Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

16-Feb-2009

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND
THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

ENV/JM/TG/RD(2009)9
For Official Use

Test Guidelines Programme

DRAFT NEW TEST GUIDELINE FOR A COLLEMBOLAN REPRODUCTION TEST IN SOIL: COMMENTS AND RESPONSES

21st Meeting of the Working Group of National Coordinators of the Test Guidelines Programme

31st March-2nd April 2009, OECD Headquarters, Paris, France

Jukka AHTIAINEN
Tel: +358 400 148621; Fax: +358 20 2591; Email: jukka.ahtiainen@ymparisto.fi

JT03259723

English - Or. English

This document presents the comments received from Canada, Czech Republic, France, Germany, Switzerland, the Netherlands and BIAC on the draft Test Guideline that was made available for comments in September 2008. The revised draft including track changes is available on the protected website in "Draft Test Guidelines and Guidance documents", under “WNT 21”.

ACTION REQUIRED: ***The Working Group of the National Coordinators of the Test Guidelines Programme is invited to take note of the comments and responses.***

ANNEX 2

Compilation of Member countries' and BIAC's comments on the September 2008 draft Test Guideline for a Collembolan reproduction test, and responses from the Secretariat to those comments.

GENERAL COMMENTS		
Country	Comment	Secretariat's responses
Canada	Comments were received from: - Environment Canada (as commenter 1) - PMRA (as commenter 2) Commenter 1: • Suggest modifying the statement that only adult females be included as part of total observed mortality since *F. fimetaria* requires both female and male adults throughout the test, and are present in a 50:50 ration. • Suggest providing greater detail on timing for feeding of the *F. candida*, to be consistent with level of detail provided under the *F. fimetaria* procedure description. Commenter 2: This proposed OECD Guideline is designed to assess the effects of chemical substances on the reproduction of the collembolan species *Folsomia fimetaria* L. and *Folsomia candida* Willem in soil. Both species are well selected to represent sexual and asexual modes of reproduction of collembolans in different habitats. They have world-wide distribution, and are present in Canada. Their easy accessibility makes them ideal species for the standard tests.	The text has been revised There are equal details on feeding for *F. fimetaria* compared to *F.candida*.

	The guideline presents detailed information on how to perform toxicity tests with *F. fimetaria* and *F. candida.* It includes a range finding test and a definitive reproduction test. The endpoints for test substances, validated by a reference substance, are determined to be LC_x , EC_x, and/or NOEC. This protocol is feasible and practical. Comments below are made to improve the clarification and reliability of the guideline: 1. Subtitle of the guideline: The guideline is titled the Collembolan reproduction test. The effect of test substances is assessed by the total number of juveniles produced by parent insects and the survival of the parents. However, the number of parents available for the actual reproduction test may vary among treatments due to the initial acute mortality of the parents, even though the same numbers of insects were used at the beginning of the test. Therefore, the effect determined by the guideline actually is the cumulative effect of both acute mortality to the parents and the reproduction of the survivals of parents. As a suggestion, sublethal concentrations identified in the range-find test could be used for reproduction tests. The use of sublethal concentrations may also reduce the high variation for the *F. fimetaria* test as described below. 2. High variation in effects in the *F. fimetaria* test: The guideline proposes a set of reasonable validation criteria for the test. According to these criteria, approximately half of the tests for *F. fimetaria* failed due to the inexperienced participants as described in the ring test report. However, these tests were conducted by highly experienced professional laboratories and research laboratories. Other unidentified factors might be expected to affect the reliability of this protocol. One of them might be the establishment of consistent test conditions, such as maintaining soil moisture content and providing food for insects without molds. In the guideline, soil moisture content is maintained by replenishing water to compensate the weight loss of test vessels. A more effective method, such as maintaining test vessels in an environmental test chamber with an effective ventilation and relative humidity equivalent to the expected soil moisture content, should be specified.	1. How to use the range-finding results is not prescribed, range-finding will give information on adult mortality and reproduction. 2. Yes, the inexperience was with respect to working on *F. fimetaria*, not the general experience of the ringtest labs. There was no clear indication of what made a lab being able to run *F. fimetaria* sucessfully. The approach to ensuring moisture is left open, as this may be done equally efficiently in many ways. However, the suggestion is good, and most labs would react on the recommendation to ensure high moisture.

Czech Republic	There should be a recommendation in the Guideline that the chemical analysis should be performed to check the real concentration in the soil, because some losses during substance application can be expected. The statistical analysis recommendations should be unified throughout the Guideline and make clearer.	The OECD test with *E. fetida*, Enchytraeids and *H. aculeifer* did not include such specifications in contrast to the plant test, OECD 208 in the section "Verification of test substance concentration/rate" this may be adopted for the springtail guideline. Statistical advice has been kept brief and are identical to the *H. aculeifer* test guideline no. 226. Since OECD released the statistical guidance document no. 54 in 2006, to which reference has been made, statistical guidance should be kept to a minimum, if not omitted.
Germany	We highly appreciate the proposed guideline for a Collembolan reproduction test. As the draft is largely based on the existing ISO 11267 guideline with proven suitability under practical conditions, the basic technical requirements of the new proposal are indisputable. However, two major points need further discussion at this stage: **a)** The suitability of *Folsomia fimetaria* as a second test species equivalent to *Folsomia candida* is highly questioned by the results of the ring test reported by Krogh et al. (2008): • As described in the report, in 56% of all tests performed with *Folsomia fimetaria* the control mortality was > 20%. In the tests with *Folsomia candida* only 21% showed a control mortality above 20%. • In 50% of all tests performed with *Folsomia fimetaria* the mean control reproduction was below 100 Juveniles whereas in the tests with *Folsomia candida* only 3% showed a lower mean control reproduction than 100 juveniles. • In comparison to *Folsomia candida* the egg hatching rate in the stock cultures was generally much lower. Therefore, to start a test with *Folsomia fimetaria* a lot more synchronisation cultures have to be initiated to achieve a sufficient number of animals to be introduced into the test	a) The overall conclusions from the ring-test report are correct. However, the main purpose of introducing *F. fimetaria* is due to its differences from *F.candida*

	vessels. • Selection of 10 male and 10 female *Folsomia fimetaria* for introduction into the test vessels is much more complicated and laborious than taking 10 juvenile *Folsomia candida* without any kind of selection process. Most severely, the validity criteria for mortality and reproductive output given in the guideline were not fulfilled by at least half of the tests performed with *F. fimetaria.* Krogh et al. (2008) excluded these "low performance results" of *F. fimetaria* testing by disqualifying them as outliers. This is considered as scientifically not justified. Rather one should discuss different acceptability levels / validity criteria for the two species. In view of the problems accounted during ring test with *F. fimetaria* and taking into account the long history of *F. candida* as test organism (especially due to its ease of handling), *F. candida* should be the main test organism (could be already pointed out in the title of the guideline as in guidelines for other species). For particular cases (i.e. endocrine disrupting substances) the use of *F. fimetaria* as sexually reproducing species could be an alternative. Still, there are questions to be answered regarding the relevance of the different reproductive strategies of these species: if the assumption is that *F. candida* and *F. fimetaria*, because of their parthenogenetic or gonochoristic reproduction, have different sensitivities to compounds with endocrine-modulating properties, this hypothesis should be proved, during the validation process, by comparative testing of appropriate compounds. Obviously, this has not been done until now. The compounds tested so far caused fairly consistent responses in the two species. If that should hold true for EDs as well, a second sexually reproducing test species is deemed not necessary. Summarizing, the test guideline should be described mainly for *F. candida.* The usage of and performance with alternative species like *F. fimetaria* can be described in an annex (as it is already done in Annex 2, Point 3 “Alternative Collembolan species”).	The collembolan ring test report has **not** excluded results due to low performance. Outliers has been identified objectively by employing the *h* statistic (section 4.10) The validity criteria should be kept as they are for any species in order not to violate the statistical performance. The title reflects the option to use other alternative validated species collembolan species. The reasons for this may be plenty; this was wisely introduced with the enchytraeid guideline 220. The guideline do not state requirements for which test species to choose, so the species preference is left to be decided by environmental managers etc. The purpose of the guideline is to provide the methodology to perform quality tests with collembolans with either a parthenogenetic or sexual mode of reproduction and with a different habitat origin. The introduction of a sexually reproducing species includes additional biological receptors that ensures that these potentially will be exposed to chemicals. This introduction cannot await detailed research for precautionary reasons. Such research would hardly be done in a ringtest exercise. The organisation of the guideline may be done according to different preferences.

	b) These test guidelines are intended for testing of chemicals, i.e. testing the intrinsic toxicity of a chemical substance. For Collembola, the major exposure route of interest is via the soil substrate (i.e. application of the test substance by mixing into the test substrate). This clear intention is partly lost and the comparability of test results is hampered by alternatively proposing the application of chemicals by spraying onto the soil surface (para 18 and 22). This mode of application is much less clear regarding the relevant route of exposure / bioavailability (relevance of dermal exposure from direct overspray versus contaminated feed (yeast) versus contaminated soil) and even avoidance of the surface soil contaminated by the test chemical may occur. It should be discussed whether surface application is generally appropriate for testing the intrinsic toxicity of chemicals via soil exposure, or if this is an option for a higher tier testing (of pesticides), only. (This difficulty already exists with other soil testing guidelines, e.g. earthworm and should be considered carefully at this stage.)	b) If OECD does not provide guidance on how to test surface sprays, then these sections should be deleted. However, the test protocol as such is appropriate for running test with surface application, if this should be preferred by a member country. It is in the interest of OECD members states that compounds that have the most common application on the soil surface can be performed according to an OECD guideline. We agree with this comment and suggest to include that the main purpose is to provide the intrinsic toxicity or basic toxicity of a compound and retain guidance on how to test surface applied compounds such as pesticides. The comment points to a general need for updating and harmonising the soil test guidelines. We think both options should be kept.
The Netherlands	This guideline looks very good, and is well supported by the ring test performed. The test on *Folsomia candida* is in line with the existing guideline ISO 11267, and ring-test results show this method functions quite well. Only concern is the variation seen in the *Folsomia fimetaria* test. This may ask for relaxation of the validity criterion of variation, maybe increase the allowed variation to 40% in order not to frustrate routine application of the test.	As demonstrated in the ring test report, the *F. fimetaria* CV is about 30% and similar to *F. candida* if succesfully performed. Relaxing the CV in validity would make toxic endpoint estimates less precise.

	Generally, the proposal is written clearly and in a coherent way.	
BIAC	The proposal for a new guideline for a Collembolan reproduction test has to be restricted to the Collembolan species, *Folsomia candida*, because 1.) only a small – insufficient – number of laboratories (only 6 laboratories) participated in ring-testing of the Collembolan species, *Folsomia fimetaria*, 2.) the failure rate for the proposed test guideline validity criteria for *Folsomia fimetaria* is > 50% (as given in the ring-test report (Krogh 2008)) is far too high. This means it is impossible to generate such data in a reliable manner for *Folsomia fimetaria*. 3.) no evidence is given that the proposed new guideline will work for the proposed "alternative Collembolan species" (given in section 3 of Annex 2). Based on the ring-test data provided for *Folsomia fimetaria* it is highly unlikely that successful testing with the proposed alternative Collembolan species is possible at all and that e.g. the proposed validity criteria could be fulfilled. – For any other Collembolan species added or proposed in the test guideline further validity criteria need to be established first. It would also be useful to provide the option for aged residue tests, and suggest methodology (for example to use a microbially active characterised field soil such as LUFA 2.2).	1) 8 laboratories were performing the *F. fimetaria* test. This information will be available now in the DK EPA ring test report. 2) There is no criteria that specifies a failure of 50% leading to an invalidation of a test protocol and the intralaboratory analysis provide strong validation of the test protocol. The test guideline is an extension of the *F.candida* test and testing of other collembolan species is allowed. The ringtest demonstrates that it is possible to produce reliable data, but it is not surprising that some labs will not be successful in employing a new test species without long-term practise that is available for *F.candida*. 3) In many other TGs it has been mentioned that alternative species can be used if they can fulfill the validity criteria. The proposal would keep the validity criteria as they are for both the species mentioned or apply them for any alternative species in order not to violate the statistical power of the test. The ring test has proven the different performances of the two species, but these

	Can advice be given on sources of organisms?	results cannot be generalised and included in the TG. If a laboratory is not able to perform a test with alternative species the end user of the test results must take appropriate precautions. Proposal would be also to leave the species names out of the title. The sources of the organisms could be given if this is agreed.
Paragraph 1		
Switzerland	**Expert 1** (Paragraph 1-5) This paragraphs describes well the rationale for the choice of test species, the ecological importance and links this guideline to exisiting procedures.	
BIAC	Restrict section to *Folsomia candida*. – Delete all other sections mentioning *Folsomia fimetaria.*	See the responses above. There is valuable information in how to test a sexual collembolan species
Paragraph 2		
Switzerland	**Expert 2** (Paragraph 2) …and represent **arthropods** species with a different …": Unclear: earthworms and enchytraeids are not arthropods. **Arthropods species** should be replaced by **soil invertebrate species**.	No comment
Paragraph 3		
Switzerland	**Expert 2** (Paragraph 3) "Adults typically measure 0.5 - 5 mm**, their** contribution to total soil animal biomass and respiration is low, estimated between 1% and 5%": 2 sentences should be made: Adults typically measure 0.5 - 5 mm**. Their** contribution to total soil animal biomass and respiration is low, estimated between 1% and 5%".	The text has been revised

BIAC	It should read: "….commonly reach **up to** 10^5 m^{-2} in soil…" (add: up to)	"up to" is implied in "commonly reach"
Paragraph 4		
Switzerland	**Expert 2** (Paragraph 4) "and may **influence mycorrhiza**,": In which way: positively/negatively? Precisions should be given.	Should no be specified – this is why "may" is used.
Paragraph 5		
Czech Republic	Replace "1 per mille" with "1 per million"	"mille" means per 1000
Switzerland	**Expert 2** (Paragraph 5) "The ecological role is similar to the role of *F. fimetaria*": The ecological role of ***F. candida*** is similar to the role of *F. fimetaria* or **its** ecological role	Done
Paragraph 6		
Canada	Commenter 2: Some wording changes: Use consistent numbers and time units throughout the guideline, such as for "two weeks for *F. fimetaria* and 3 weeks for *F. candida*" in paragraph 6; "Food is added … after 14 days" in paragraph 28 while "Food is added … after about 2 weeks" in paragraph 29" etc.	Revised
Czech Republic	Replace "an artificial soil according to OECD 207" with "an artificial soil prepared according to OECD 207" Use the same units (either days or weeks) for duration of both F. fimetaria and F. candida tests.	Revised
Germany	3rd paragraph, 2nd sentence: "… or 28 days for F. candida."	Changed to weeks
The Netherlands	Why OECD soil with 5% sphagnum peat content instead of 10% is advised? OECD guidelines 222 and 207 advise medium with 10% peat. Both species tolerate, and *Folsomia candida* even prefers, habitats with high organic matter content. Having	There has been a growing agreement that this is more appropriate because it is closer to normally occuring soil O.M. content (adopted in OECD 226)

	similar composition of OECD medium for more tests would make comparison of results easier.	
Switzerland	**Expert 1:** No rationale for the choice of 5 % organic matter is presented. Why is the range finding test shorter than the effective test, maybe long term responses can only be seen after 2 weeks and are missed like that if the range finding is only two weeks. Which exact endpoints are used for "reproduction"?	The purpose of the range-finding is to indicate the level of toxicity. For this purpose the test period may be shorter
BIAC	Restrict section to *Folsomia candida*	Not agreed as discussed and proposed above.
Paragraph 9		
Czech Republic	Delete "per replicate" in the last sentence. The CV is not for replicate but for treatment, control etc.	Deleted
Germany	4th sentence: "… per vessel should be…"	Done
The Netherlands	As argued above, it may be considered to relax the third validity criterion (CV of nr of juveniles/replicate) and make it 40% instead of 30% (only for *Folsomia fimetaria*) See paragraph 6	Not recommended. See responses above.
Switzerland	**Expert 2:** "Mean **adult mortality** should not exceed 20% at the end of the test;": See paragraph 40. Changes should be made if appropriate.	Done
Paragraph 10		
Czech Republic	Add "dry weight soil" after "at about 100 mg/kg"	Revised
Germany	Experiences with boric acid as test substance is not convincing. In two of three tests we failed in getting a dose-response relationship and in the most recent test the lowest concentration tested was 320 mg/kg indicating a threshold level higher than 50 % reduction. It is therefore recommended to include effect levels for at least one	The experience by the commenter is not representative to all labs in the commenting country.

	more chemical substance. Boric acid (CAS 10043-35-3) is classified by the EU Commission staff working document SEC(2007) 1635 on the implementation of the "Community strategy for Endocrine Disrupters" as a "substance with evidence of endocrine disrupting effects (Category I)". It should be carefully checked whereas this may concern operators in the laboratory and/or may question the general suitability or acceptability of boric acid as a reference substance. We strongly propose to align the text with the ones on "reference substances" in other OECD test guidelines (if there are not good reasons not to do that). To our understanding, a clear guidance should be given how to design a test with the reference substance, how many test concentrations to be included, when and how often to perform such a test, how to interpret any results in the light of existing information (historical database), and which consequences to obey (e.g. any links to validity criteria?). Terms like "at regular intervals" or "about 100 mg/kg)" are not considered to be very helpful and may easily lead to misunderstandings.	The comment may be addressed in generally updating OECD guidelines. OECD recently included boric acid. Dimethoate may be employed at a dose of about 1 mg kg^{-1}. The details on how to ensure a proper test-system performance through a positive control/reference substance may vary from lab to lab and should not be too strict. However, please note that it is strictly required to run a reference compound test.
Switzerland	**Expert 2:** "which should reduce reproduction by 50% [10, 11] at about **100 mg/kg**.": Precisions should be added: 100 mg/kg (**dry weight**) for **both species.**	Revised
Paragraph 11		
Czech Republic	The use of plastic containers should be avoided because of possible sorption of some chemicals to the walls which decreases final exposure. This notice should be added.	Revised
Germany	1st sentence: "… 30 g of <u>dry</u> soil …"	"fresh soil", as this is the final condition of the test soil
Switzerland	**Expert 2:** "inert plastic (non-toxic)": Example could be given: inert plastic (non-toxic; e.g. polyethylene (PE) or polypropylene (PP)) "The test vessels should have a cross-sectional area allowing the actual soil depth within the test vessel to be 2-4 cm":	The performing lab should ensure the specified soil depth – volume is optional.

	More precisions could be given: e. g. about 100 ml capacity, 5 cm in diameter	
Paragraph 13		
Czech Republic	Second sentence: "a natural standard soil" should be changed to "a natural soil" because it is not defined what "natural standard soil" is. In case LUFA is meant, it should be specified. In the part beginning with "approximately 74% air-dried industrial sand ....", there is doubled information that sand amount depends on CaCO3 addition (firstly in the parentheses and the sentence at the end. In my opinion the simplest way is to set 75% for sand and then specify CaCO3 because < 1% of sand cannot make any differences Note 1. There should be notice that pH should be measured in equilibrated soils after being moist for few days. Measuring of pH in dry artificial mixture gives results far from real pH when the soils are at 40-60% WHC for few days. Note 2. Not only C/N and CEC should be reported for artificial soil used, but TOC itself. Note 3. Add CEC among the properties of natural soil which should be determined.	"standard" is deleted We don't see the problem here. See section 15 CEC and TOC added
Germany	For purposes of testing intrinsic properties of chemical substances a unique test substrate should be used. Taking into account the existing data base artificial soil should be taken. (See also general comment b) above.)	No comment
Switzerland	**Expert 2:** "(a particle size of 2 +/- 1 mm is acceptable)": Replace by: A particle size of 2 ± 1 mm is acceptable "Note 2: It is recommended to measure the pH and optionally the C/N ratio and **CEC** of the soil": **Cation Exchange Capacity** (CEC) should be added	Revised Revised
BIAC	Check English of the final sentence	Revised
Paragraph 15		
Canada	Commenter 2: Specify the "required range" of soil pH value.	See §13

Czech Republic	Some advice should be added on which inorganic acid to use.	No specific advice available, e.g. HCl
Germany	4th sentence: "… a harmless inorganic acid." (Please specify.)	Revised
Switzerland	**Expert 2:** Precisions should be added regarding the "storage" of the soil during the equilibration time (equilibration at room temperature?) When the pH of the soil must be verified? Before equilibration? Precisions should be added "by the addition of a harmless **inorganic acid**" An example of inorganic acid could be given	"2-7 days" was already specified See §13 and Annex 1
BIAC	Give an example of a "harmless inorganic acid"	No example available yet. Probably HCl.
Paragraph 16		
Switzerland	**Expert 2:** No precisions are given regarding when and how much water has been added to obtain the final required moisture content of the soil. Same comments regarding the preparation of test concentration and mixing with pre-moistened soil (paragraph 19, 20).	See §15
Paragraph 17		
Czech Republic	(and 32) Please change ANNEX to Annex throughout the text	
Germany	4th sentence: The selection of the biggest females violates the statistical principle that a representative (random) sample should be collected. The described selection method may lead to bias in the test results. The authors are encouraged to present a different approach how to make sure that the collembola are sexed correctly.	"biggest" changed to "big"
The Netherlands	1) Editorial comment: in the 3rd line the word 'added' is a bit confusing (as if for *F. candida* in total 10 males and 20 females are to be used). Suggestion to replace	Revised

	'added' with 'used'. 2) In the last sentence should be probably 20/10 as *F.fimetaria* is mentioned before *F.candida* 3) Last sentence: '…..and the biggest females of *F. fimetaria* are selected to ensure a proper distinction from the *F. fimetaria* males.' The question is whether this selection of the biggest females will not influence the test results? From Annex 2 (3rd alinea), it appears that the emphasis in par. 17 on selecting the biggest females may be redundant, since males are reported to be clearly smaller than females. Selecting the biggest of the already bigger females therefore seems a bit redundant and may be an unnecessary influence on the test results.	OK "Biggest" was changed
Switzerland	**Expert 1:** The test is described well (Paragraphs 11-22). No rationale is given for the different ages chosen for the two test species. **Expert 2:** "and the **biggest females** of F. fimetaria are selected to ensure a proper distinction from the *F. fimetaria* males." Care should be taken that a sufficient number of "big" females is available for the whole experiment (and for all the test containers) before the introduction in test containers. Otherwise a variation in the size of the females could be observed between the first test containers and the last containers (the last test containers would have smaller females).	*F. fimetaria* has to older tro enable sexual discrimination OK
Paragraph 18		
Germany	See general comment b) above. (Additional comment: Collembolans don't disappear from the soil surface as quickly as worms do. Consequently springtails may be contaminated directly during application resulting in a type of exposure that is not to be expected under field situations.)	Added: "collembolans should be allowed to enter the soil"
The Netherlands	Including measurements to verify the test concentration in the soil would be a useful addition to the test protocol, especially for (future) pesticide registration procedures.	The OECD test with *E. fetida*, Enchytraeids and *H. aculeifer* did not include such specifications in contrast to the plant test, OECD 208 in the section

		"Verification of test substance concentration/rate" this may be adopted for the springtail guideline.
Paragraph 19		
Czech Republic	There should be specified: The amount of water which is added with the dissolved substance should correspond to the water necessary to adjust soil moisture to 40-60% WHC.	This is implicitly suggested: about half of the water is available! See §15
Switzerland	**Expert 2:** Should the volume of the solution (water + substance) bring the pre-moistened soil to the final moisture content (i.e. complete the volume of water already added in the pre-moistened soil to reach the final water content)? Or should an appropriate amount of water be added to the pre-moistened soil (to reach the moisture content) after the application of the solution?	Yes to question one! But the addition of the compound depends on the chemical properties.
Paragraph 20		
Canada	Commenter 2: Paragraph 20 and 22: Organic solvent evaporation should be conducted in a manner where soil moisture content can be maintained as well.	Quartz sand is used as a carrier, so there will be no loss of water.
Czech Republic	There should be a recommendation that the chemical analysis should be performed to check the real concentration in the soil, because some losses during substance application can be expected.	The OECD test with *E. fetida*, Enchytraeids and *H. aculeifer* did not include such specifications in contrast to the plant test, OECD 208 in the section "Verification of test substance concentration/rate" this may be adopted for the springtail guideline.
Switzerland	**Expert 2:** The use of a fume hood for solvent evaporation could be recommended. More details (as for the 3 other application methods) regarding mixing of the sand with pre-moistened soil should be given as well as precision about the volume of water to be added to reach the final water content.	Volume of water was specified in §14 The evaporation method is optional.
Paragraph 21		
Czech Republic	Why 2.5 g of sand is recommended here, I think "part of the total sand added to the soil" is appropriate enough.	Changed!

Paragraph 22		
Czech Republic	"and then weighted" should be changed to "and then the test containers are weighted"	Revised
Germany	See general comment b) above. (See para 22. comment to Additional comment: It is not clearly specified, whether baker`s yeast as a food source is introduced before or following spray application of the test substance.)	Added. Yeast is kept uncontaminated.
The Netherlands	It should be noted that in the opinion related to the revision of Annex II and III of Directive 91/414/EC, EFSA-PPR does not support tests with sprayed soil surfaces. EFSA-PPR states that '*the risk assessment should be based on concentrations, and it has not been demonstrated that both application methods will yield the same results. From the scarce literature available it appears that the method of application has influence on the response (Van Gestel et al., 1995). The PPR Panel is therefore concerned that the assessment based on overspray is not informative and the decision making between products will not be uniform. It is the view of the PPR Panel that testing should be performed with test substances homogenously incorporated into the soils.*' EFSA. 2007a. Opinion of the Scientific Panel on Plant protection products and their Residues on a request from the Commission related to the revision of Annexes II and III to Council Directive 91/414/EEC concerning the placing of plant protection products on the market - Ecotoxicological studies. The EFSA Journal 461: 1-44.	The surface application is optional. No requirements are stated in the guideline. More general guidelines will provide recommendations as indicated in the comment.
Switzerland	**Expert 2:** "When the test substance is a pesticide, it **may be appropriate to apply it onto the soil surface by spraying**": Only if it is the case during agricultural use "The application should take place at a temperature within ± 2 °C of variation and for aqueous solutions, emulsions or dispersions at a water application rate according to the risk assessment recommendations and the test soil WHC requirements" Unclear. Please reformulate.	OK
BIAC	Test vessel shape or material may not be suitable to act as a liner in the way	These details have been deleted.

	suggested. Give other options (e.g. metal foil) or remove the sentence as it could be restrictive. "Application..... according to the risk assessment recommendations and the test soil WHC requirements". What does this mean? Does the moisture content of the test soil need to be adjusted to take account of the spray volume? This should be quite negligible up to about 400 L/ha, and this could be indicated to avoid unnecessary petty calculations. More consideration of seed treatment/granule/furrow/chisel applications would be useful. These can be problematic with Collembola tests because the test vessels are so small that putting in a single seed or granule or furrow can exceed the realistic treatment rate many-fold and cause toxicity when it would not be seen in reality, and can hold the Collembola closer to the treatment than would be the case in the field with spaces between rows, or seeds sown deeper than 2-4 cm of soil in these test vessels. How should this be dealt with? Larger, deeper vessels with treatment at one edge? What is the largest viable vessel and how many collembola should be used? At least an open option for alternative (higher tier?) designs to deal with such situations could be allowed in the guideline. Test containers left overnight – is this wise? Moisture loss and even Collembola loss could occur, depending on the shape of the test vessels and the environmental conditions.	Deleted. These considerations are presently out of scope of a standard single species test guideline. Collembolans are added after the addition of the test substance.
Paragraph 23		
Canada	Commenter 2: Include a reference or rationale to support the statement "To discourage collembolans from escaping from the soil, the test is carried out under controlled light-dark cycles…".	Sentence deleted
Czech Republic	The first sentence about temperature is confusing: "The test mean temperature must be 20 ± 1 °C with a temperature range of 20 ± 2 °C" So which one is correct?	Please note the difference between a mean and a range!
Germany	"The test mean temperature must be 20 ± 1 °C with a temperature range of 20 ± 2	Please note the difference between a mean and a range!

	°C." change to "The test temperature should be 20 ± 2 °C."	
Switzerland	**Expert 2:** "The test mean temperature must be 20 ± 1 °C with a temperature range of 20 ± 2 °C." …with an **instantaneous temperature range**	
Paragraph 24		
Czech Republic	According to our experience, very high air humidity in the incubator may lead to the negative effects on collembolans. It is probably caused by some infections.	Collembolans need saturated air at least in the soil. This is optional.
The Netherlands	It is advisable to open the test containers at least once a week for aeration; this may be combined with the regular checking of soil humidity and/or feeding.	Agreed
Paragraph 26		
Czech Republic	It should be noticed here how the soil should be placed into the vessels. It should be as a layer on the bottom of the vessel, not as clod, but the soil must not be compacted but keep its natural structure/porosity.	This was described in §11 an §14
Germany	"…an amount of test soil corresponding to 30 g fresh weight…" change to: "… an amount of test soil corresponding to 20 g dry weight…"	30 g moist soil is an appropriate amount
BIAC	Paragraph 26 is not clear if vehicle and water/untreated controls are needed, Paragraph 45 seems to clarify that solvent and plain controls are needed. Best to clarify fully at paragraph 26.	Water control has been specified
Paragraph 27		
Canada	Commenter 1: "For efficient transfer of the animals, a low-suction air flow device can be used." – insert comma	Revised
Paragraph 28		
Canada	Commenter 1: "For the F. fimetaria test, twenty adults,…" – insert comma	Revised
Czech Republic	On paragraphs 28, 29, 30 and 31: 1) There is repeated information "food is added to the test vessels at the beginning of	

	the test and then after about 2 weeks" in paragraph 28 and 29 and then, again, it is mentioned in the paragraph 30. In my opinion all information about food should be concentrated in the paragraph 30 only. 2) In my opinion, all information about the test duration and the extraction should be moved and concentrated in the paragraph 30 only.	Deleted accordingly OK, retained.
Germany	See comment on para 17 (appropriate guidance for discriminating sex in *F. fimetaria*).	OK
Switzerland	**Expert 2:** "For the F. fimetaria test twenty adults": The age should be repeated: between 23-26 days "10 males and 10 females, should be used per test-vessel".: Replace should by must "On day 21 the soil samples should be extracted and counted.": replace by: collembolans should be extracted from the soil samples and counted Last sentence ("For *F. fimetaria* the gender are discriminated…than the males."): Could be inserted in paragraph 31.	Revised
Paragraph 29		
Canada	Commenter 1: • 1st sentence – "For the *F. candida* test, 10 juveniles…" – insert comma • 2nd sentence – reference to food addition "after about 2 weeks" – recommend that the time of food addition be more specific, similar to the recommended *F. fimetaria* procedure; there is no reason for an approximation; the test is only 28 days in duration, so food addition after 14 days is acceptable. • 3rd sentence – "On day 28, the soil samples…" – insert comma	Revised For practicability reasons flexibility is allowed.

		Revised
Switzerland	**Expert 2:** "On day 28 the **soil samples** should be extracted and counted.": replace by: **collembolans** should be extracted from the soil samples and counted.	Revised
Paragraph 30		
Canada	Commenter 1: • 1st sentence – "As a suitable food source, a sufficient…" – insert comma • Provide specificity for time of food addition, similar to *F. fimetaria* procedures, that is, after 14 days.	Revised
Germany	In the beginning of the development of the test with *F. candida* we did some testing for the appropriate amount of yeast to be spread on the substrate surface. 1.5 to 2.5 mg/30 g gave the best results. Adding considerably more yeast granules resulted in stronger growth of fungi on the yeast. The proposed amount therefore should be reduced.	The amount is optional.
The Netherlands	Why is the food added only once per 2 weeks? This seems contradictive with what is stated in Annex 2 (e.g. under 1.c: Fresh food is provided either once or twice a week, to avoid moulding. Or 2.a: Lack of food may lead the animals to defecate on the eggs resulting in fungal growth on the eggs or *F. candida* may cannibalize its own eggs.)	The recommendation is for testing in soil and in Annex 2 is valid for culturing on plaster of Paris.
Switzerland	**Expert 2:** "…beginning of the test and after **about 2 weeks**". Replace by:… after **14 days.**	Flexibility is allowed to aid in practicability and feasibility
Paragraph 31		
Canada	Commenter 1: • 2nd sentence – "After 21…days**, the Collembola** are extracted from the test soil…" – suggested wording change	Revised

Germany	For the heat extraction method, the wording from the new OECD Predatory Mite Guideline should be taken: Extraction efficiency must be validated once or twice a year in controls with known numbers of adults and juveniles. Efficiency should be above 90% on average combined for all developmental stages. The floatation method is not an 'extraction' method and no animal get lost during this process; therefore, it is questionable if this method has to be validated. Is validation of directly counting or manually counting of a digital image necessary? However, for all cases guidance for such an efficiency check should be given in the respective Annex.	This is a quality assurance issue pertaining to the validity of equipment, methods and personnel, which is controlled by GLP compliance requirements, which is implicit in OECD guidelines. Flotation is an extraction method just as heating/drying is an extraction method. The flotation is not 100%. Any method should be validated to be GLP compliant.
The Netherlands	The second sentence is a bit unclear: suggestion to replace the word 'they' with e.g. 'the animals' or 'the adults and juveniles'.	Revised
BIAC	Can advice be added on how to validate this?	Done: "e.g. by adding a known number to soil"
Paragraph 33		
Canada	Commenter 1: • 1st sentence – the comma after "for example" is underlined – this is not necessary	Revised
Czech Republic	Replace "mg/kg (dry weight of soil)" with "mg/kg dry weight soil"	Revised
Germany	The 2nd sentence in para 34 "Additional information, from tests… may also be useful in deciding on the range of concentrations to be used in the range-finding test." Should be included in para 33 instead.	Revised
Switzerland	**paragraphs 33- 38, Expert 1:** I plead for three replicates instead of duplicates. No rationale is given for the difference in test duration between the range finding and definitive test and the different durations for the different species.	Replicate number is optional. Rationale added: "to ensure one clutch of juveniles has been produced". Range finding results are just to aid in selecting concentrations for the definitive test – this is not guideline results so there is no strict requirements.
Paragraph 34		

The Netherlands	It is not clear why the range-finding test should last shorter than the final test. If the test substance causes a delay in hatching then the results of a shorter test can give wrong indication for concentration range in the final test. If the reason is the reduction of work then in my opinion would be better to reduce the number of the replicates in the intermediate concentrations to one. An alternative would be a combined limit/range-finding approach, where for the intermediate concentrations one replicate is used and for the control and the limit concentration eight and four replicates are used, respectively. The study should then last as long as in the final test. The advantage of such approach is reduction of time spend on testing and reduction of usage of the animals	There is no way of ensuring a true result of a range-finding running shorter than the definitive test. The duration of a range-finding is optional. Replicate number is optional.
Paragraph 35		
Czech Republic	Why exactly 1.8 factors are recommended here? I would regard factor of 2 sufficient.	This is inherited from other guidelines. The exact spacing factor size is optional.
BIAC	The proposed spacing factor of 1.8 is too restrictive, instead more flexibility should be allowed and no limits should be fixed. E.g. for substances with flat dose-response curves require larger spacing factors. Guidance can be found in Grimm et al 2001 J. Pest Science 74: 72-84 and should be cited.	Now changed to: "The spacing factor may vary depending on the dose-response pattern"
Germany	In regulatory ecotoxicology, EC10 is generally considered as a suitable surrogate of the NOEC. It is difficult to understand why the determination of an EC10 should use a totally different testing scheme than that of a NOEC determination. No explanation is provided why 12 (!) test concentrations to be used for ECx determination; we strongly recommend to keep to the "usual design" of at least 5 concentrations. Similarly, no explanation is provided why spacing factors are allowed to vary over the tested concentration range. We believe that these types of "explorative suggestions" undermine the existing paradigms of ecotoxicological tests as validated over many years within the OECD test guidelines programme. We believe that the sections 35-38 need to be (re-thought and) re-written, with the aim to provide a robust	This was inherited from the recently accepted OECD guideline 226. At least 5 concentrations is feasible too. As this approach seems to have been accepted we leave to general updating of the guideline to come up with harmonized and updated methodologies.

	guidance how to design chronic tests with collembolans, irrespective of ECx or NOECs (or mixed approaches) to be determined. More specifically: Why are 12 test concentrations proposed for the derivation of an ECx in the one case (ECx-design / para 35) and only eight test concentrations in the other case (mixed ECx/NOEC-design / para 37)? The statistical reasoning behind this proposal remains cryptic.	
Paragraph 35		
Canada	Commenter 1: • 1st sentence – "If no effects…1000 mg/kg), the reproduction test…" – insert comma	Revised
Paragraph 37		
Switzerland	**Expert 2:** " For the **two** approaches…": replace by: for this **combined** approach…	Revised
BIAC	2nd sentence should say "For the combined approach..." How is 8 concs with 4 reps equivalent to 12 concs with 2 reps at parag 35? This should mean that it is also an option to use 8 concs of 4 reps at parag 35?	Revised
Paragraph 38		
BIAC	It should not be necessary to specify the limit test concentration here. It may not be necessary for the risk assessment to have a NOEC at 1000 mg/kg. For example, if a NOEC of 100 mg/kg is ample to pass all risk assessments with a test substance and is likely from literature, then the limit test could be done at 100 mg/kg. The 1000 mg/kg upper dose in the rangefinder is only an example.	The guideline just gives figures as examples. The approach mentioned is not prevented by the guideline
Paragraph 39		

Canada	Commenter 1: • 1st sentence – "**The main test endpoint** is reproductive output." – suggested re-phrasing • 3rd sentence – "…ANOVA procedures such as the Student t-test, Dunnett's test, or Williams' test…"	Revised
Switzerland	**Expert 1:** The proposed statistics is valid.	OK
Czech Republic	Mortality is the endpoint of the test too, and statistics for this endpoint should be done the same way as for reproduction. Many statistical softwares do not need means and variance as input for ANOVA or range tests. Rather original results of individual replicates serve as input. Thus, I would mention here only the treatment of primary results as mean and variance calculation. All other statistics should be discussed in the next paragraph (e.g. 40)	Included in §41 Means and variance terms are deleted.
Paragraph 40		
Canada	Commenter 1: • 1st sentence – "The number of surviving females in the untreated controls…" – recommendation to change to "**number of surviving adults**" to include the total number of adults added for *F. fimetaria*, which include both females and males – the total number surviving is also reported in Krogh (2008) for both species, and not on a female basis (i.e., for *F. fimetaria*; note that the majority of *F. candida* will likely be female).	Revised
Czech Republic	This should be moved the part describing reporting – paragraph 46	Retained
The Netherlands	First sentence: It is unclear why only the number of females is mentioned. For *F. fimetaria*, should not also the males be included in the untreated control?	Yes. Done
Switzerland	**Expert 2:** It is said that "the number of surviving females in the untreated controls is a major validity criterion". However, this is not specified in the validity criteria given for the test (paragraph 9). In the validity criteria it is specified that the **mean adult mortality** should not exceed 20% at the end of the test. Precisions and explanations	Changed to "adults"

	should be given. Does it mean that only female mortality has to been assessed?	
Paragraph 41		
Canada	Commenter 1: • 1st sentence – "**The ECx, including the associated**…for the parameter, are calculated…"	Revised
Czech Republic	Last sentence: Regressions often use all data (i.e. results of individual replicates) which enables to include the variance within the same concentration into the total error of the model. Thus, I would not limit the regression only for calculated means.	Yes, true. Changed.
Germany	(Applies to surface application only:) It is questionable whether the ECx and NOEC are the appropriate abbreviations of the toxicity measures, since the collembolans are not really exposed to a concentration like aquatic organisms. The 'effective application rate, ERx' or the 'no-observed effect rate/level, NOER or NOEL' should be preferred.	NOEC and ECx is used by convention. Pertains to generally updating OECD guidelines
Paragraph 42		
Germany	If it is the intention to test substances with hormonal activity, one might a priori assume reproduction either to go up or to go down. There are certainly cases where one cannot anticipate if reproduction will be enhanced or impaired. A double-sided hypothesis testing may be more appropriate in these cases. Again, it is strongly recommended to enlarge the knowledge on the experimental effects of those compounds before all important details of the test procedure will be finally fixed.	This is a comment that pertains to many OECD test guidelines. The present collembolan guideline is crafted from existing accepted guidelines.
Paragraph 43		
Canada	Commenter 1: • 2nd sentence – "Conchran's test, Levene's test, …"	Revised
Czech Republic	This paragraph should be rewritten and carefully revised by some statistician. E.g.: Shapiro-Wilk test at $p<0.05$ as mentioned here means NOT normal distribution and test of variance homogeneity at $p<0.05$ means NOT equal variances. It is better to	The indicated thresholds have been deleted. This § was inhereted from previous

	specify alpha error as criterion than p <>.	guidelines.
Germany	Since according to the OECD Guidance Document No. 54 the Shapiro-Wilk and Levene test are the most preferable test on normal distribution and variance homogeneity as well as the the R/s test is a very rough method the first two sentences should be reworded: "Normal distribution and variance homogeneity can be tested using an appropriate statistical test, e.g. the Shapiro-Wilk test and Levene test, respectively ($p \leq 0.05$)." Further the phrase: "…or step-down trend tests (e.g. Williams' test in case of a monotonous dose-response relationship) can be used…". is misleading since the "monotonous dose-response relationship" related to the theoretical type of response in the sample population. Due to the sampling error in a given experiment e.g. the monotonous decrease in a variable can be violated. And this is exactly the reason why Williams developed his test, in which the maximum-likelihood means are used with testing which replace the original means. The maximum-likelihood means are obtained by a moving average method which makes the series of responses monotonously decreasing (or increasing depending on the endpoint). In the case under consideration, the principally monotonous relationship can be assumed.	Deleted "in case of a monotonous dose-response relationship"
Paragraph 46		
Canada	Commenter 1: • 2nd bullet in "Test results" – "number of adult females and adult mortality…" – suggest change to "**number of adults** and adult mortality" to accommodate for the total number of adults added to a test for *F. fimetaria*, which includes both females and males.	Agree. Text revised.
Czech Republic	The formulation of the test substance and the additives should be specified if not the pure chemical is tested (e.g. pesticides) The sentence: „The LC50, the NOEC and/or ECx (e.g. EC50, EC10) for reproduction	Included Revised accordingly

	with 95% confidence intervals, and a graph of the fitted model used for its calculation, the slope of the dose-response curve and its standard error (See [9])“ should be replaced with „The NOEC values, LCx for mortality and ECx for reproduction (mostly LC50, LC10, EC50, and EC10) together with 95% confidence intervals. A graph of the fitted model used for calculation, its function equation and its parameters“. The slope of the dose-response curve is not the only parameter of the model, it is dependent on the function used.	
France	In the ”test results” part of the report, the validity of the test according to the validity criteria should be added.	Added
Germany	Test results: In case the NOEL is determined, in accordance to reference 14 (OECD Guidance Document No. 54), the MDD (minimal detectable difference) between treatment and control should be given in order to give a hint on the test power and the risk of a type-II error.	Added
The Netherlands	Test results, second point: Why are only the females mentioned? For *F. fimetaria*, should not also the surviving adult males be included in the test results?	Changed to “adults”
BIAC	“- physico-chemical properties of the test substance….., **if available**.” (add: if available)	Added
LITERATURE		
Switzerland	**Expert 1:** Few references are cited, however the test is based on other existing OECD protocols, therefore it is not needed to always cite the original references	Retained
BIAC	(13) Fjelberg 1980: provide full citation	Included
Annex 1		

The Netherlands	The schedule is a bit misleading. On day -9 to -12 a preparation of synchronous *F.candida* culture should start. According to the proposal, 9-12 day old animals should be used in test. So the timing given is the emergence of first offsprings. The start of the synchronous cultures should then be scheduled for day -23 to -26 for *F. candida.*	Correct. Changed.
Switzerland	**Expert 2:** Here it is specified soil is pre-moistened 14 days before the beginning of the test. In paragraph 15 it is specified artificial soil is pre-moistened 2-7 days before the test start. Please check. Range finding tests do not have the same duration for F. fimetaria and F.candida: Please precise that after 14 days the end of the range-finding test is for F. fimetaria and add range-finding test actions for F. candida. Day -2 to -1: Sorting Collembola for testing and day -1: Distribute into batches: that mean that 2 transfer occur. Is that not too much stress? When are water soluble substances applied? (day-1 or day 0?)	True. Corrected Done "Sorting deleted" Day 0. Now specified
Annex 2		
Canada	Commenter 1: • 1st paragraph, 2nd sentence – "Basically, the incidence of oviposition…" – insert 'the' • Section 2, 2.a. Synchronisation 7th bullet, 2nd sentence – "In particular for *F. fimetaria*, the following measures should be taken…" – insert comma	Revised
Czech Republic	The second and third paragraphs are related to F. fimetaria only, whilst it should be the description for both species.	*F. fimetaria* deleted – valid for *F.candida* too.
The Netherlands	Under paragraph 2, point 2a, 4th bullet: collecting eggs may be time-consuming and may also cause damage to the eggs. Alternatively, adult animals may be removed from the synchronization culture boxes after 2 or 3 days of producing eggs. In that	This alternative approach has been added.

	way, the eggs don't have to be touched.	
Switzerland	**Expert 2:** Would it not be more appropriate to use bigger breeding containers instead of Petri dishes (90 x 15 mm). This will reduce the number of the breeding containers of the stock culture (not so time consuming). By using Petri dishes as breeding containers, how do you resolve the problem of "Collembola escape" from the Petri dishes? 1. Rearing It would also be appropriate to give the "proportion" of the substrate constituents : "The culturing substrate is plaster of Paris (calcium sulphate) with activated charcoal (mass ratio 8:1) 1.b. Breeding "Collembolans are held in containers such as Petri dishes (90 mm x 13 mm)" : replace by: …Petri dishes (about 90 mm x 15 mm) 1.c. Food source A sentence about removing of uneaten beaker's yeast should be added. 2a. Synchronisation "After 10 days the eggs are carefully collected with a needle and spatula and moved to "egg-paper" (small pieces of filter paper dipped in plaster of Paris/charcoal slurry) which is placed in a container with fresh plaster/charcoal substrate. A few grains of yeast are added to attract the juveniles and make them leave the egg-paper". Precise that: …grains of yeast are added on the substrate of the container to attract…	Collembolans do not really escape from Petri dishes. Using Petris dishes is just a suggestion, obviously. Already included for mouldy yeast Done
BIAC	1A: Clarify that the mixtures are poured and allowed to set before being moistened with water. The text suggests that the PoP mixture is used wet, without setting. 1B: How long can each such container be maintained. Is a replacement cycle necessary? 1C: Should say "fresh food is provided once or twice a week", i.e. remove "either" as	Done Depends on the preferences of the lab. Breeding is described not just keeping a stock culture.

	this is too strict.	Done
Annex 3		
Czech Republic	Part 1.a. How to prevent collembolans from escaping the test vessel and collecting vessel? There is not mentioned that the soil must be transferred to some container with the mesh in the bottom! Part 1.b. Replace the second sentence „This involves emptying the tube of soil into a 250 ml vessel and adding approx. 200 ml of distilled water“ with „For that purpose the test is performed in the vessels of approximately 250 ml volume. At the end of the test approx. 200 ml of distilled water are added“	The extractor prevents collembolans to do so! The guideline does not include the constrution of extraction equipment. Done
Germany	1.a. Last sentence: "… is increased automatically every 12 h by 5°C. After 12 h at 40°C the extraction is finished." (Please clarify: i.e. 48 h total extraction time.)	Changed
Annex 4		
Czech Republic	Reference (3) is not correct here.	Reference deleted
BIAC	Is there a need to repeat the methods here? This means the current guidelines would be outdated if the other guidelines were amended. It may be better just to reference them accurately.	The method is completely described in annex 4 and do not need the reference. The method is sufficient even if the ISO guideline is changed.
Annex 5		
Czech Republic	Reference (15) is not correct here.	