensitive life-stages that are potentially exposed. It should be noted that while a spring or summer experiment might reflect a *worst-case* in terms of effects, it does not necessarily do so in terms of recovery. It may be expected that in spring or summer the disappearance of substance is faster because of the raised temperatures and that the chances for re-colonisation are higher.

5.3.4.4 Type of ecosystem

The ecosystem in the test should be relevant for the situation to be assessed. For registration in the Netherlands, this means that ditches with a low stream velocity are more relevant than stagnant ponds, fast streaming creeks or rice fields. Abiotic factors also include the characteristics of sediment and water, for instance nutrient availability. Temperature, rainfall and other climatic parameters should be in the range normally expected during the time of application in the Netherlands. For terrestrial field studies, the soil type must comply with the characteristics that are also used for the assessment of fate and behaviour in soil [29]. As exposure is greatly determined by the physical environment that surrounds the individual non-target organisms, this should reflect the situation to be assessed (see also Section 5.3.4.7).
With respect to the *biotic* part of the system, the Health Council gives as a ground rule that test organisms must include species from different taxonomic groups, from different trophic levels, with different ecological functions and with a different life history [3]. The chosen test

species should include species that are common in the Netherlands or species that have shown to be relevant [44] and should include susceptible life-stages [15]. With respect to aquatic systems, the CLASSIC-workshop advised not to include fish in aquatic field tests because they will lead to unbalanced situation given the scale of most tests systems [10]. The inclusion of macrophytes is advised, but growth should be managed. For non-target arthropods, the choice of species is governed by the results of the first-tier assessments, whereas the species in an earthworm field test are mainly determined by the choice of the test location (grassland or arable land). As registration generally concerns agricultural areas over a whole country, various agro-ecosystems may be involved. This implies that there could be a need to perform tests in different systems depending on the intended use. An example for this is the risk assessment for birds of a product that is applied on fruit and on maize. The bird species that are found in fruit orchards are different from those found on maize fields and adjacent grassland, and field tests for this product should thus include different species.

5.3.4.5 Location and isolation of the test system

The type of habitat or ecosystem that should be protected determines the location of the test. Within this context, the EPPO Panel on Environmental Risk Assessment has recently proposed to distinguish between the various areas in in-field and off-field areas. The in-field area is defined as the land intended for agricultural production activities, and the off-field area is defined as all area surrounding the area of application, including natural and semi-natural habitats, in particular hedgerows and woodland [12]. In some cases, specific protection goals determine the test location. Because of the wide spread cultivation of maize in the Netherlands, concern was raised that links between ecologically valuable regions (in Dutch: Ecologische Hoofdstructuur) would be affected by the use of glufosinate ammonium. The applicant therefore initiated a study into the effects of this herbicide on non-target plants in nearby verges when used in maize [35]. However, the environmental policy with respect to refining and differentiation of protection goals is still under development.
As was mentioned in Section 5.3.3.2, the degree of isolation of a site partly determines the potential for recovery after treatment. In this respect, separated aquatic mesocosm often reflect a *worst-case* situation in comparison with the connected ditches of the field situation. The same goes for a test with non-target arthropods where the different plots are separated by shields. On the other hand, re-colonisation of affected species may be reduced in practice because of the treatment of adjacent field in the same period or because the distance to reach the site cannot be covered between two applications.

5.3.4.6 History of the test system

An applicant may choose to perform a field test on an existing (agricultural) site instead of using an experimentally constructed mesocosm. This is often the case for non-target arthropods and earthworms. In this case, it should be clear that the populations were not previously exposed to the test substance or structure analogues and that no adaptation or resistance had been developed.

5.3.4.7 Crop and crop stage

This item is particularly relevant for the assessment of effects on non-target species within a crop. The crop type can determine the species that live in the area. EPPO and SETAC/ESCORT give guidance on the relevant indicator species for different crop types [12,13,15]The crop stage determines the interception of the applied dose and thus is a determining factor for the exposure of ground-dwelling arthropods. The test should therefore preferably be performed on the intended crop, but other crops can be used when it can be made clear that the test system covers the exposure and the species groups that are expected

in the field. An exception must be made for earthworms. The condition for this test that abundance in the control must at least be 100 individuals/m^2 is often not met on arable land because of the tillage and grassland may be used as an alternative [21]. As *Allolobophora chlorotica*, *A. longa* and *A. caliginosa* are the species encountered most in arable land, these species must at least be present in a grassland test.

5.4 Risk assessment on the basis of the $NOEC_{FIELD}$

After a proper evaluation of all submitted first-tier and additional higher-tier test(s), the Environmental Risk Assessment (ERA) is performed. It should be noted that sometimes the ERA can only be performed for part of the proposed uses because based on the similarity aspects, the supplied field data are not relevant to all different use patterns. In this case, additional field tests may be requested that do cover the appropriate use.
The ERA focuses on the question: *How likely is it that the $NOEC_{FIELD}$, as derived from field test(s) will be exceeded under the proposed conditions of use?*
This question addresses the comparison of the $NOEC_{FIELD}$ with the actual *exposure* under the conditions of use. In some cases, the $NOEC_{FIELD}$ will be determined in a test that was performed under the conditions of actual use with respect to the dimensions of the system and the way of application. This is for instance the case with arthropod and earthworm field tests. The result of such tests may be given as a dosage of the product in $kg.ha^{-1}$. This $NOEC_{FIELD}$ may then be compared directly with the intended field application rate, provided that the other similarity aspects that determine fate and behaviour are taken into account (see Section 5.3.4). In aquatic tests, the test substance may also be added by overspraying the mesocosms. In this case, however, the applied dose is often higher than the intended field dose to ensure that the target concentration in the water column is reached within a reasonable time span. In this case it is more relevant to refer to initial concentrations in the column for derivation of the $NOEC_{FIELD}$. When the $NOEC_{FIELD}$ is given as a concentration in soil or water, a comparison should be made with the modelled PEC. In this case it should be clear, that the potential risks that were identified in the first-tier and that triggered the performance of a field trial, were not concerned with the predicted concentration. When effects occur at the intended field rate, or when the $NOEC_{FIELD}$ is lower than the PEC, it cannot be excluded that under field conditions adverse effects will occur.

> If both the first- and the second-tier ERA (*i.e.* based on a reliable and useful field test) indicate adverse effects for (part of) the proposed applications, a risk cannot be excluded.

More often it will happen that in a second-tier field test no effects are observed at an application rate that indicated risks in the firs-tier assessment. One of the likely factors to cause this difference is exposure. The first-tier assessment is based on data from relatively short-term standard laboratory tests. These tests aim at maximum and continuous exposure of the organisms, whereas in the field test sorption and degradation lower exposure. Furthermore, because the distribution of the substance in the test system will be inhomogeneous, organisms have the possibility to avoid the most contaminated sites. From a comparison of laboratory and field experiments with chlorpyrifos [45] it appeared that the short term effects observed in field tests can be accurately predicted by laboratory data, provided that the actual exposure in the field tests is known. Although the models used to estimate the PEC do take many factors into account, it is likely that long-term estimates of the exposure concentration deviate from the actual concentration encountered in the field. A second important factor that differs between the first- and second-tier assessment is of course that the latter is based on systems with a higher degree of biological complexity, which may introduce resilience in the ecosystem that is not part of the single-species toxicity tests. It is

advised that the possible reasons for the difference between the first- and second-tier risk assessment are identified.

The final decision as to whether an ERA on the basis of a reliable and useful field test overrules the first-tier ERA, should be made in the light of all available information, including (higher-tier) laboratory data. This decision should be documented.

To conclude this Section, the route for the use of ecotoxicological field tests is schematically given in Figure 5.3.

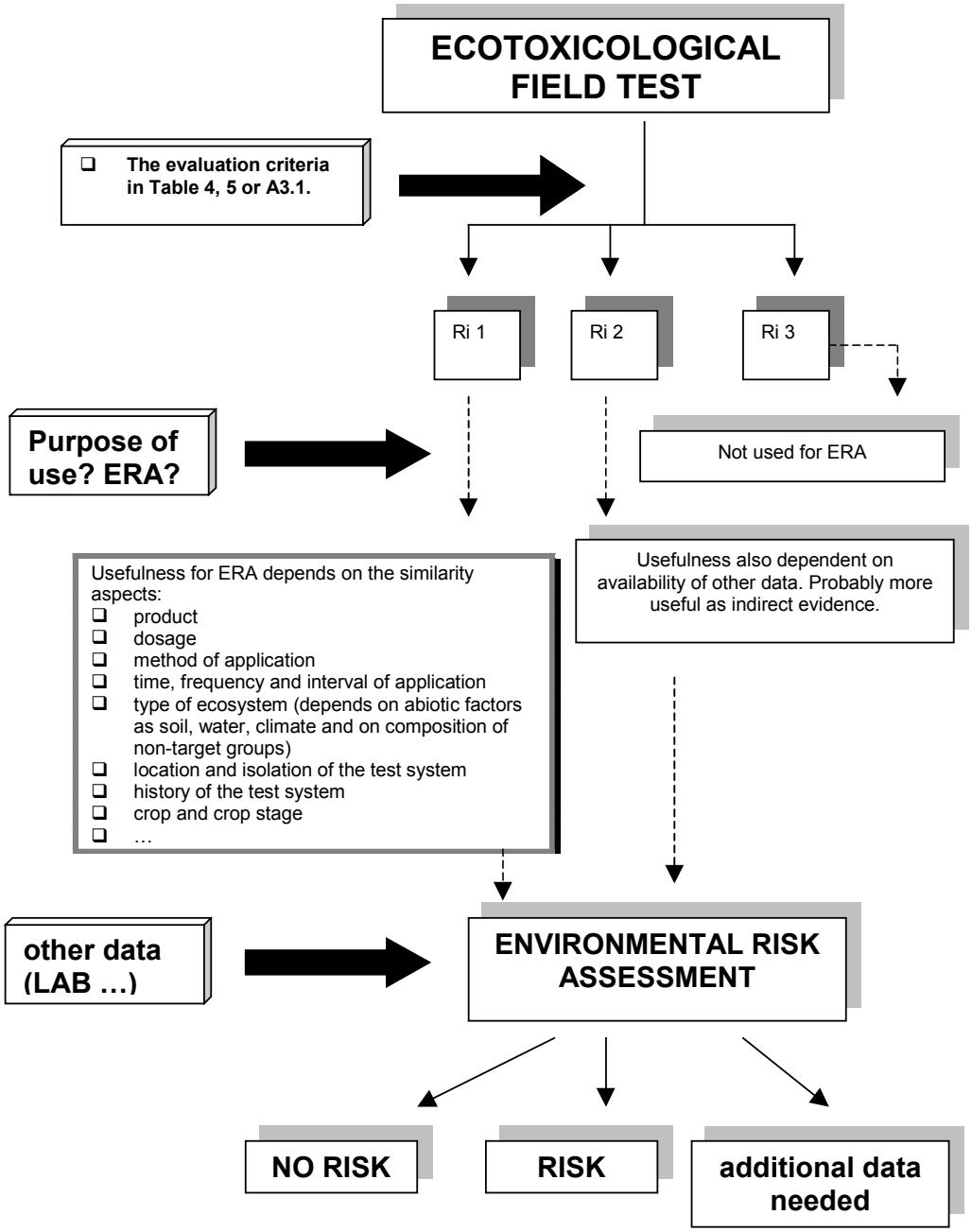


Figure 5.3. The process of environmental risk assessment and the position of field studies.

5.5 Concluding remarks, working agreements and recommendations

5.5.1 Concluding remarks

Guidance for interpretation and evaluation of field tests is limited. This is partly due to a lack of experience with field test systems other than freshwater micro- and mesocosms. It is also due to the complexity of such systems, and to the *ad hoc* character as they may specifically focus on the type of use and conditions that are under registration, when required for higher-tier risk assessment. Various sources confirm that there is a high need for the development of field test evaluation tools. The present factsheet aims at integrating expertise by giving as much guidance as possible in interpreting and evaluating field tests. Guidance is given to judge the scientific reliability of field tests and their usefulness for the registration of pesticides.

Whether a single-dose or a multi-dose design should be preferred depends on the questions that should be answered. If the design should allow a higher-tier risk assessment, then the assessment requires a proper risk characterisation based on a statistically valid $NOEC_{FIELD}$. This approach requires a multi-dose design. If the test should only demonstrate that recommended dosages should not affect the aquatic wildlife, then a single-dose design may be sufficient. Such a design may be useful when the dosage is *e.g.* 10 times higher than the highest recommended dosage and no adverse effects are found. In case of the occurrence of adverse effects, however, test refinement may be necessary, as a 10 times higher dosage is actually an unrealistic *worst-case*, and therefore not reflecting a realistic agricultural practice. These examples show that different rationales may trigger different test designs.

Field tests may represent a more realistic exposure scenario, but the effects are generally more difficult to interpret than in lower-tier tests. Therefore, field tests do not necessarily give more or better results than laboratory tests. In view of this, the prominent role of field tests in current European legislation to substitute first-tier test data may be an overrating. They should not be preferred simply because they claim to better reflect the circumstances under natural or agricultural conditions.

The guidance in this factsheet refers primarily to freshwater micro- and mesocosms, as most experimental expertise is with these field tests. However, it is recommended to update and integrate the expertise in other areas as well. Especially in the case of non-target arthropods this is deemed possible [7]. The expertise in terrestrial field systems is *e.g.* rapidly expanding. There are, however, still various areas with data gaps. Some examples [41]:

- there are almost no data available on the recovery of species with a relatively long life-cycle after the application of insecticides;
- there are almost no data on the ecotoxicological effects of pesticides in flowing water;
- ecotoxicological field data seems to be available only for a limited number of pesticides, as organophosphate esters and pyrethroids; there are not much field test data available for fungicides in particular.

5.5.2 Working agreements

1. Ri3 tests are not used for any purpose. Both Ri1 and Ri2 tests can be used for risk assessment. It depends on the overall data availability whether only Ri1 of both Ri1 and Ri2 tests are used.
2. A test with one concentration that does not result in effects, will only be used for risk assessment when a “positive” control is included, either by application of a very high

concentration of the test compound, or by using a toxic reference compound with known activity. The result is expressed as NOEC ≥ test concentration.

3. Where possible, the data are used to establish a concentration response curve and a comparison between EC_{50} and NOEC is made. If NOEC > EC_{50}, the result will not be used for risk assessment. For NOECs between EC_{10} and EC_{50}, a decision is made as to whether the tests provides sufficient evidence that there are "no unacceptable effects under field conditions".
4. If multivariate analysis is performed by the applicant, a specialist will be consulted to evaluate the reliability of the methods and results. If no multivariate analysis is performed, a specialist will be consulted to judge whether the data are suitable for such an analysis. In this case, the regulatory authority will be advised to have such an analysis performed, either by the applicant or by an independent specialist.
5. Where possible, the effect classification for field tests as listed in the STOWA-reports is used to derive a NOEC. EACs are not automatically regarded as NOECs.
6. Reliable field tests (Ri1 or 2) are only used for risk assessment when their usefulness in terms of similarity with the situation that is to be assessed is sufficiently demonstrated.
7. The final decision as to whether an ERA on the basis of a reliable and useful field test overrules the first-tier ERA, is made in the light of all available information, including (higher-tier) laboratory data. This decision is documented.

5.5.3 Recommendations

1. It is recommended to take the various statistical options into account prior to the actual field test performance and to weigh the expected scientific relevance of the data against the costs. Where possible, the following questions should be addressed
   a) one-dose *versus* multiple-dose design,
   b) using an experimental protocol, how large is the probability P (=1-β) that a difference of size δ will be detected,
   c) how many replicates are needed to be able to observe a difference of δ with probability P,
   d) when is a difference of size δ ecologically relevant [3,34])?
2. It is recommended that industries consult the regulatory authorities as early as possible about their plans how and where to perform field tests, when required. A proper communication in this respect will be fruitful to both parties. When industries think that they have already field test data available, that can fulfil the requirements of the regulatory authorities, it is recommended to have a pre-consultation to discuss such items. This factsheet may be a facilitating tool for such (pre)consultations.
3. Most expertise and experience in field testing has been gained with freshwater model ecosystems (micro- and mesocosms) and with honeybees and non-target arthropods. Therefore it is recommended to follow closely the developments in field testing with other non-target groups and to integrate these developments into the Summary Tables.

References

1. EU. 1997. Doc. 397L0057. Council Directive 97/57/EC of 22 September 1997 establishing Annex VI to Directive 91/414/EEC concerning the placing of plant protection products on the market. Official Journal L265: 0087-0109.
2. Campbell PJ, Arnold DJS, Brock TCM, Grandy NJ, Heger W, Heimbach F, Maund SJ, Streloke M, eds. 1999. Guidance document on higher-tier aquatic risk assessment for pesticides (HARAP). From the SETAC-Europe/OECD/EC Workshop, held in Lacanau Océan, France, 19-22 April 1998. Brussels, Belgium: SETAC-Europe. 179 pp.
3. Health Council of the Netherlands. 2000. Field research for the authorisation of pesticides. Den Haag, The Netherlands: HCN. Report no. 2000/07 (in Dutch, with English summary).
4. Hill IR, Heimbach F, Leeuwangh P, Matthiessen P, eds. 1994. Freshwater field tests for hazard assessment of chemicals. Ann Arbor, USA: Lewis Publishers.
5. Sijm D, Luttik R. 1999. Gaps in the risk assessment. Contacts outside RIVM. Bilthoven, The Netherlands: National Institute of Public Health and the Environment. Report no. CSR 06836a00 (in Dutch).
6. Van Dijk HFG, Brussaard L, Stein A, Baerselman F, De Heer H, Brock TCM, Van Donk E, Vet LEM, Van der Gaag MA, Van Gestel CAM, Van der Hoeven N, De Jong FMW, Van der Linden AMA, Van Noort PCM, Oomen PA, Van Vliet PJM. 2000. Field research for the authorisation of pesticides. Ecotoxicology 9: 377-381.
7. De Jong FMW, De Knecht J, Montforts M, Smit CE. 2001. Report of the workshop Problems and solutions concerning higher tier studies with pesticides in the framework of the “unless”-condition. Bilthoven, The Netherlands: National Institute for Public Health and the Environment. Report no. CSR 08640a00. 19 pp.
8. Mensink BJWG. 2002. How to use ecotoxicological field test data for derivation of harmonised maximum permissible concentrations (MPCs). (An addendum to RIVM report no. 601501012). Bilthoven, The Netherlands: RIVM/CSR. Report no. 08659 (draft). 26 pp.
9. De Jong FMW. 2001. Terrestrial field trials for side-effects of pesticides. Thesis Leiden University.
10. Giddings J, Heger W, Brock TCM, Heimbach F, Maund SJ, Norman S, Ratte HT, Schäfers C, Streloke M. 2001. Proceedings of the CLASSIC-workshop (Community Level Aquatic System Studies Interpretation Criteria). Brussels, Belgium: SETAC-Europe.
11. EPPO. 2001. Soil organisms. Draft version 31-01-2001. In: EPPO, Decision making scheme for the environmental risk assessment of plant protection products, EPPO. Chapter 7.
12. EPPO. 2000. Decision making scheme for non-target terrestrial arthropods. Draft version 5/2000. In: EPPO, Decision making scheme for the environmental risk assessment of plant protection products, EPPO. Chapter 9.
13. Candolfi MP, Barrett KL, Campbell P, Forster R, Grandy N, Huet M-C, Lewis G, Oomen PA, Schmuck R, Vogt H, eds. 2001. Guidance document on regulatory testing and risk assessment procedures for plant protection products with non-target arthropods. Report of the SETAC/ESCORT2 Workshop, 21-23 March 2000, Wageningen. Brussels, Belgium: SETAC-Europe.
14. Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs MA, Reber B, Schmuck R, Vogt H, eds. 2000. Guidelines to evaluate side-effects of plant protection products to non-target arthropods. IOBC, BART and EPPO Joint Initiative. Gent, Belgium: IOBC/WPRS. 158 pp.
15. Barrett KL, Grandy N, Harrison EG, Hassan S, Oomen PA. 1994. Guidance document on regulatory testing procedures for pesticides with non-target arthropods. Report of the SETAC/ESCORT Workshop, 28-30 March 1994, Wageningen. Brussels, Belgium: SETAC-Europe. 51 pp.
16. Bestrijdingsmiddelenwet. 2001. Code 34. Band 1, 2 en 3. Lelystad, The Netherlands: Koninklijke Vermande Publishers. (in Dutch).
17. Arnold D, Hill I, Matthiesen P, Stephenson R, eds. 1991. Guidance document on testing procedures for pesticides in freshwater static mesocosms - from the workshop “A meeting of experts on guidelines for static field mesocosm tests, held at Monks Wood Experimental Station, Abbotts Ripton, Huntingdon, UK. 3-4 July 1991. Brussels, Belgium: SETAC-Europe. 46 pp.
18. EPPO. 1992. EPPO Guideline 170. Guideline on test methods for evaluating side-effects of plant

protection products on honeybees. EPPO Bulletin 22: 203-215.

19. Felder DP, D'Surney SJ, Rodgers Jr. JH, Deardorff TL. 1998. A comprehensive environmental assessment of a receiving aquatic system near an unbleached kraft mill. Ecotoxicology 7: 313-324.
20. Jak RG. 1997. Toxicant-induced changes in zooplankton communities and consequences for phytoplankton development. Thesis Vrije Universiteit Amsterdam.
21. Greig-Smith PW, Becker H, Edwards PJ, Heimbach F, eds. 1992. Ecotoxicology of earthworms. Hants, United Kingdom: Intercept Ltd.
22. ISO. 1999. ISO 11268-3. Soil quality - effects of pollutants on earthworms. Part 3: Guidance on the determination of effects in field situations.
23. EPPO. 1993. Decision making scheme for the environmental risk assessment of plant protection products. EPPO Bulletin 23: 1-165.
24. Candolfi MP, Bigler F, Campbell P, Heimbach U, Schmuck R, Angeli G, Bakker F, Brown K, Carli G, Dinter A, Forti D, Forster R, Gathman A, Hassan S, Mead-Briggs M, Melandri M, Neumann P, Pasqualini E, Powell W, Reboulet J-N, Romijn K, Sechser B, Thieme Th, Ufer A, Vergnet Ch, Vogt H. 2000. Principles for regulatory testing and interpretation of semifield and field studies with non-target arthropods. J Pest Sci 73: 141-147.
25. AEDG. 1992. Improving aquatic Risk Assessment under FIFRA. Report of the Avian Effects Dialogue Group, 1990-1991. The RESOLVE program of the World Wildlife Fund.
26. Fairbrother A, Landis WG, Dominquez S, Shiroyama T, Buchholz P, Roze MJ, Matthews GB. 1998. A novel nonmetric multivariate approach to the evaluation of biomarkers in terrestrial field studies. Ecotoxicology 7: 1-10.
27. RIVM, VROM, VWS. 1999. Uniform System for the Evaluation of Substances 3.0 (USES 3.0). The Netherlands: National Institute of Public Health and the Environment (RIVM), Ministry of Housing, Spatial Planning and the Environment (VROM), Ministry of Health, Welfare and Sport (VWS). Report no. 601450 004.
28. Van der Hoeven N. 2000. Investigation into the number of replicas needed for the desired power in a field experiment with 4 treatments and a control. Leiden, The Netherlands: ECOSTAT. Report no. 00/002.
29. RIVM/CSR. 2001. Manual for the evaluation of the environmental aspects of substances. Bilthoven, The Netherlands: National Institute for Public Health and the Environment, SOP CSR/236 rev. 0 (confidential).
30. TOXSTAT® [computer program]. Version 3.3. Wyoming: Fish Physiology and Toxicology Laboratory, Department of Zoology and Physiology, University of Wyoming, Jan 1991.
31. Haanstra L, Doelman P, Oude Voshaar JH. 1985. The use of sigmoidal dose respons curves in soil ecotoxicological research. Plant Soil 84: 293-297.
32. Van Brummelen TC, Van Gestel CAM, Verweij RA. 1996. Long-term toxicity of five polycyclic aromatic hydrocarbons to the terrestrial isopods *Oniscus asellus* and *Porcellio scaber*. Environ Toxicol Chem 15: 1199-1210.
33. Van Straalen NM, Verkleij JAC, eds. 1993. Handboek oecotoxicologie. Amsterdam, The Netherlands: VU Uitgeverij. Chapter III-1 (in Dutch).
34. Van der Hoeven N. 1998. Power analysis for the NOEC: what is the probability of detecting small toxic effects on three different species using the appropriate standardized test protocols? Ecotoxicology 7: 355-361.
35. De Snoo GR, De Jong FMW, Van der Poll RJ, Van der Linden MGAM. 2001. Effects of glufosinate-ammonium on off-crop vegetation: interim results. Proceedings of the International Symposium on Crop Protection, Gent, Belgium:
36. Weyers A, Schuphan I. 1998. Variation of effect endpoint parameters in a terrestrial model ecosystem. Ecotoxicology 7: 335-341.
37. Van den Brink PJ. 1999. Ecological and statistical evaluation of effects of pesticides in freshwater model ecosystems. Thesis Landbouwuniversiteit Wageningen.
38. Smit CE, Schouten AJ, Van den Brink PJ, Van Esbroek MLP, Posthuma L. 2002. Effects of zinc contamination on a natural nematode community in outdoor soil mesocosms. Arch Environ Contam Toxicol 42: 205-216.
39. SANCO. 2001. Working Document. Guidance document on aquatic ecotoxicology in the frame of the Directive 91/414. Brussels, Belgium: EU (DG Health and Consumer Protection). Report no.

8075/VI/97 (revision 8). 41 pp.

40. Lahr J, Van den Brink PJ, Brock TCM. 1998. Ecological risks of pesticides in freshwater ecosystems. Part 1: herbicides. Utrecht, The Netherlands: STOWA. Report no. 98.30 (in Dutch, with English summary). 55 pp. with Annexes.
41. Van Wijngaarden RPA, Van der Geest GJ, Brock TCM. 1998. Ecological risks of pesticides in freshwater ecosystems. Part 2: insecticides. Ùtrecht, The Netherlands: STOWA. Report no. 98.31 (in Dutch, with English summary). 58 pp. with Annexes.
42. Brock T, Crum S, Leeuwangh P, Lucassen W, Van Wijngaarden R. 1993. Modelecosystemen: brug naar het veld. Landschap 10: 23-36. (in Dutch).
43. De Jong FMW, Bergema WF. 1994. Field bioassays for side-effects of pesticides. Leiden, The Netherlands: Centre of Environmental Sciences. Report no. CML 112.
44. De Jong FMW. 1995. Framework for field trials for side-effects of pesticides. Leiden, The Netherlands: Centre of Environmental Sciences. Report no. CML 117.
45. Van Wijngaarden RPA, Van den Brink PJ, Oude Voshaar JH, Leeuwangh P. 1996. Effects of the insectide Dursban®4E (a.i. chlorpyrifos) in outdoor experimental ditches. I. Comparison of short-term toxicity between laboratory and field. Environ Toxicol Chem 15: 1133-1142.
46. Crossland NO, LaPoint TW. 1992. Symposium on aquatic mesocosms in ecotoxicology. Environ Toxicol Chem 11.

Annex 1. Abbreviations & definitions

AChE	Acetyl cholinesterase
AEDG	Avian Effects Dialogue Group
BART	Beneficial Arthropod Regulatory Testing Group
CLASSIC:	Community Level Aquatic System Studies Interpretation Criteria. An initiative of SETAC-Europe in co-operation with the OECD, the European Commission, Umweltbundesamt, and Biologische Bundesanstalt für Land- und Forstwirtschaft [10]. This initiative is on formalising and harmonising the evaluation and interpretation of aquatic micro and mesocosms. The CLASSIC workshop was the follow-up of the HARAP workshop
EAC	Ecologically Acceptable Concentration. The concentration at or below which no ecologically adverse effects were observed in a particular study [2,10].
EPPO	European Plant Protection Organisation
ERA	Environmental Risk Assessment
EROD	7-ethoxy resorufin-O-deethylase. A phase I metabolising enzyme on the cytochrome P450 (CYP) 1A1. It is involved in biotransformation reactions, and its activity is measured with 7-ethoxy resorufin as the model substrate
ESCORT	European Standard Characteristics of beneficial Regulatory Testing. In co-operation with SETAC, two workshops on regulatory testing with beneficial arthropods were organised.
GAP	Good Agricultural Practice, *i.e.* the recommended agricultural practice as stated in the legal regulation and usage instructions, in Dutch: Wettelijk Gebruiksvoorschrift en GebruiksAanwijzingen (WG/GA).
HARAP	Higher-tier Aquatic Risk Assessment for Pesticides. A SETAC-Europe initiative on formalising higher-tiered tests, including field tests [2].
IOBC/WPRS	International Organisation for Biological and integrated Control of noxious animals and plants, West Palearctic Regional Section
macrocosm	Natural environment (definition by [46]).
mesocosm	Outdoor experimental tanks/ponds greater than 15 m^3 water volume or experimental streams greater than 15-m length (definition by [46], see also model ecosystem).
microcosm	Experimental tanks/ponds less than 15 m^3 water volume or experimental streams less than 15-m length (definition by [46], see also model ecosystem).
model ecosystem	Man-made study system containing associated organism and abiotic components that is large enough to be representative of a natural ecosystem, yet small enough to experimentally manipulated. There is some subjective differentiation between larger, outdoor model ecosystems (mesocosms) and smaller, generally indoor model ecosystems (microcosms) [2].
NCAA	Nonmetric Clustering and Association Analysis
$NOEC_{FIELD}$:	A No Observed Effect Concentration in a field test under semi- or realistic conditions, respecting one or more species or functional endpoints.
PEC	Predicted Environmental Concentration
PRC	Principal Response Curve

Ri	Reliability Index (pl.: Indices). An index that can be 1, 2 or 3 for scientifically reliable, less reliable or unreliable (field) tests, respectively. Together with the usefulness of a (field) test for a particular purpose, the Ri denotes the test quality
SAVE	An initiative of RIVM, Alterra, RIZA & TNO on formalising ecotoxicological higher-tier tests including field tests
SETAC	Society for Environmental Toxicology And Chemistry
Summary Table:	A checklist used at CSR to evaluate experiments for regulatory reasons (e.g. the registration of pesticides). A Summary Table comprises a wide array of test items that may influence the scientific reliability of the test under evaluation. It also focuses on the test methodology and description, the results and on the additional remarks of the evaluator
Tier	Reference to the different levels of data requirements and therefore to the specificity of the evaluation. The first-tier refers to the basic test data that should be submitted to the regulatory authorities for chemical registration. They generally consist of a particular set of relatively simple laboratory test data. A second or higher-tier refers to the test data that are required after the first-tier evaluation had indicated adverse environmental effects. As the first-tier risk assessment is generally based on a worst-case scenario, *e.g.* respecting the exposure, second or higher-tier test data may be required to assess the exposure more realistically. Field tests may be used for the second-tier evaluation in this way.
TME	Terrestrial Model Ecosystem
TOXSTAT	A computer program for the use of statistics in (eco)toxicology [30]

Annex 2. A summarised overview of available guidance and literature concerning field trials.

Below an overview is given of the state of the art concerning available guidelines and other relevant literature concerning performance and assessment of higher-tier studies. A distinction is made in guidance or literature for the aquatic (§ A1.1) and for the terrestrial environment (§ A1.2). For every document referred, a brief indication of the content is given, and for every group of organisms it is briefly indicated what the Dutch Board for the Authorisation of Pesticides prescribes in the Handbook for Authorisation of Pesticides. In § A1.3 the references are given. This Annex is not meant to be exhaustive, and may overlap with information that is already given in the main text of this factsheet.

A1.1 Aquatic environment

A1.1.1 General guidance for aquatic higher-tier studies

Guidance Document on Higher-tier Aquatic Risk Assessment for Pesticides (HARAP), Campbell, et al., 1998
For the aquatic environment the HARAP document can be seen as a co-ordinating document. The document describes a procedure, which distinguishes four possible types of higher-tier studies, applicable after exceeding a trigger for the 'unless'-statement:
- Interrogation use of core (tier 1) data, with the aim to design hypotheses for further testing
- Further single species studies (additional species, modified exposure, population-level studies)
- Indoor multi species tests, microcosm
- Field studies

A procedure is described, and a large number of case studies are handled. It is clearly indicated for what type of question a specific type of test is suitable and the implementation in the assessment procedure is giver per test-type. A number of more general criteria for evaluating of test reports are given. For the assessment of and acceptable effect the concept of an Ecological Acceptable Concentration (EAC) is introduced. Experts should establish the height of this concentration, however.
Concerning guidelines for performance of higher-tier studies, it is stated that the presently available guidelines are sufficient, and that experts should be used at crucial points.

Boxall et al., 2001. Higher-tier laboratory aquatic toxicity testing.
In this British study higher-tier laboratory studies are handled, and this study could be seen as a further refinement of the HARAP document for this type of studies. A number of test-types are distinguished:
- realistic exposure scenario’s
- additional species
- recovery of organisms
- recovery of the system
- sensitive life-stages
- population level studies
- indoor multi-species studies

For these types the possibilities and the place in the procedure are indicated and the handling of the results is described. It is described in detail which test is suited for which problem, and how the tests should be designed to answer the specific question, including statistics. The back coupling to the first-tier safety factors is handled as well, among others as a table.

Hill, I.R., Heimbach, F., Leeuwang, P. and Matthiessen, P. (eds.), 1994. Freshwater field tests for hazard assessment of chemicals.
Extended report of the European Workshop on freshwater field tests (EWOFFT), the subjects:
Endpoints in aquatic field tests
Methodologies of aquatic field tests
Case studies of aquatic field tests
Extrapolation and hazard assessment
Modelling and other issues
The report contains a lot of relevant information with a lot of case studies.

Report of workshop Interpretation of aquatic (semi-)fieldstudies for the approval of pesticides, 15-11-1996 (Brock)
A number of aspects concerning endpoints and interpretation are discussed. This discussion is more or less overruled by the HARAP and CLASSIC workshops.

A1.1.1.1 Guidance for specific test types

Postuma, L. et al., 2002. Species sensitivity distributions in ecotoxicology.
More information about the use of data for additional species to establish the level at which for instance 95% of the species is protected.

OECD monograph 59.
Summary to be included in the next version of this factsheet.

SETAC, 1994. Pesticide risk and mitigation: Final report of the aquatic risk assessment and mitigation dialog group. Pensacola, Florida, US.
Summary to be included in the next version of this factsheet.

A1.1.1.2 Microcosms

Crossland, N.O. et al., (eds.) 1992. European Workshop on Freshwater Field Tests (EWOFFT). Summary and recommendations. Potsdam, June 25-26, 1992.
Several types of systems are defined:

Outdoor microcosm:	Experimental tanks/ponds less than 15 m^3 water volume or experimental streams less than 15-m length.
Mesocosm:	Outdoor experimental tanks/ponds greater than 15 m^3 water volume or experimental streams greater than 15-m length.
Macrocosm:	Natural environment.

SETAC/RESOLVE, 1991. Workshop on aquatic microcosms for ecological assessment of pesticides. Report from a meeting held in Wintergreen, Virginia, USA.
Summary to be included in the next version of this factsheet.

A1.1.1.3 Mesocosms

Crossland, N.O. and T.W. La Point (eds.). 1992. Symposium on aquatic mesocosms in ecotoxicology. Env. Toxicol. Chem. 11: Issue 1.
Summary to be included in the next version of this factsheet.

OECD, 1996. Draft proposal for a guidance document Freshwater Lentic Field Tests. OECD guidelines for testing of chemicals.
Provides guidance concerning aquatic mesocosm studies, including design, testing, interpretation and reporting.

Graney, R.L., Kennedy, J.H., and Rodgers J.H. (eds.) 1994. Aquatic mesocosm studies in ecological risk assessment. Lewis Publishers, London, UK.
Summary to be included in the next version of this factsheet.

Arnold, D. et al. 1991. Guidance document on testing procedures for pesticides in freshwater static mesocosms.
Mesocosm are defined in detail, including the content. It is indicated what should be measured before and during the experiment, how samples should be obtained and how the data should be processed to obtain conclusions from the results. Dosing is handled as well. Endpoints for shorter and longer-term studies are handled as well. General starting points for interpretation are given, but concrete criteria for interpretation or acceptability of effects are lacking.

Touart, L.W. 1988. Aquatic mesocosm test to support pesticide registrations. Hazard evaluation division technical guidance document. US-EPA, Report no. USEPA/540/09-88-035, Washington DC, 35 pp.
Summary to be included in the next version of this factsheet.

A1.1.1.4 Field

Graney, R.L., Giedy, J.P. and Clark, J.R., 1995. Field studies. In: G.M. Rand (ed.), Fundamentals of aquatic toxicology. Taylor and Francis Ltd, Londen, UK: pp. 257-306.
Gives an extensive overview, including examples of different types of field studies. Not specially aimed at pesticides. Useful as a reference for other types of field studies than handled in the other documents.

RIZA/BKH 1996. Criteria for assessment of aquatic (semi-)field studies of pesticides and other chemicals.
Describes a number of criteria for aquatic field studies as mentioned in the title. Comparable with Annex 3 of this factsheet. Useful at points where criteria are given in more detail.

De Jong, F.M.W. 2000. Usefulness of a duckweed-test as indicator for herbicides in surface water.
Describes the use of a field test with duckweed to trace effects of herbicides in surface water. It is concluded that the test is too insensitive to trace effects in the field. Useful to compare field tests with aquatic plants. An overview of laboratory toxicity data for duckweed is given.

A1.2 Terrestrial environment

A1.2.1 General guidance for terrestrial higher-tier studies

In application form A of CTB, part H: "toxicity for organisms living in the natural environment" data are required for birds, water organisms, non-target arthropods divided into honeybees and other non-target arthropods, soil organisms divided into nitrification (soil micro-organisms) and earthworms, influence on wastewater cleaning and bio-accumulation in water organisms. Further data can be required, for instance concerning secondary poisoning. Higher-tier studies are indicated in the case a compound is "not permissible, unless …." in the case of birds, water organism, non-target arthropods, soil micro-organisms and earthworms.

In other EU-countries in the past guidelines for higher-tier studies have been developed for other organism groups, such as ground beetles and other guidelines are under development such as in the case of vascular plants. These data are, as far as they are available, handled below as well.

EPPO, 1993. Decision-making scheme for the environmental risk assessment of plant protection products. EPPO Bulletin. 23/24.
The EPPO Decision –making schemes give some guidance for the selection of the appropriate test for a certain problem (Chapter 2). Further more per subject decision-making schemes are divided in soil, groundwater and surface water and effect on aquatic organisms, soil micro-organisms, earthworms, natural enemy arthropods, honeybees and terrestrial vertebrates. Guidance for the actual performance of the tests is not incorporated in the decision-making schemes. The specific decision schemes are handled with the specific organism groups further on.

EU Working document Guidance Document on Terrestrial Ecotoxicology.
In the document it is stated that it is not clear what effects are deemed acceptable, and for this aspect guidance and agreement exist, but that the general scope is on the protection of populations rather than individuals, with the goal that non-target populations can survive on the long-term. Structural and functional parameters are deemed equally important. NOEC is replaced by EC_x which has consequences for the statistics (see also Chapman *et al.,* 1995).

A.1.2.2 Guidance for specific test types

A1.2.2.1 Birds and Mammals

The HTB only handles risk for birds, although the uniform principles clearly mention other vertebrates. A distinction is made in different exposure routes: by food (granules or treated seeds), bate, crops (also insects), drinking water and fish and worms as food. Indirect effects (habitat destruction, lack of food) are mentioned, but not incorporated. No guidance for higher-tier studies is given.

EPPO, 1993. Decision-making scheme for the environmental risk assessment of plant protection products. Chapter 11. Terrestrial vertebrates. EPPO bulletin 24: 37-87.
Scheme aims at effects on individuals, because criteria for assessing population effects are lacking. For laboratory studies it is indicated in detail how No Observed Effect Levels

(NOELs) and Doses with 50% Effect (ED_{50}'s) should be calculated, for a large number of exposure routes. Field trials can be used to confirm the predicted effect. For conducting field trials it is referred to experts and general literature (e.g. Sommerville & Walker). Eventually useful to estimate the exposure via different routes.

RESOLVE, 1994. Assessing pesticide impacts on birds. Final report of the avian effects dialogue group, 1988-1993. RESOLVE, Washington, DC.
In 1992 a change in the position of field trials have taken place: before 1992 field trials were part of a tiered approach, after 1992 field trials are recommended only for specific conditions, such as a new mode of action. The document extendedly describes the set-up, choice of test species, post-registration testing and additional information and even handles the indirect effects. Concerning the acceptability of effects a starting point is that repeated mortality of individuals without compensating measures is in itself unacceptable. It cannot be expected that population effects are being studied within the framework of an approval. Both effects (individual and population) however are deemed relevant and sorrowful. An overview of relevant ecological endpoints is given. In general the report contains a lot of information that can be of importance while assessing a field trial with birds.

In the United States legislation a balance of risk and benefits is made. Therefore a field trial can have two aims: proving the absence of effects in practice, or quantifying the expected effects. Here the aspect how to prove the absence of effects is handled, also solutions for the statistical side of this problem are given. For instance, the number of replications and trial fields is handled, and other solutions are discussed. Number of statistical aspects is handled in detail.

EU Working document Guidance Document on Terrestrial Ecotoxicology.
A working document aimed to obtain more uniformity and transparency for the assessment of terrestrial ecology parts of the Annex II, III, and IV. For endpoints it is referred at Suter, 1993.

Jong, F.M.W. de et al., 1990. Field trials for the-side effects of pesticides.
Gives guidance for field trials with birds, based on the UK guideline, which is summarised as well. Gives an overview of field studies with birds before 1990. See also De Snoo & Canters, 1990.

De Snoo, G.R. and Canters, K.J., 1990. Side-effects of pesticides on terrestrial vertebrates.
The report gives an extended overview of incidents and field studies with terrestrial vertebrates and pesticides.

Somerville, L. and Walker, C.H.,, 1990
In this book a number of aspects are handled, concerning field testing with birds and mammals. A number of aspects are discussed, including examples, *e.g.* objectives, distribution and accumulation, modelling exposure, and all kind of effect parameters and studies, including monitoring. A very useful background document.

A1.2.2.2 Honeybees and bumble bees

The HTB refers to EPPO guideline 170. For decision-making follows the EPPO decision scheme as well. In practice it is not always possible to make a statistically sound lay out. By means of exposure studies, however (pollen-analyses, feeding behaviour) such trials can give

interpretable results. Discussion is going on about the representativeness for bumblebees, for this aspect a recent version of the HTB should be consulted.

EPPO guidelines for Honeybees (170, 1997)
Describes the design of cage, field and tunnel studies, including the use of an untreated control (or with a harmless compound) and a positive control (harmful reference).

EPPO decision-making scheme, Chapter 10 honeybees (1998)
Decision scheme for authorisation, basis for the Dutch procedure.

Jong, F.M.W. de et al., 1990. Field trials for the side effects of pesticides.
Gives guidance for honeybees based on the EPPO guideline, with slight modifications: fields of 3 ha instead of 1.5 ha and a higher frequency of observations. A summery of available guidance and field studies in 1990 is given.

A1.2.2.3 Earthworms

The HTB refers to ISO guideline 11268-3, and for assessment of the results expert judgement is mentioned.

ISO 11268. Soil quality – Effects of pollutants on earthworms – Part 3. Guidance on the determination of effects in field situations.
The guideline describes design, performance and reporting of the test in detail. When assessing a field trial conducted according to this guideline, the guideline should be used as a checklist. Indications for the assessment of results are lacking.

EPPO guideline for soil organisms 2001. Environmental risk assessment for plant protection products. Chapter 7.2.1 Soil organisms.
For performing of the trial it is referred to the ISO standard. Indications for assessment of results are given.

Jong, F.M.W. de et al., 1990. Field trials for the side effects of pesticides.
A guideline for earthworms based on the BBA guidance, adapted according to the UK guideline. In the mean time superseded by the ISO guideline, but useful information is given concerning available guidelines and information about field studies before 1990.

A1.2.2.4 Other soil organisms

EPPO guideline for soil organisms 2001. Environmental risk assessment for plant protection products. Chapter 7.2.1 Soil organisms.
For other soil organisms than earthworms the non-target arthropods are used as triggers with HQ=9 (see below), except for soil incorporated products. In that circumstances testing on relevant soil dwelling species should be conducted instead. If continued and repeated exposure is expected, further single species testing with a relevant soil invertebrate (e.g. *Folsomia candida* or *Hypoaspis aculiefer*) is indicated. For higher-tier studies a number of considerations are given, but decisions about trial and assessing of trial results are still expert judgement.

A1.2.2.5 Decomposition

EPPO guideline for soil organisms 2001. Environmental risk assessment for plant protection products. Chapter 7.2.1 Soil organisms.
A relative empty guideline, indicating that decomposition is an important process and that methods are under development. An *ad hoc* group with members from industry, academia and regulatory bodies and chaired by the BBA has developed a draft guideline. A workshop on Effects of Plant Protection Products on Functional Endpoints in Soil (EPFES) is to be held in Lisbon, April 2002 with participants from industry, regulatory bodies, independent institutes and academia. Aim of the workshop is to produce a guideline that will published by SETAC-Europe.

A1.2.2.6 Non-target arthropods

SETAC, 1994. Barrett, K.L., Grandy, N., Harrison, E.G., Hassan, S. and Oomen, P.A. (eds.). Guidance document on regulatory testing procedures for pesticides with non-target arthropods. From the ESCORT Workshop, 28-30 march 1994, Wageningen.
The document provides a number of general starting points for semi-field and field testing. Most points have been filled in further in the more recent documents (see below).

EPPO, 1994. Decision-making scheme for the environmental risk assessment of plant protection products. Chapter 9. Arthropod natural enemies. EPPO bulletin 24: 17-35.
Tiered approach based on the risk for arthropod natural enemies. The indication of an intermediate or high risk forms the basis for further testing, from basic laboratory tests, to further laboratory tests, semi-field tests and field tests with natural populations. The benchmark for medium or high risk is 25 to 30% effect, and has to be made specific if more information becomes available.

EPPO guideline for non-target terrestrial arthropods (EPPO Standards Draft version 5/2000).
A new scheme for non-target arthropods is under development. A major difference is that a distinction is made between within-field and off-field. As indicator species *Aphidius rhopalosiphi* and *Typhlodromus pyri* are used. The between the assessment of within-field and off-field effects is expressed in calculation of the PEC. Within field the PEC = application rate x MAF (in which MAF is a correction factor for multiple application). Off-field the PEC = application rate x MAF x drift factor x uncertainty factor / vegetation dilution factor. An uncertainty factor of 10 is used based on a larger diversity of species outside the treated field, but a vegetation dilution factor is used at the same time. The drift factor is a measure for the drift at a certain distance. The benchmarks for medium or high risk are based on the Hazard Quotient HQ = PEC / LR50, in which LR50 is the dose of active substance per surface area in g product per ha that is estimated to result in 50% mortality. For *A. rhopalosiphi* the Benchmark is 8 and for *T. pyri* the benchmark is 12. If the benchmark is exceeded higher-tier studies could be performed. Categories for risk classification are given. In-field and off-field effects should be assessed separately. In the case of off-field risks, mitigating measures are possible, such as buffer zones, windbreaks or drift-reducing application techniques.

SETAC, 2001. Candolfi 2001 Guidance for non-target arthropods.
Summary to be included in the next version of the factsheet

Candolfi et al., 2000. Guidelines to evaluate side-effects of plant protection products to non-target arthropods.
Provides guidance for a number of laboratory and semi-field trials with the rove beetle *Aleochara bilineata,* the parasitic wasp *Aphidius rhopalosiphi*, larvae of *Chrysoperla carnea*, *Coccinella septempunctata,* the predatory bug *Orius laevigatus*, the spider *Oardisa*, the carabid beetle *Poecilus cupreus*, *Trichogramma cacoeciae* and on the predatory mite *Typhlodromus pyri*. As far as the test species are not incorporated in the first-tier assessment yet, they can be seen as additional data. For some of the species guidance is given for semi-field or field trials. Test methods are described in detail, including validity criteria and data analysis and reporting.

A1.3 References

A1.3.1 Aquatic

AEDG (1992). Improving aquatic Risk Assessment under FIFRA. Report of the aquatic effects dialogue group, 1990-1991. The RESOLVE program of the World Wildlife Fund.

Arnold D, Hill I, Matthiesen P, Stephenson R, eds. 1991. Guidance document on testing procedures for pesticides in freshwater static mesocosms - from the workshop "A meeting of experts on guidelines for static field mesocosm tests, held at Monks Wood Experimental Station, Abbotts Ripton, Huntingdon, UK. 3-4 July 1991. Brussels, Belgium: SETAC-Europe. 46 pp.

Boxall, A., Brown, C. & Barrett, K., 2001. Higher-tier laboratory aquatic toxicity testing. Cranfield Centre for EcoChemistry research report No. JF4317E for DETR, 70 pp.

Campbell, P.J. *et al.*, (eds), 1998. Guidance Document on Higher-tier Aquatic Risk Assessment for Pesticides (HARAP). From the SETAC-Europe/OECD/EC Workshop, held at Lacanau Océan, France, April 1998.

Crossland, N.O. and T.W. La Point (eds.). 1992. Symposium on aquatic mesocosms in ecotoxicology. *Env. Toxicol. Chem.* 11: Issue 1.

Crossland, N.O. *et al.,* (eds.) 1992. European Workshop on Freshwater Field Tests (EWOFFT). Summary and recommendations. Potsdam, June 25-26, 1992.

Graney, R.L., Kennedy, J.H., and Rodgers J.H. (eds.) 1994. Aquatic mesocosm studies in ecological risk assessment. Lewis Publishers, Londen, UK.

Graney, R.L., Giedy, J.P. and Clark, J.R., 1995. Field studies. In: G.M. Rand (ed.), Fundamentals of aquatic toxicology. Taylor and Francis Ltd, onden, UK: pp. 257-306.

Hill, I.R., Heimbach, F., Leeuwang, P. and Matthiessen, P. (eds.), 1994. Freshwater field tests for hazard assessment of chemicals. Lewis Publishers, Londen, UK.

Jong, F.M.W. de, Snoo, G.R. de, Canters, K.J. & Voet, E. van der, 1990. Field trials for the side-effects of pesticides. CML report 60, Leiden.

Jong, F.M.W. de, 2000. Usefulness of a duckweed-test as indicator for herbicides in surface water. Stowa-report 2000-32, Utrecht.

OECD, 1996. Draft proposal for a guidance document Freshwater Lentic Field Tests. OECD guidelines for testing of chemicals.

OECD, 1998. Report of the OECD Workshop on Statistical Analyses of Aquatic Toxicity Data, OECD series on testing and assessment No. 10, 1998.

Odum, E.P., 1984. The mesocosm. Bioscience 34: 558-562.

RIZA/BKH 1996. Criteria voor de beoordeling van aquatisch (semi-)veldonderzoek aan bestrijdingsmiddelen en andere chemische stoffen.

SETAC/RESOLVE, 1992. Workshop on aquatic microcosms for ecological assessment of pesticides. Report from a meeting held in Wnitergreen, Virginia, USA.

SETAC, 1994. Pesticide risk and mitigation: Final report of the aquatic risk assessment and mitigation dialog group. Pensacola, Florida, US.

Solomon, K.R. and Liber, K., 1988. Fate of pesticides in aquatic mesocosm studies – an overview of methodology. Brighton Crop Prot. Conf. Pests and Diseases: 139-148.

Touart, L.W. 1988. Aquatic mesocosm test to support pesticide registrations. Hazard evaluation division technical guidance document. US-EPA, Report no. USEPA/540/09-88-035, Washington DC, 35 pp.

Verslag workshop Interpretatie van aquatische (semi-)veldstudies bij de toelating van bestrijdingsmiddelen, 15-11-1996 (Brock)

A1.3.2 Terrestrial

Beelen, P. van, Luttik, R., Montforts, M.H.M.M., draft version. Criteria for field studies on earthworms and birds. Guidance in the registration process for pesticides and biocides.

Candolfi, M.P., Blümel, S., Forster, R., Bakker, F.M., Grim, C., Hassan, S.A., Heimbach, U., Mead-Briggs, M.A., Reber, B., Schmuck, R. & Vogt, H. (eds.), 2000. Guidelinies to evaluate side-effects of plant protection products to non-target arthropods. IOBC/WPRS, Gent.

Candolfi, M.P., Barrett, K.L., Campbell, P., Forster, R., Grandy, N., Huet, M.-C., lewis, G., Oomen, P.A., Schmuck, R. & Vogt, H. (in press). Guidance document on regulatory testing and risk assessment procedures for plant protection products with non-target arthropods. SETAC/ESCORT2 Workshop Report, 21-23 March 2000, Wageningen, The Netherlands.

EC, 2000. Working Document. Guidance Document on Terrestrial Ecotoxicology. European Commission, DG Agriculture. 2021/VI/98 rev. 7.

EPA, 1996. Ecological effects test guidelines. OPPTS 850.2500. Field testing for terrestrial wildlife. Public Draft.

EPPO, 1993. Decision-making scheme for the environmental risk assessment of plant protecion products. EPPO Bulletin. 23/24.

EPPO guideline for soil organisms (Eppo Standards Draft version 31-01-2001.5/2000). Eppo Standards. Environmental risk assessment for plant protection products. Chapter 7. Soil organisms.

EPPO, 1994. Decision-making scheme for the environmental risk assessment of plant protection products. Chapter 9. Arthropod natural enemies. EPPO bulletin 24: 17-35.

EPPO, 1992. Guideline on test methods for evaluating the side-effects of plant protection products on Honeybees. No. 170. Bulletin OEPP/EPPO Bulletin 22: 203-216.

EPPO guideline for non-target terrestrial arthropods (Eppo Standards Draft version 5/2000).

EPPO, 1998. Decision-making scheme for the environmental risk assessment of plant protection products. Chapter 10. Honeybees. EPPO bulletin 24: 37-87.

EPPO, 1993. Decision-making scheme for the environmental risk assessment of plant protection products. Chapter 11. Terrestrial vertebrates. EPPO bulletin 24: 37-87.

ISO 11268. Soil quality – Effects of pollutants on earthworms – Part 3. Guidance on the determination of effects in field situations. ISO 11268-3: 1999(E), first edition 1999-04-01.

Jong, F.M.W. de, Snoo, G.R. de, Canters, K.J. & Voet, E. van der, 1990. Field trials for the side-effects of pesticides. CML report 60, Leiden.

Jong, F.M.W. de, 2001. Terrestrial field trials for side-effects of pesticides. Thesis, Leiden University.

Posthuma, L., Suter, G.W. and Traas (eds.), T.P., 2002. Species sensitivity distributions in ecotoxicology. Lewis Publishers, London.

RESOLVE, 1994. Assessing pesticide impacts on birds. Final report of the avian effects dialogue group, 1988-1993. RESOLVE, Washington, DC.

SETAC, 1994. Barrett, K.L., Grandy, N., Harrison, E.G., Hassan, S. and Oomen, P.A. (eds.).Guidance document on regulatory testing procedures for pesticides with non-target arthropods. From the ESCORT Worrkshop, 28-30 march 1994, Wageningen.

Snoo, G.R. de and Canters, K.J., 1990. Side-effects of pesticides on terrestrial vertebrates. CML-reports 35.

Somerville, L. and Walker, C.H. (eds.), 1990. Pesticide effects on terrestrial wildlife. Taylor and Francis, London, New York, Philadelphia.

Suter, G.W., 1993. Ecological risk assessment. Lewis Publishers, Boca Raton.

A1.3.3 Other relevant literature

Brink, P.J. van den, 1999. Ecological and statistical evaluation of effects of pesticides in freshwater model ecosystems. Thesis Wageningen Agricultural University.

Chapman, P.F., Crane, M., Wiles, J.A., Noppert, F. and McIndoe, E.C., (Eds.) 1995. Asking the right questions: ecotoxicology and statistics. Report of a workshop held at Royal Holloway University of Londen, Egham, Surrey, UK, 26-27 april 1995. SETAC-Europe, Brussel.

Health Council of the Netherlands: Committee on pesticides and field research. Field research for the authorisation of pesticides. The Hague: Health Council of the Netherlands , 2000; publication no. 2000/07E.

Jong, F.M.W. de, J.W. Deneer, & W.L.M. Tamis, 2000. Veldbioassays. Ontwikkeling van een richtlijn voor veldbioassays met watervlooien en waterplanten voor het aantonen van bestrijdingsmiddelen in oppervlaktewater. CML report 150/Alterra rapport 061, Leiden, Wageningen.

Jong, F.M.W. de & W.F. Bergema, 1994. Field bioassays for side-effects of pesticides: final report. CML report 112, Leiden.

Jong, F.M.W. de, 1995. Framework for fieldtrials for the side-effects of pesticides. CML report 117, Leiden.

Lahr, J., Brink, P.J. van den and Brock, T.C.M., 1998. Ecologische risico's van bestrijdingsmiddelen in zoetwater ecosystemen. Deel 1: herbiciden. STOWA rapport 98-31, Utrecht.

Lahr, J. (Ed.). Methoden voor de analyse van effecten van toxische stoffen in veldsituaties. Verslag van de SAVE-workshop in Bilthoven, 10 maart 1998.

OECD, 1999. Revised consensus document on The application of the GLP principles to field studies. OECDSeries on Principles of GLP and Compliance Monitoring Number 6 (Revised) ENV/JM/MONO(99)22

Traas, T.P. et al., (1998). Modelling of ecosystem function in aquatic microcosms as influenced by chronic levels. RIVM 607601002, Bilthoven.

Wijngaarden, R.P.A. van, Geest, G.J. van and Brock, T.C.M., 1998. Ecologische risico's van bestrijdingsmiddelen in zoetwater ecosystemen. Deel 2: insecticiden. STOWA rapport 98-31, Utrecht.

Annex 3. A Summary Table for the data evaluation of field tests with aquatic organisms

Relevant guidance for pesticide evaluation has been condensed to Summary Tables. These are used for pesticide evaluation by RIVM/CSR. Summary Tables comprise a wide array of test items that may influence the test quality, they function as a checklist and they refer to scientific guidelines, when necessary. Apart from widely accepted guidelines, they represent additional criteria or calculation methods developed or agreed upon by RIVM and/or collaborating institutions. Actually, a Summary Table pretends to contain the most relevant field test items that should be dealt with for an adequate evaluation. However, expert judgement may always be necessary for final conclusions.

The interpretation of field tests is summarised in Table A1. Some important underlying ideas can be listed as follows:

1. A field test without actual analyses of the a.i. in the medium is considered unreliable, as the statistically or biologically significant effects cannot be linked with a proper exposure analyses.
2. The more questions (see the Notes in Table A3.1) are answered with "yes" (Y in the last column of Table A3.1) the less likely it is assumed that the field test is reliable. However, it depends on expert judgement whether *e.g.* 1-4, 5-10, and ≥11 of such "yes" items designate the test as a whole, as reliable, less reliable or unreliable, respectively. This is primarily because the test items may have a different "weight" in the overall judgement. Nevertheless, some sole, important test items — in Table 1 with "**Y [⇨Ri3]**" —designate a test already as unreliable: *e.g.* if a test design does not allow a proper statistical analysis this is already sufficient to label the whole test as unreliable (see also 4.1 test design, statistics in Table A3.1).
3. Table A3.1 refers to the adverse effects of pesticides to multi-species systems under outdoor conditions. Therefore, lower-tier studies as modified exposure multi-species tests, population-level studies, simple indoor multi-species tests, indoor defined microcosms, indoor semi-realistic microcosms, and single-species field tests are not included.
4. Table A3.1 refers to outdoor conditions: therefore the (micro)climatological conditions at the start and during the test are important. Wind during application may explain low recoveries in the water. Temperature and light conditions may influence dissipation processes to a large extent, if dissipation rates are temperature or light dependent. And even heavy rainfall — when diluting the a.i. substantially — may bias the test results.

For further guidance see [29].

Table A3.1. Test items for the data evaluation of field tests with pesticides (agricultural or industrial) : acute or chronic toxicity to aquatic organisms under outdoor conditions.

Table A3.1

	TEST ITEMS	NOTES	RELIABILITY LOWER?
METHODOLOGY & TEST DESCRIPTION	**1. substance**	1. **improperly characterised or reported?**	1.
	1.1 concentration	1.1 [*identity and amount of a.i. per litre test water: initially and frequently during the test, if required; see also 3.2 (application) rate and extraction/analysis*]	1.1 **Y [⇨RI 3]**
	1.2 purity	1.2 [*purity? identity and amounts of impurities?*]	1.2 **Y**
	1.3 formulation	1.3 [*name? type? composition — e.g. quantities and function of non-active ingredients, e.g. wetting agents? preferably the same formulation as the one under registration*]	1.3 **E**
	1.4 vehicle	1.4 [*in case a vehicle — other than in the formulation — is used, this and the concentration used should be reported*]	1.4 **E**
	2. duration/observation time	2. **not reported?** [*for testing chronic toxicity, preferably until there is an equilibrium between external and internal concentration*]	2. **Y [⇨RI 3]**
	3. application	3. **improperly characterised or not reported?**	3.
	3.1 type I	3.1 [*e.g. spraying with a spraying boom, a knapsack, a plane or just dissolving and homogenising the a.i. into the test medium*]	3.1 **Y [⇨RI 3]**
	3.2 rate(s)	3.2 [*see 1.1 concentration; actual concentrations during the test are most important; however, additional data on the nominal application rate may give an indication of the agronomical or industrial similarity between the field test and the pesticide use under registration*]	3.2 **E**
	3.3 type II	3.3 [*when spraying, the type of boom should be reported; this is necessary to get a picture of the type of exposure*]	3.3 **Y**
	3.4 date, year and location	3.4 [*necessary to make a link between the effects and local environmental and climatic conditions; season indicate light regime as well*]	3.4 **Y**
	3.5 (micro)climate	3.5 [*at the start of an outdoor application e.g. wind speed and temperature may be relevant for the exposure assessment as they may explain low recoveries in the water*]	3.5 **E**
	4. test design I	4. **improperly designed/improperly or not reported?**	4.
	4.1 statistics	4.1 [*the design should allow appropriate statistical analysis: all tests require accurate statistical analysis for proving significant differences between the treatment group(s) and the control(s); NOECs preferably derived by regression or by multivariate techniques as PRC*[2]]	4.1 **Y [⇨RI 3]**
	4.2 type & size	4.2 [*e.g. outdoor microcosm, outdoor pond or mesocosm*]	4.2 **Y**
	4.3 dose-response	4.3 [*≥2 test concentrations for finding a dose-response relation (controls excl.); the design should allow appropriate statistical analysis*]	4.3 **Y**

[2] PRC is Principle Response Curve, a clustering technique based on the Redundancy Analysis Ordination technique (see [33]).

Table A3.1 (contd.)

	TEST ITEMS	NOTES	RELIABILITY LOWER?
METHODOLOGY & TEST DESCRIPTION	**4.4 test system I**	4.4 [*the test system should mimic a (semi-)natural outdoor ecosystem with individual(within the species), population, community and ecosystem interactions; the system should be pre-tested and proven to be "stable": it should not collaps due to other reasons than applying a pesticide (e.g. by overloading the system with fish); no additional feeding*]	4.4 **Y**
	4.5 test system II	4.5 [*biological and physicochemical properties of the test system (incl. the water and sediment) should be determined frequently— also the equilibration can then be checked — and also during the main test; monitoring pH, hardness, oxygen content, temperature and the occurrence of other toxic chemicals will help to interpret test results; the thickness of a sediment layer, if included, should be reported*]	4.5 **Y**
	4.6 pre-treatment	4.6 [*the system should be well equilibrated; the intactness of the system should be measured or monitored adequately;equilibration is of particular importance for the benthic stratum, if included; how long and at what temperature are the test organisms pre-treated?*]	4.6 **Y**
	5. test organisms I	5. **not representative or improperly reported?** [*e.g. genus and species of all the involved organisms should be reported; strains or clones need not necessarily be reported; the community/ecosystem should resemble a Dutch natural habitat to a reasonable extent: e.g. macrophytes, algae, fish, crustaceans and sediment-dwellers are preferably involved*]	5. **Y**
	6. test organisms II	6. **insufficiently or not monitored/improperly or not monitored?**	6.
	6.1 general features	6.1 [*length, weight, behaviour, health condition, reproductive features should be reported satisfactorily when necessary to determine the endpoints*]	6.1 **Y**
	7. control	7. **sufficient number, invalid or improperly reported?**	7.
		7.1 [*the control is very important in a field test; for the number of control replicates see 4.1 statistics*]	7.1 **Y [⇨RI 3]**
		7.2 [*preferably the results in the control are represented with confidence limits*]	7.2 **Y**
	8. extraction/analysis	8. **inappropriate or improperly reported?** [*actual concentrations in the test medium should be determined properly; actual concentrations in other compartments or biota may be helpful when interpreting test results*]	8. **Y**

Table A3.1 (contd.)

	TEST ITEMS	NOTES	RELIABILITY LOWER?
RESULTS	**9. endpoint** **9.1 type**	9. **inappropriate or improperly reported?** 9.1 **inappropriate**? [*relevant endpoint are particularly those that are closely related with population, community or ecosystem dynamics: e.g. biochemical,morphological, behavioural, mortal, respecting growth or reproduction, intraspecific (e.g. population growth), interspecific,respecting biodiversity, primary and secondary production, food web interactions, resilience; $NOEC_{FIELD}$ therefore can be based on different structural and functional parameters; it should be clearly reported on which endpoint the $NOEC_{FIELD}$ is based: therefore all investigated endpoints and those that are the most sensitive, whereupon the $NOEC_{FIELD}$ is finally based, should be reported*]	9. 9.1 **Y**
	9.2 value	9.2 **improperly reported?** [*as there are generally less replicates than in lab tests, the range of the $NOEC_{FIELD}$ is more important; therefore the 95% confidence limits of endpoints should be reported*]	9.2 **Y [⇨RI 3]**
	10. effect class	10. **properly derivable?** [*are there no effects (class 1) or are the effects slight (class 2), large, acute/short-term (class 3), large in short-term test (class 4) or large, long-term (class 5)*]	10. **Y [⇨RI 3]**
	11. biological meaning of statistically significant differences	11. **insufficiently explained?** [*expert judgement is a must. A significant reduction in e.g. the number of crustaceans in a field test can only be explained on a case-by-case basis by taking into account*: *A lower-tiered laboratory tests with crustaceans;* *B the presence of a dose-effect relation: if such a relation exists it is much more probable that a consistent significant difference is actually due to the substance;* *C the specific test conditions that — besides the substance itself — might have influenced population dynamics (therefore the use of PRC is a better tool to determine an NOEC than the conventional regression analysis: PRC may indicate whether population changes are due to other causes than the chemical itself: e.g. pH, oxygen, temperature or a particular population build-up)*]	11. **Y**
	12. verification of endpoint	12. test results are not verificable? [*verification of conclusions by data recalculation increases the reliability of the field test; however, this is not possible when raw data are not reported; TOXSTAT may be useful for verification*]	12. **Y**
	13. actual concentrations	13. **not measured?** [*actual concentrations in the test medium should be measured frequently throughout the test; data on vapour pressure, water solubility, photolysis, hydrolysis, biodegradation, sediment and plant sorption can be helpful in explaining the substance dissipation*]	13. **Y [⇨RI 3]**

Table A3.1 (contd.)

NOTES	
R E M A R K S	**14. similarity aspects**(see Section 2.5) [*the usefulness of a field test depends on the scientific reliability and purpose of use (where should the test results be used for?). Similarity may be checked for at least six test items: product, dosage/concentration, type and time of application, application frequency and interval, geographical location and the crop (stage). The more similarity aspects are found between the field test and the product under registration and its proposed conditions of use, the more likely it is that the field test is useful for ERA. general guidance however is difficult to give. Expert judgement therefore is decisive, as the appraisal of the usefulness may differ from pesticide to pesticide. As an example: a field test with spraying without a spray-free edge zone is not necessarily useful when the application under registration requires such a spray-free zone. Such a field test may be useful in case no adverse effects were found. When, however, effects were demonstrated these may not be found under the practice circumstances that require a spray-free margin. Ideally, the application in the field test resembles the one under registration — as notified in the application form — as much as possible. However, a less reliable field test with a different application rate or type may be used as circumstantial evidence*]

Appendix 1 Mailing list

1. Directeur-Generaal RIVM, H.A.P.M. Pont
2. Dr. H. Roelfzema, GZB/CO, Den Haag
3. Prof. Dr J.J. Sixma, Gezondheidsraad, Den Haag
4. Dr. W.F. ten Berge, DSM, Heerlen
5. Prof.dr. J.S.M. Boleij, CTB, Wageningen
6. Dr. S. Bosman-Hoefakker, CTB, Wageningen
7. Dr. Ir. P.C. Bragt, VWS/HIGB, Den Haag
8. Dr. J. Bridges, CSTEE, Brussels, België
9. Ir. G. de Bruijn, BRD, Wageningen
10. Dr. F.M. Carpanini, ECETOC, Brussel, België
11. Dr. M. Clook, PSD, York, United Kingdom
12. Dr. J. Costa-David, CSTEE, Brussels, België
13. Dr. C. Cuypers, VWS/GZB, Den Haag
14. Dr. J.I. Delic, HSE, Bootle, United Kingdom
15. Dr. C. DeRosa, ATSDR, Atlanta, USA
16. Dr. R. van Doorn, GGD, Rotterdam
17. Drs. J.W. Dornseiffen, GZB/VVB, Den Haag
18. Dr. M.L. Dourson, TERA, Cincinnati, USA
19. Dr. W. van Eck, VWS/GZB, Den Haag
20. Dr. S. Fairhurst, HSE, Bootle, United Kingdom
21. Dr. H.E. Falke, CTB, Wageningen
22. Dr. R. Fielder, Department of Health, Londen, United Kingdom
23. Dr. R. Garret, EPA-OPPTS, Washington DC, USA
24. Dr. J. Gift, EPA-NCEA, Washington, USA
25. Dr. D.G. Groothuis, ST.TZ/W&V, Den Haag
26. Prof. dr. U. Gundert-Remy, BGVV, Berlijn, Duitsland
27. Dr. S. Hanai, JCIA, Tokyo, Japan
28. Dr. B. Hansen, ECB, Ispra, Italië
29. Dr. J. Hart, EPA, Kopenhagen, Denmark
30. Prof. dr. P.T. Henderson, RUL, Maastricht
31. Dr. R.F. Hertel, BGVV, Berlijn, Duitsland
32. Dr. J. Herman, IPCS, Genève, Switserland
33. Dr. A. Jarabek, EPA-NCEA, Washington, USA
34. Dr. G. Joermann, BBA, Braunschweig, Germany
35. Drs. H.W. Jeuring, ST.TZ/W&V, Den Haag
36. Dr. D.W.G. Jung, DGM/SAS, Den Haag
37. Dr. F. Kalberlah, FoBiG, Freiburg, Duitsland
38. Prof. dr. I. Rietjens, LUW, Wageningen
39. Dr. ir. M. Ruijten, GHOR, Rotterdam
40. Prof. dr. R. Kroes, RITOX, Utrecht
41. Dr. D. Kroese, TNO-ATRA, Zeist
42. Dr. C. Lally, P&G, Strombeek-Bever, België
43. Dr. E. Löser, Bayer, Duitsland
44. Dr. B. Meek, Health Canada, Ottawa, Canada
45. Dr. L. Messchendorp, CTB, Wageningen
46. Dr. G. Moore, KEMI, Solna, Zweden
47. Dr. S. Munn, ECB, Ispra, Italië
48. Dr. P. van Noort, RIZA, Lelystad
49. Dr. E.V. Ohanian, OST, EPA, Washington, USA
50. Dr. W.F. Passchier, Gezondheidsraad, Den Haag
51. Dr. H.R. Pohl, ATSDR, Atlanta, USA
52. Ir. ing. W. Pol, CTB, Wageningen
53. Dr. A.G. Renwick, University of Southampton, United Kingdom
54. Dr. M. Reynier, INRS, Parijs, Frankrijk
55. Prof. dr. W. Seinen, RITOX, Utrecht
56. Dr. Y. Sekizawa, NIHS, Tokyo, Japan

57. Dr. C. Sonich-Mullin, IPCS, Cincinnati, USA
58. Ir. J.M. de Stoppelaer, VWS/VVB, Den Haag
59. Ing. J. Stronkhorts, RIKZ, Den Haag.
60. Drs. G.T.J.M. Theunissen, LNV/VVM, Den Haag
61. Dr. R. Tomerlin, NoviGen Sciences, Washington, USA
62. Dr. H.G. Verschuuren, Dow Europe, Horgen, Switserland
63. Dr. E. van Vliet, Gezondheidsraad, Den Haag
64. Drs. P. van Vliet, CTB, Wageningen
65. Dr. V. Vu, NCEA, EPA, Washington, USA
66. Dr. A. Wagner, DHSH, Canberra, Australië
67. Mr. J.A.M. Whyte, VWS/GZB, Den Haag
68. Dr. A.A.E. Wibowo, Coronel Laboratorium, Amsterdam
69. Dr. A. Wilson, HSE, Bootle, United Kingdom
70. Dr. G.Würtzen, Coca Cola Nordic and Baltic, Hellerup, Denmark
71. Dr. M. Younes, IPCS, Geneve, Switserland
72. Dr. J. van Zorge, DGM/SAS, Den Haag
73. Depot Nederlandse Publikaties en Nederlandse Bibliografie
74. Directie RIVM
75. Sector Directeur milieurisico's en Externe Veiligheid, Dr. ir. G. de Mik
76. Sector Directeur Voeding en Consumenten Veiligheid, Prof. Dr. Ir. D. Kromhout
77. Sectordirecteur Volksgezondheidsonderzoek, Dr. G. Elzinga
78. Hoofd Centrum voor Stoffen en Risicobeoordeling, Dr. W.H. Könemann
79. Hoofd Laboratorium voor Geneesmiddelen en Medische Hulpmiddelen, Dr. ir. J.F. van Sonderen
80. Hoofd Laboratorium voor Pathologie en Immunobiologie, Prof. Dr. J.G. Vos
81. Hoofd Laboratorium voor Blootstellingsonderzoek en milieu-epidemiologie, Dr. ir. E. Lebret
82. Hoofd Laboratorium voor Ecotoxicologie, Drs. J.H. Canton
83. Hoofd Laboratorium Nationaal Vergiftigingen Informatie Centrum, Dr. J. Meulenbelt
84. Hoofd laboratorium voor Effectenonderzoek, Dr. A. Opperhuizen
85. Hoofd Laboratorium voor Bodem- en Grondwateronderzoek, Ir. R. van den Berg
86. Hoofd Laboratorium voor Organische-analytische Chemie, Dr. P. van Zoonen
87. Dr. A.J. Baars (CSR)
88. Dr. P. van Beelen (ECO)
89. Dr. R.B. Beems (LPI)
90. Dr. J. van Benthem (LEO)
91. Dr. ir. S. de Boer (LGM)
92. Dr. A.B.T.J. Boink (LEO)
93. Dr. L. van Bree (LEO)
94. Drs. M. Draaijer (CSR)
95. Dr. Ir. J.G.M. van Engelen (CSR)
96. Dr. H.E.M.G. Haenen (LGM)
97. Drs. P.H. van Hoeven-Arentzen (CSR)
98. Dr. A.A.J. van Iersel (SB1)
99. Drs. J. Janus (CSR)
100. Drs. A.G.A.C. Knaap (CSR)
101. Dr. J.W. van der Laan (LGM)
102. Dr. F.X.R. van Leeuwen (CSR)
103. Ir. A.M.A. van der Linden (LBG)
104. Ir. J.B.H.J. Linders (CSR)
105. Dr. H. van Loveren (LPI)
106. Dr. W.C. Mennes (CSR)
107. Dr. J. Meulenbelt (VIC)
108. Drs. M.H.M.M. Montforts (CSR)
109. Ing. J.J.A. Muller (CSR)
110. Dr. B.C. Ossendorp (CSR)
111. Dr. A.H. Piersma (LEO)
112. Dr. Ir. M.N. Pieters (CSR)
113. Dr. E. Putman (CTG)
114. Ir. G.J. Schefferlie (CSR)
115. Dr. A.J.A.M. Sips (LBM)
116. Prof. Dr.W. Slob (LEO)

117. Dr. G.J.A. Speijers (CSR)
118. Ir. H. Stevenson (CTG)
119. Dr B.M. van de Ven (CSR)
120. Ir. A.J. Verschoor (LBG)
121. Dr. W. Vleeming (LEO)
122. Dr. P.W. Wester. (LPI)
123. Drs. I. De Vries (VIC)
124. – 134. Auteurs
135. Bureau Voorlichting & Public Relations
136. Rapportenregistratie
137. Bibliotheek RIVM
138. Bibliotheek Centrum voor Stoffen en Risicobeoordeling
139. – 163. Bureau Rapportenbeheer
163. – 170. Reserve exemplaren