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

31-Aug-2007

English - Or. English

ENV/JM/MONO(2007)26
Unclassified

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

SERIES ON TESTING AND ASSESSMENT
Number 79

VALIDATION REPORT OF THE FULL LIFE-CYCLE TEST WITH THE HARPACTICOID COPEPODS NITOCRA SPINIPES AND AMPHIASCUS TENUIREMIS AND THE CALANOID COPEPOD ACARTIA TONSA – PHASE 1

JT03231276

English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 79

VALIDATION REPORT OF THE FULL LIFE-CYCLE TEST WITH THE HARPACTICOID COPEPODS NITOCRA SPINIPES AND AMPHIASCUS TENUIREMIS AND THE CALANOID COPEPOD ACARTIA TONSA – PHASE 1

Environment Directorate

Organisation for Economic Co-operation and Development

2007

Also published in the Series on Testing and Assessment:

No. 1, *Guidance Document for the Development of OECD Guidelines for Testing of Chemicals (1993; reformatted 1995, revised 2006)*

No. 2, *Detailed Review Paper on Biodegradability Testing (1995)*

No. 3, *Guidance Document for Aquatic Effects Assessment (1995)*

No. 4, *Report of the OECD Workshop on Environmental Hazard/Risk Assessment (1995)*

No. 5, *Report of the SETAC/OECD Workshop on Avian Toxicity Testing (1996)*

No. 6, *Report of the Final Ring-test of the Daphnia magna Reproduction Test (1997)*

No. 7, *Guidance Document on Direct Phototransformation of Chemicals in Water (1997)*

No. 8, *Report of the OECD Workshop on Sharing Information about New Industrial Chemicals Assessment (1997)*

No. 9, *Guidance Document for the Conduct of Studies of Occupational Exposure to Pesticides during Agricultural Application (1997)*

No. 10, *Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data (1998)*

No. 11, *Detailed Review Paper on Aquatic Testing Methods for Pesticides and industrial Chemicals (1998)*

No. 12, *Detailed Review Document on Classification Systems for Germ Cell Mutagenicity in OECD Member Countries (1998)*

No. 13, *Detailed Review Document on Classification Systems for Sensitising Substances in OECD Member Countries 1998)*

No. 14, *Detailed Review Document on Classification Systems for Eye Irritation/Corrosion in OECD Member Countries (1998)*

No. 15, *Detailed Review Document on Classification Systems for Reproductive Toxicity in OECD Member Countries (1998)*

No. 16, *Detailed Review Document on Classification Systems for Skin Irritation/Corrosion in OECD Member Countries (1998)*

No. 17, *Environmental Exposure Assessment Strategies for Existing Industrial Chemicals in OECD Member Countries (1999)*

No. 18, *Report of the OECD Workshop on Improving the Use of Monitoring Data in the Exposure Assessment of Industrial Chemicals (2000)*

No. 19, *Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation (1999)*

No. 20, *Revised Draft Guidance Document for Neurotoxicity Testing (2004)*

No. 21, *Detailed Review Paper: Appraisal of Test Methods for Sex Hormone Disrupting Chemicals (2000)*

No. 22, *Guidance Document for the Performance of Out-door Monolith Lysimeter Studies (2000)*

No. 23, *Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures (2000)*

No. 24, *Guidance Document on Acute Oral Toxicity Testing (2001)*

No. 25, *Detailed Review Document on Hazard Classification Systems for Specifics Target Organ Systemic Toxicity Repeated Exposure in OECD Member Countries (2001)*

No. 26, *Revised Analysis of Responses Received from Member Countries to the Questionnaire on Regulatory Acute Toxicity Data Needs (2001)*

No 27, *Guidance Document on the Use of the Harmonised System for the Classification of Chemicals Which are Hazardous for the Aquatic Environment (2001)*

No 28, *Guidance Document for the Conduct of Skin Absorption Studies (2004)*

No 29, *Guidance Document on Transformation/Dissolution of Metals and Metal Compounds in Aqueous Media (2001)*

No 30, *Detailed Review Document on Hazard Classification Systems for Mixtures (2001)*

No 31, *Detailed Review Paper on Non-Genotoxic Carcinogens Detection: The Performance of In-Vitro Cell Transformation Assays (2007)*

No. 32, *Guidance Notes for Analysis and Evaluation of Repeat-Dose Toxicity Studies (2000)*

No. 33, *Harmonised Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures (2001)*

No. 34, *Guidance Document on the Development, Validation and Regulatory Acceptance of New and Updated Internationally Acceptable Test Methods in Hazard Assessment (2005)*

No. 35, *Guidance notes for analysis and evaluation of chronic toxicity and carcinogenicity studies (2002)*

No. 36, *Report of the OECD/UNEP Workshop on the use of Multimedia Models for estimating overall Environmental Persistence and long range Transport in the context of PBTS/POPS Assessment (2002)*

No. 37, *Detailed Review Document on Classification Systems for Substances Which Pose an Aspiration Hazard (2002)*

No. 38, *Detailed Background Review of the Uterotrophic Assay Summary of the Available Literature in Support of the Project of the OECD Task Force on Endocrine Disrupters Testing and Assessment (EDTA) to Standardise and Validate the Uterotrophic Assay (2003)*

No. 39, *Guidance Document on Acute Inhalation Toxicity Testing (in preparation)*

No. 40, *Detailed Review Document on Classification in OECD Member Countries of Substances and Mixtures Which Cause Respiratory Tract Irritation and Corrosion (2003)*

No. 41, *Detailed Review Document on Classification in OECD Member Countries of Substances and Mixtures which in Contact with Water Release Toxic Gases (2003)*

No. 42, *Guidance Document on Reporting Summary Information on Environmental, Occupational and Consumer Exposure (2003)*

No. 43, *Draft Guidance Document on Reproductive Toxicity Testing and Assessment (in preparation)*

No. 44, *Description of Selected Key Generic Terms Used in Chemical Hazard/Risk Assessment (2003)*

No. 45, *Guidance Document on the Use of Multimedia Models for Estimating Overall Environmental Persistence and Long-range Transport (2004)*

No. 46, *Detailed Review Paper on Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances (2004)*

No. 47, *Detailed Review Paper on Fish Screening Assays for the Detection of Endocrine Active Substances (2004)*

No. 48, *New Chemical Assessment Comparisons and Implications for Work sharing (2004)*

No. 49, *Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs (2004)*

No. 50, *Report of the OECD/IPCS Workshop on Toxicogenomics (2005)*

No. 51, *Approaches to Exposure Assessment in OECD Member Countries: Report from the Policy Dialogue on Exposure Assessment in June 2005 (2006)*

No. 52, *Comparison of emission estimation methods used in Pollutant Release and Transfer Registers (PRTRs) and Emission Scenario Documents (ESDs): Case study of pulp and paper and textile sectors (2006)*

No. 53, *Guidance Document on Simulated Freshwater Lentic Field Tests (Outdoor Microcosms and Mesocosms) (2006)*

No. 54, *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application (2006)*

No. 55, *Detailed Review Paper on Aquatic Arthropods in Life Cycle Toxicity Tests with an Emphasis on Developmental, Reproductive and Endocrine Disruptive Effects (2006)*

No. 56, *Guidance Document on the Breakdown of Organic Matter in Litter Bags (2006)*

No. 57, *Detailed Review Paper on Thyroid Hormone Disruption Assays (2006)*

No. 58, *Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals (2006)*

No. 59, *Report of the Validation of the Updated Test Guideline 407: Repeat Dose 28-Day Oral Toxicity Study in Laboratory Rats (2006)*

No. 60, *Report of the Initial Work Towards the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1A) (2006)*

No. 61, *Report of the Validation of the 21-Day Fish Screening Assay fort he Detection of Endocrine Active Substances (Phase 1B) (2006)*

No. 62, *Final OECD Report of the Initial Work Towards the Validation of the Rat Hershberger Assay: Phase-1, Androgenic Response to Testosterone Propionate, and Anti-Androgenic Effects of Flutamide (2006)*

No. 63, *Guidance Document on the Definition of Residue (2006)*

No. 64, *Guidance Document on Overview of Residue Chemistry Studies (2006)*

No. 65, *OECD Report of the Initial Work Towards the Validation of the Rodent Uterotrophic Assay – Phase One (2006)*

No. 66, *OECD Report of the Validation of the Rodent Uterotrophic Bioassay - Phase 2: Testing of Potent and Weak Oestrogen Agonists by Multiple Laboratories (2006)*

No. 67, *Additional Data Supporting the Test Guideline on the Uterotrophic Bioassay in Rodents (2007)*

No. 68, *Summary Report of the Uterotrophic Bioassay Peer Review Panel, including Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2006)*

No 69, *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models (2007)*

No. 70, *Report on Preparation of GHS Implementation by the OECD Countries (2007)*

No. 71, *Guidance Document on the Uterotrophic Bioassay – Procedure to Test for Antioestrogenicity (2007)*

No. 72, Guidance Document on Pesticide Residue analytical Methods (2007)

No. 73, *Report of the Validation of the Rat Hershberger Assay: Phase 3: Coded testing of Androgen Agonists, Androgen Antagonists and Negative Reference Chemicals by Multiple Laboratories. Surgical Castrate Model Protocol (2007)*

No. 74, *Detailed Review Paper for Avian Two-generation Testing (2007)*

No. 75, *Guidance Document on the Honeybee (Apis Mellifera L.) Brood Test under Semi-field Conditions (2007)*

No. 76, *Final Report of the Validation of the Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances: Phase 1 – Optimisation of the Test Protocol (2007)*

No. 77, *Final Report of the Validation of the Amphibian Metamorphosis Assay: Phase 2 – Multi-chemical Interlaboratory Study (2007)*

No. 78, *Final Report of the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances. Phase 2: Testing Negative Substances (2007)*

No. 79, *Validation Report of the Full Life-cycle Test with the Harpacticoid Copepods Nitocra Spinipes and Amphiascus Tenuiremis and the Calanoid Copepod Acartia Tonsa – Phase 1 (2007)*

ABOUT THE OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 30 industrialised countries in North America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in ten different series: Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides and Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and the Safety of Manufactured Nanomaterials. More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (http://www.oecd.org/ehs/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The participating organisations are FAO, ILO, OECD, UNEP, UNIDO, UNITAR and WHO. The World Bank and UNDP are observers. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

FOREWORD

This document presents the report of the Phase 1 of the validation of the copepod development and reproduction test.

A Standard Project Submission Form, proposing to develop a Test Guideline on the copepod development and reproduction test, was presented by Sweden at the 14th meeting of the Working Group of National Coordinators of the Test Guidelines Programme (WNT) in 2002. The project was included on the work plan of the Test Guidelines Programme.

An OECD Invertebrate Expert Group was established in 2002 to follow the development and validation of test methods on invertebrate species. The group met for the first time in October 2003 and developed a proposal for the pre-validation of the test, using 3 species of harpacticoid copepods and one species of calanoid copepods, both groups of copepods being of ecological interest. Pre-validation work was led by Sweden and Denmark. The outcome of the pre-validation was presented to the Invertebrate Expert Group for discussion at its second meeting in November 2005. The protocol was streamlined for the validation exercise and a reference substance: 3,5-dichlorophenol was chosen by the Invertebrate Expert Group.

The Phase 1 experimental work was undertaken on two species of harpacticoid copepods (*Amphiascus tenuiremis* and *Nitocra spinipes*) and on one species of calanoid copepods (*Acartia tonsa*). The outcome of the validation was presented to the Validation Management Group on Ecotoxicity testing (VMG-eco) in January 2007. The Task Force on Endocrine Disrupters Testing and Assessment and the WNT endorsed the report in March 2007. The validation work will continue in 2007-2008 to fully demonstrate the reproducibility of the test.

This document is published on the responsibility of the Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology.

TABLE OF CONTENTS

ACKNOWLEDGEMENT

This report is the fruit of collaboration between laboratories in several countries:

- Department of Applied Environmental Science (ITM), Stockholm University, Sweden
- IVL Swedish Environmental Research Institute, Sweden
- Toxicon AB, Sweden,
- Block Environmental Inc., Pleasant Hill, California, USA,
- NOAA Center for Coastal Environmental Health and Biomolecular Research (CCEHBR), Charleston, South Carolina, USA,
- Institute of Environment & Resources, Technical University of Denmark,
- DHI Water and Environment (DHI), Agern Allé 5, 2970 Hørsholm, Denmark,
- Opus Plus Ltd, (OPUS), Flotta, Stromness, Orkney, KW16 3NP, United Kingdom
- ECT Oekotoxikologie GmbH, Boettgerstr, Germany,

and has been prepared by [1]M. Breitholtz, [1]M. Linde [1]J. Minten and [2]J-O. Persson.

[1]Department of Applied Environmental Science (ITM), Stockholm University, S-106 91 Stockholm, Sweden

[2]Mathematical Statistics, Stockholm University S-106 91, Stockholm, Sweden

ABSTRACT

With the purpose to develop an OECD Test Guideline (TG) for full-life cycle tests with sexually reproducing small crustaceans, ring test activities with the two harpacticoid copepods *Nitocra spinipes* and *Amphiascus tenuiremis* (OECD 2005b) and the calanoid copepod *Acartia tonsa* (OECD 2005a) were carried out during 2006 with participation of 10 laboratories from Denmark (2), Germany (1), UK (1), USA (3) and Sweden (3). Eight of these laboratories contributed with developmental data, whereas only six laboratories contributed with reproduction data. 3,5-dichlorophenol (DCP) was used as a single reference substance in the following concentrations: 6.6, 20, 60, 180 and 540 μg/L.

The present ring test activity shows that most of the participating laboratories have succeeded to follow one of the two draft TGs and deliver both developmental and reproduction data, which have been treated and analysed at the Department of Applied Environmental Science (ITM) and Mathematical Statistics at Stockholm University, Sweden. However, the chemical analyses, which were also performed at ITM, showed that there was a significant decline of 3,5-DCP-levels in the test medium over time for all laboratories, which makes it difficult to draw thorough conclusions about exact concentration effect levels. Nevertheless, since there was a consistent decline evident for all laboratories, comparisons of LOECs were still possible based on the initial measured concentrations.

The interpretational difficulties due to the exposure declines aside, the developmental data were generally consistent among all species and laboratories. For the harpacticoids, the LOEC for development ranged from 6.6 to 20 μg/L whereas for the calanoids the range was somewhat larger between the three participating laboratories; i.e. 6.6 to 180 μg/L. For the harpacticoids the reproduction data was also consistent, at least for two out of three "reproductive" endpoints (i.e. "time to first clutch" and "offspring production"). The reproductive data for the calanoids varied more, with LOECs ranging from 6.6 to 540 μg/L. Overall, the LOECs for the developmental data were lower than the corresponding LOECs for reproduction. Whether this actually reflects the sensitivity of the species exposed to 3,5-DCP in the present ring test is difficult to say since the concentrations were not stable over time. On the other hand, the chemical analyses of the test medium samples for some laboratories showed that the levels of 3,5-DCP increased as the reproductive phase was reached, but the LOEC for development was still consistently lower for the developmental endpoints.

The test time for the two laboratories that ran the harpacticoid full lifecycle test using *N. spinipes* was over 40 days, which indicates that this species should not be tested over the full life cycle. However, the test time for *A. tenuiremis*, for which the current experimental design has been more commonly used, the lifecycle test time was about 25 days. This indicates that *A. tenuiremis* is more appropriate to use for full life cycle testing in the sense that about 15 days reduced test time will significantly reduce the costs of testing.

Taking into consideration the above mentioned findings, and if it is preferred that all the species used in the current ring tests are to be included in further validation work, then it is recommended that a partial life cycle test, which focuses on the larval and juvenile developmental phases, be used. However, if there is considered to be a greater regulatory need for a full life cycle test which includes reproduction endpoints, *A. tenuiremis* should be used as a single harpacticoid species, owing to its much shorter test time and more consistent reproductive output and success. For the calanoid *A. tonsa*, the large inter-

laboratory variability for reproduction illustrates that a partial life cycle test may be a better alternative. Other additional reference chemicals should be used for continuing validation work; at least until the reason(s) for the 3,5-DCP concentration decline using the current brackish and marine species has been thoroughly investigated. More laboratories should also participate in the ring test activities to avoid random errors and variability that may influence the overall judgement of the ring test results. Finally, in order to estimate more statistically robust EC_{50} values instead of LOEC/NOECs, non-linear models should be fitted to the harpacticoid and calanoid (partial) life cycle data. The Phase 2 validation work should consider the difficulties encountered in the Phase 1 studies in the planning of the test protocol and overall testing program.

1. INTRODUCTION

1.1 Rationale for the Proposed Test Method Validation

Copepods in general are often key species in freshwater as well as marine ecosystems. They constitute the food and energy link between primary producers and higher-level heterotrophic organisms such as macro crustaceans and fish larvae. Thus, they play an important role in the function of aquatic ecosystems.

Harpacticoid copepods comprise more than 3.000 species, are mostly free-living benthic organisms, and are usually the second most abundant group of animals (after nematodes) in marine benthic communities (Huys *et al.*, 1996). They almost always dominate the gut contents of bottom or phytal feeding larval and juvenile fish (Hicks and Coull, 1983). They generally have short generation times, which mean that they can be easily studied for the duration of a full life cycle, focusing on all crucial ontogenetic stages and life history traits, such as larval development, reproduction, etc.; this is essential for characterization of both lethal and sublethal effects of toxicants (Ingersoll *et al.*, 1999; Preston and Snell, 2001). As copepods in general, harpacticoid copepods have a sexual reproduction; the copulation almost always involves the attachment of a sperm sac - *spermatophore* - by the male to the copulatory pore of the female, immediately after the last moult (Wesenberg-Lund, 1937; Huys *et al*., 1996). Consequently, harpacticoid copepods have become one of the most studied metazoan groups in chronic ecotoxicity tests (*e.g.* Hutchinson *et al.*, 1999a, b; Chandler et al. 2001; 2004a, b; Breitholtz and Bengtsson 2001; Breitholtz and Wollenberger 2003; Breitholtz et al. 2003). Also, an international standard method (ISO 1997) for acute toxicity testing with both harpacticoid and calanoid copepods exists since 1997, including the species *Acartia tonsa, Tisbe battagliai* and *Nitocra spinipes*. Within the regime of the ASTM (American Society of Testing and Materials) a full life cycle test using *Amphiascus tenuiremis* has also been published recently (ASTM 2004).

At the moment there are no internationally harmonized (i.e. OECD) chronic toxicity test methods for marine invertebrates, although they are an ecologically important and large group of organisms that need to be protected. Therefore, the development of a test on reproduction and development of marine copepods has been added to the work plan of the OECD Test Guideline Program. The copepod life-cycle test is urgently needed in several multi-national regulatory systems. It will be of wide applicability and as such facilitate international harmonization of risk and hazard assessments. It addresses important endpoints and environmental compartments (marine/estuarine/brackish) that are poorly covered by existing methods.

Two common and regularly used species in toxicity tests were originally proposed (SPSF, 2002) as model test organisms, namely the harpacticoid copepods *Nitocra spinipes* and *Tisbe battagliai*. Later one further harpacticoid species, *Amphiascus tenuiremis, and* the calanoid copepod *Acartia tonsa* were included. The new guidelines were intended for evaluation of adverse long-term effects (incl. endocrine disruptive effects) of various types of chemicals to aquatic invertebrates – e.g. industrial chemicals, pesticides and pharmaceuticals, as well as compounds used in off shore oil industry.

1.2 Experiences from the pre-validation using 3,5-Dichlorophenol

At the time of the copepod test pre-validation, the draft OECD guideline for life-cycle tests also included the harpacticoid copepod *Tisbe battagliai*. Due to difficulties in finding interested laboratories for the ring-test activities, this species was excluded after the pre-validation (OECD 2005c).

Two types of copepod test methods were conducted by five laboratories using 3,5-dichlorophenol as a reference compound. Other candidate substances (e.g. fipronil, bisphenol-A, fenoxycarb and methoprene), which were believed to serve as possible reference compounds with a specific mode of action, were also tested using *N. spinipes,* but none of these candidate chemicals fulfilled the criteria for suitable reference substances (OECD 2005c). Fenarimol, which has been suspected of having ecdysteroid disruptive effects in freshwater daphniids (Mu et al 2000; Mu and LeBlanc 2002, 2004) was later also excluded from this list owing to "lack" of effects in *A. tenuiremis* (T. Chandler personal communication).

1.2.1 Acute toxicity

For both calanoids and harpacticoids (considering late juvenile or adult stage) the acute toxicity (96-hr LC_{50}) of 3,5-DCP is in the range of 1 mg l^{-1} In *A. tenuiremis*, the most sensitive life stage was the nauplius stage, which resulted in a 96-hr LC_{50} of 0.38 mg l^{-1} (95%-CI= 0.16-0.54 mg l^{-1}). This difference in juvenile and adult sensitivity to 3,5-DCP was considered when the concentration range for the full life cycle tests was set (see 1.2.2 - 1.2.4. below).

1.2.2 Development

To compare the sensitivity related to development between *A. tonsa* and the harpacticoid species, we have to make clear that *larval/juvenile development ratio*, which is often a sensitive test variable in toxicity testing using copepods, is measured in two different ways in these groups of species. This is due to differences in the experimental set up; *A. tonsa* is exposed in cohorts and the morphological distinction between the nauplius and copepodite stages (metamorphosis occurs between these two life stages) is used to measure effects on *development* (termed larval development ratio). The *larval development ratio* gives a picture of the mean developmental status among a large group of animals, but may not necessarily represent the mean development *rate* since only the *ratio* between the nauplius and copepodite stages at an arbitrarily chosen point in time is calculated. In the harpacticoid tests exposed and control copepods are held and tested individually. This means that the exact number of days for an individual to develop into the copepodite stages or the adult stage, respectively, is observed and this represents a rate. The larval development ratio may be presented using such data, but since the copepods are kept individually, one must calculate the ratio of copepodite- to naupliar-stage individuals in each replicate microplate at some fixed time interval (e.g., at day 6 for *N. spinipes* and at day 7 for *A. tenuiremis*) averaged across microplates. Consequently, such data collected at the fundamental level of replication (i.e. the microplate or the beaker for calanoids) may be used to do further statistical analysis. The number of microplate or beaker replicates used, the number of copepods tested, and the statistical sampling error (variance) will determine statistical power to reject the null hypothesis of no toxicant concentration effect relative to the toxicant-free control.

If stage-specific mortality from a test compound is the same for calanoids and harpacticoids, then these two test endpoint variables indicate the same types of effects on development (a non-mortality based endpoint). It is therefore possible to compare the lowest observed effect concentrations (LOECs) for larval development ratio (in F_0-calanoids or harpacticoids) and/or mean development rates (in F_0-harpacticoids). The results from the tests with 3,5-DCP show that there are species differences but these differences cannot be generalized to indicate different sensitivity in the two taxonomic groups (i.e. calanoids vs harpacticoids). In *A. tonsa* the two (DTU vs DHI) LOECs for inhibition of larval development ratio was

100 and 220 µg/L, respectively. Among the harpacticoids, the LOEC for the test with *A. tenuiremis* was 18.4 µg/L, which caused a significant delay in the development from the first nauplius stage to the first copepodite stage. Mean larval and juvenile developmental rate was also suppressed in *N. spinipes* but only in the highest concentration tested (i.e. 200 µg/L). In *T. battagliai* none of the two test rounds showed significant responses on delayed larval development up to and including 560 µg/L. This means that related to delay of mean larval/juvenile development, *A. tenuiremis* was the most sensitive species, followed by *A. tonsa* and *N. spinipes*. *T. battagliai* was the least sensitive species in these tests.

1.2.3 Reproduction

Since the test variables related to reproduction differ widely between the two protocols, it is difficult to make comparisons. In the harpacticoid protocol, the main test variables related to reproduction are *fertilization success* (i.e. number of mating pairs able to produce two successive clutches), *total viable offspring production per mating pair* (i.e. offspring through two clutches averaged over the number of mating pairs created and tested), *time to production of first clutch* and *time interval between successive clutches*. In *A. tonsa*, it is mainly the *egg production rate* that is used as a measure of reproductive effects. However, the use of this test variable with *A. tonsa* caused some concerns because of the low number of females present in each group making the statistics quite uncertain due to large confidence intervals. Thus, it was not possible from these experiments to evaluate if this was a proper test variable to examine with *A. tonsa*. As a consequence the method for measuring egg production is changed in the next draft guideline.

As for reproduction in the harpacticoid copepods, none of the < 200 µg/L 3,5-DCP concentrations had significant effects on fertilization success or total viable offspring production per female in *A. tenuiremis*. However, the reproductive success (and thereby offspring production) of *A. tenuiremis* was significantly lowered (75%) at 200 µg l^{-1} with only one of four mating pairs able to produce two clutches. Likewise, for *N. spinipes,* total offspring production was significantly lowered when considering that in 200 µg l^{-1} no mating pairs were able to produce offspring. Otherwise the time to hatch and number of offspring produced per hatch at lower concentrations were not significantly different from the control. In *T. battagliai* significant reductions in offspring production at nominal 18, 180 and 560 µg l^{-1} test concentrations were observed compared with the control. However, no effects were observed at 56 µg l^{-1}. Irrespective of the deviation from dose response at 56 µg l^{-1}, *T. battagliai* was the least sensitive species with regard to fertility (i.e., highest ability to produce clutches under 3,5-DCP exposure), but the most sensitive species with regard to offspring production by those females able to reproduce.

1.2.4 Conclusions

The comparison of results between the different harpacticoid and calanoid species should be treated with caution as such comparisons are based on a very small data set which was analyzed using different statistical techniques that makes direct comparison difficult. The sensitivity to 3,5-DCP between the four species differed depending on specific biological (e.g. physiology, metabolism, morphology) and ecological (e.g. life strategies) features of the respective species, and the endpoints chosen for analysis. Still, the results were within the same order of magnitude and the same concentration range was chosen for the ring test, i.e. *6.6, 20, 60, 180 and 540 µ*g/L 3,5-DCP.

To be able to find laboratories that were interested in participating (i.e. to reduce the work load) it was after the pre-validation at the OECD Invertebrate expert meeting in Paris 2005 decided that we should only proceed with 3,5-DCP as a reference substance.

1.3 Scheduled work plan

In June 2003 a Nordic Expert group under NordUtte prepared a proposal for a new guideline "OECD draft Guideline for Testing of Chemicals – Copepod Development and Reproduction Test". This guideline describes testing of effects on development and reproduction of marine harpacticoid and calanoid copepods in full life-cycle tests. The method is intended also to cover effects of potential endocrine disrupters.

This draft guideline was discussed at the first OECD meeting of the *ad hoc* expert group on invertebrate testing in November 2003 in Paris (OECD 2003). At this meeting it was decided to separate the first draft into two – one with the benthic species (*Nitocra spinipes, Amphiascus tenuiremis* and *Tisbe battagliai*) and one with the pelagic species *Acartia tonsa.* This was a consensus due to differences in biology leading to differences in methods, especially with regard to handling of animals, feeding and media, which made it difficult to include both harpacticoid and calanoid copepods in the same guideline. It was decided that the drafts should be ready by the end of 2003 and be commented within the first two months of 2004. A pre-validation test should start in spring 2004. OECD planned to finalise the pre-validation work by June 2004 followed by an inter-laboratory study on the two draft guidelines. The suggested time schedule could however not be met owing to a number of unforeseeable reasons.

Revised drafts for the two test guidelines (with harpacticoid and calanoids) were prepared in December 2003 by Magnus Breitholtz, Stockholm University, and K. Ole Kusk and Leah Wollenberger, Technical University of Denmark, respectively, and sent out for commenting in January 2004 by OECD. The received comments on the method with harpacticoid were compiled and responded to in May 2004. Owing to a late response from one of the invertebrate experts, the draft was not fully responded to until December 2004. After the pre-validation small changes was made to the harpacticoid draft (assembled with some specific recommendations). The calanoid draft was also updated: some corrections were made in the validity criteria (e.g. mortality rate in controls was increased from 20% to 30%) and the egg production was observed by the end of the test (day 15-18 instead of day 7-12).

1.4 Ring test

OECD sent out an invitation letter (see Appendix C) to potential participants in the ring test. Labs from five countries (i.e. UK, USA, Denmark, Sweden and Germany; see below for details) have been involved in the ring test of the copepod Test Guideline.

The actual ring-test started in the spring of 2006 and followed the schedule outlined below (see Table 1).

Table 1

	2006									
	March	April	May	June	July	August	September	October	November	December
Sending out 3,5-DCP	x	x	x							
Nitocra										
Testing	x	x								
Sending raw data to ITM (deadline)			x (May 31)							
Handling of raw data			x	x						
Sending test medium to ITM			x	x						
Analyses of test medium ITM										
Amphiascus										
Testing		x	x							
Sending raw data to ITM (deadline)				x (June 10)						
Handling of raw data				x	x					
Sending test medium to ITM				x						
Analyses of test medium ITM				x		x				
Acartia										
Testing			x	x						
Sending raw data to ITM (deadline)				x (June 30)						
Handling of raw data						x	x			
Sending test medium to ITM						x				
Analyses of test medium ITM						x	x			
Writing report: Harpacticoid part						x	x			
Writing report: *Calanoid part*							x			
Writing report: *Comparison*							x	x		
Sending report to Nord-Utte/OECD								30-nov		
Potential revision									x	
Sending final report OECD										15-dec

1.4.1 Reference chemical

At the expert meeting in OECD November 2003 (OECD 2003) it was decided to use two chemicals in the pre-validation work and perform the full test with both of them; but, according to the above-mentioned discussion (Section 1.2), only one reference substance was used in this ring-test activity:

3,5-dichlorophenol (DCP)

3,5-dichlorophenol is often used as a reference compound in various standardised toxicity test methods. It is water soluble, not ready biodegradable, non-volatile, and has a relatively low K_{ow}. This means that the compound is easy to handle and is expected to stay in the water phase for the duration of the test period.

Physical-chemical properties, toxicity and fate data of 3,5-DCP are presented below.

Physical-chemical properties:
CAS Number: 591-35-5
Molecular weight: 163.0
Water solubility: 5380 mg/L (25 °C) (Huyskens, P et al.1975)
MP (deg C): 68 (Hansch, C et al. 1995)
BP (deg C): 233 (Hansch, C et al.. 1995)
Log Kow: 3.62 (Hansch, C et al. 1995)
Vapor Pressure: 0.00842 mm Hg (extrapol.) (25 °C) (Shiu, WY et al.1994)
pKa: 8.36 (SPARC on-line calculator)

Toxicity data: (US EPA ECOTOX database except otherwise indicated)
Acartia tonsa 48 h LC50: 0.51 mg/L
Acartia tonsa 48 h LC50: 0.95 mg/L (ISO 14669)
Acartia tonsa 48 h LC50: 1.00 mg/L (Bjørnestad et al. 1993)
Crangon septemspinosa 96 h LC50: 1.50 mg/l
Carassius auratus (Goldfish) 96 h LC50: 3-5 mg/L
Zebra fish 24 h LC50: 2.9 mg/L (Zarorc-Konĉan et al 2002)
Tisbe battagliai 24 h LC50: 10.74 µM (1.75 mg/L)
Daphnia magna 48 h EC50: 2.48 mg/L (Zarorc-Konĉan et al 2002)

Fate data:
Test for ready biodegradability: 5% degradation in 28 days. (Zarorc-Konĉan et al 2002)

1.4.2 Participants in the ring test

The following laboratories participated in the ring test of the harpacticoid and calanoid draft TG:

Harpacticoids

Nitocra spinipes

- Department of Applied Environmental Science (ITM), Stockholm University, SE-106 91 Stockholm, Sweden. Contact: magnus.breitholtz@itm.su.se.

- IVL Swedish Environmental Research Institute, Box 210 60, SE-100 31 Stockholm, Sweden. Contact: ann-sofie.allard@ivl.se.

- Toxicon AB, SE-261 92 Landskrona, Sweden. Contact: anders.sjolin@toxicon.com.

Amphiascus tenuiremis

- Department of Environmental Health Sciences, Arnold School of Public Health, University of South Carolina, SC-29208, USA. Contact: tchandler@sc.edu [The University of South Carolina participated as a lead laboratory for guidance because of their experience in developing the initial protocol for *A. tenuiremis* but did not perform an experiment in the Phase 1.]

- Block Environmental Inc., Pleasant Hill, California, USA. Contact: dblock@blockenviron.com

- The NOAA Center for Coastal Environmental Health and Biomolecular Research (CCEHBR), Charleston, South Carolina, USA. Contact: geoff.scott@noaa.gov

Calanoid

Acartia tonsa

- Institute of Environment & Resources, Technical University of Denmark (E&R), Building 113, 2800 Lyngby, Denmark. Contact: K. Ole Kusk; kok@er.dtu.dk

- DHI Water and Environment (DHI), Agern Allé 5, 2970 Hørsholm, Denmark. Contact Estelle Bjørnestad; esb@dhi.dk

- Opus Plus Ltd, (OPUS), Flotta, Stromness, Orkney, KW16 3NP, United Kingdom. Contact Brenda Hudson; brenda.hudson@opusplus.co.uk

- ECT Oekotoxikologie GmbH, Boettgerstr. 2-14 D-65439 Floersheim, Germany. Contact: Michael Meller; m-meller@ect.de

2. MATERIALS AND METHOD

2.1 Full life cycle tests

For the methods used we refer to the two draft harpacticoid test guidelines (OECD 2005a, b; See also the invitation letter in Appendix C).

Before the ring test started the participating laboratories also received detailed ring test plans (see Appendix D and E).

2.2 Chemical analysis

Determination of chlorophenolics in water samples

All test medium samples that were sent to ITM were shipped frozen. The samples were stored in freezer until analyses.

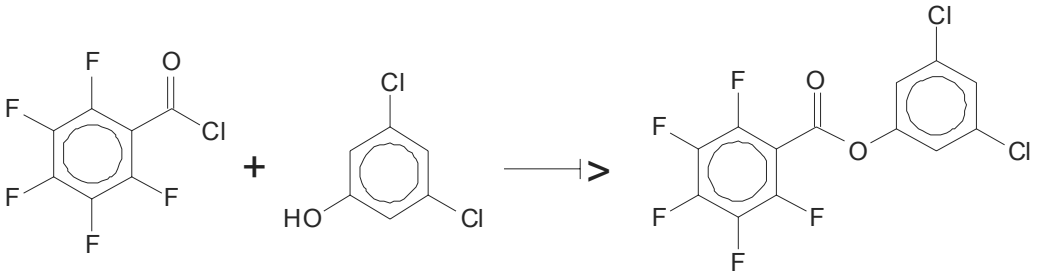
For improved detection by GC-ECD the extracts were derivatised by extractive acylation with pentafluorobenzoyl chloride (PFBCl). The derivatisation method used is based on the methods reported by Renberg 1981; Kuch and K. Ballschmiter 2001.

A surrogate standard (2,4-dichlorophenol) was added to the water sample (2mL). Then 50 µl of 2M KOH and 10 µl of a 10 % (v/v) solution of pentafluorobenzoyl chloride (PFBCl) in toluene were added. The derivatives were generated and extracted by shaking for 2 min with 2 ml of n-hexane containing a volumetric standard (CB 189). The hexane phase was analysed by GC/ECD.

Standard samples were prepared in the same way and standard solutions in five different concentrations (ranging from 5 to 40 pg/µl) were analysed in the beginning and at the end of each batch of samples. Concentrations were calculated using the surrogate standard. Calibration curves were linear over the concentration range studied with coefficients of correlation greater than 0.980. The regression plots appeared linear and the residual plots did not show any significant pattern. The residual sum squares were not greater than 0.178.

The limit of quantification (LOQ), calculated as signal to noise ratio (S/N) of 10, was 3 pg/μL and the limit of detection (LOD), calculated as an S/N ratio of 3, was 1 pg/μL.

The repeatability (inter-day) was calculated from a standard of 20 pg/μL 3,5-dichlorophenol that also was analysed as a part of the calibration curves analysed with every batch. The mean was 20 pg/μL, the standard deviation 3.2 and the relative standard deviation (RSD) 16%.



pentafluorobenzoyl chloride | 3,5-dicholorophenol | pentafluorobenzoyl ester

2.3 Statistical analysis

All data was delivered in Excel format from ITM and then transferred to, and analysed with the statistical software Stata version 8 (http://www.stata.com). Most of the statistical analyses chosen for each group of species and endpoint are discussed in the results and Discussion section. Still, specific explanations to some of the chosen statistical analyses are presented below.

For the harpacticoid copepods, the development rates had to be estimated for the different development stages (nauplii to copepodite and copepodite to adult) separately. The developmental data should also be treated as survival data where the outcome variable is copepodite/adult (yes or no), and where censored observations are present since the time to copepodite/adult is not observed for every individual. A first approach to the analysis was to fit a parametric survival function to data (lognormal or gamma), but it did not fall out well. The reason for that could be that these particular data are not pure survival data; there may perhaps be individuals that never become copepods or adults even if we follow them to infinity. In order to take this into account, a more sophisticated model would have been needed. As a simpler alternative the analysis was based on the nonparametric Kaplan Meier survival estimate.

In the analysis of development rate from nauplius to copepodite, an individual who turned out to be dead or is missing was handled as *censored* and a nauplius that did not develop to copepodite during the follow up time was also *censored*. The analysis of development rate from copepodite to adult was performed on individuals that had developed to copepodites during the follow up time. If an individual was dead, missing or did not become an adult within the follow up time, it was handled as censored at that specific time point.

There are different measures of development rate which can be used, e.g. mean time, median time, proportion developed at a particular time point, and it is not straightforward how to choose the most suitable measure for this data. The first candidate is the median value, which is an ordinary measure for survival data. For these particular data, the median will be a quite rough measure. The reason for that is the low precision of the time measurements compared to the quite small exposure effect. The time measurements are rounded up to one whole day so consequently the median has a value in whole day increments. The second measure candidate is the mean value. The mean value is calculated as the area under the Kaplan Meier curve, and that measure should be less rough and more appropriate where we have pure survival data and the follow up time is longer. For data in this study, the mean value may be sensitive to the choice of endpoint time scales. For that reason the median (50^{th} percentile) complemented with the 25^{th} percentile was chosen. In addition to that, a Log rank test was performed to test equality of the survival curves between an exposure and the control.

3. RESULTS AND DISCUSSION

Throughout the results and discussion section all the harpacticoid (i.e. both *N. spinipes* and *A. tenuiremis*) data have been treated in the same way as well as presented and discussed together, as if there were not two separate species that have been tested. The only calanoid species used in the ring test, *A. tonsa* was handled separately, both for statistics and presentation.

For further information about *within treatment variability* raw data spreadsheets are also presented in Annex F for most of the endpoints.

The following abbreviations were used to denote each laboratory in the figures and tables presented in the results and discussion section:

Harpacticoids:

Nitocra spinipes

- **ITM** (Department of Applied Environmental Science, Stockholm University, Sweden).
- **IVL** (IVL Swedish Environmental Research Institute, Sweden).
- **TOXICON** (Toxicon AB, Sweden).

Amphiascus tenuiremis

- **BEI** (Block Environmental Inc., USA).
- **NOAA** (The NOAA Center for Coastal Environ. Health and Biomolecular Research USA).

Calanoid:

Acartia tonsa

- **DTU** (Inst. of Environment & Resources, Technical University of Denmark, Denmark).
- **DHI** (DHI Water and Environment, Denmark).
- **OPUS** (Opus Plus Ltd, United Kingdom).

3.1 Harpacticoids

All five harpacticoid laboratories delivered developmental data, which have been analyzed and presented below. Adequate reproduction data were only available however for three of the laboratories, ITM, NOAA and Toxicon.

3.1.1 Development

In the developmental phase of the harpacticoid full life cycle test, the following endpoints were analysed: *mortality*, *development rate* and *sex ratio*.

Mortality

For a test to be valid, the mortality in the control must not exceed 20% at the end of the exposure period (OECD 2005b). Whether these requirements have been fulfilled was examined by the use of the non-parametric Kaplan Meier estimate of the survival function (1). In this analysis, an individual who turned out to be adult or missing was handled as censored at that specific time point. Also individuals still alive at the end of experiment were treated as censored.

The length of the experiment period in the development phase differed between laboratories (NOAA 29 days, BEI 22 days, ITM 26 days, IVL 16 or 19 days[1], Toxicon 35 days) and the mortality was examined both at the last day (i.e., overall mortality) and at day 16.

Table 2 shows the estimated survival proportions, over all, with 95% confidence intervals. Table 3 shows the estimated survival proportions, at day 16, with 95% confidence intervals. The Kaplan Meier survival curves are shown in Figure 1 in appendix A.

Table 2. Over all survival

Lab	Survival	Lower_95	Upper_95	
NOAA	0.861	0.757	0.923	*A. tenuiremis*
BEI	1.000	.	.	*A. tenuiremis*
ITM	0.724	0.154	0.946	*N. spinipes*
IVL	0.922	0.805	0.970	*N. spinipes*
Toxicon	0.965	0.865	0.991	*N. spinipes*

[1] 19 days for concentration 180 and 540

Table 3. Survival until day 16

Lab	Survival	Lower_95	Upper_95	
NOAA	0.861	0.757	0.923	*A. tenuiremis*
BEI	1.000	.	.	*A. tenuiremis*
ITM	0.965	0.867	0.991	*N. spinipes*
IVL	0.922	0.805	0.970	*N. spinipes*
Toxicon	0.965	0.865	0.991	*N. spinipes*

For ITM the overall mortality exceeds 20 % (27.6%) which exceeds the set validity criterion, but the mortality until day 16 is well within tolerance limits (i.e. 3.5 %). The other laboratories show a tolerable mortality for this particular experiment. An examination of the Kaplan Meier survival curve (see Figure 1 Appendix A) shows a massive mortality at day one (8 individuals out of 72) for the NOAA laboratory. It is difficult to ascertain whether this is an actual effect or if it is an artefact. Also, for NOAA we cannot be statistically certain that the "population mortality" is lower than 20 % (since the lower 95% confidence limit is below 80%). In this context it is important to notice that the protocols that were prepared for each laboratory to send in their detailed descriptions about the fate of each individual copepod were used differently between the laboratories. ITM apparently made much more detailed descriptions. For example, an animal that actually disappeared was classified as disappeared ("missing") and not as dead. If a laboratory experienced problems finding an animal in a single well, it is not unlikely that this animal was denoted as "missing" instead of "dead". For some laboratories this seems to have happened, which would of course have negatively skewed the mortality rates.

The *mortality among exposed* was investigated in the same manner as the control (estimation of the Kaplan Meier survival function). Figure 2 -4 in Appendix A shows the survival curves for three of the concentrations; 6.6, 20 and 540 μg/L. For some of the laboratories there is a significantly higher mortality for exposed copepods compared to the controls. When observing Figure 2-4 it becomes relatively clear that some of the non dose-related trends depend on the problematic decline of 3,5-DCP concentrations. For some laboratories (e.g. ITM and Toxicon) the lowest concentration (6.6 μg/L) has higher mortality rates than for example the highest concentration (i.e. 540 μg/L). This could either depend on random factors or on the bioavailability of the test substance, which, according to the test medium analyses (see section 3.4 below), for some reason was very low in the high concentration treatment(s).

Mortality was also investigated separately *for the different developmental stages* (nauplius to copepodite and copepodite to adult). One peculiar example is shown in Figure 5; the survival in the copepodite to adult stage, for NOAA. Almost all deaths occur one day after reaching copepodite stage, which was surprising and difficult to explain. There are probably some explanations other than randomness for that, since the same trend was also seen in the control. The last nauplius stage to first copepodite stage is a major metamorphosis requiring more extensive genetic to physiological change. It is not uncommon to see this transition as the most sensitive developmental window for *A. tenuiremis* (Chandler published and unpublished observations).

Development

The development rates have to be estimated separately for the different developmental stages (nauplius to copepodite and copepodite to adult). The results for development are presented in Table 4 and Table 5, and the corresponding Kaplan Meier failure curves are shown in Figure 6 to 9 in Appendix A (only for the development from nauplii to copepodite). (For notification: *Individuals that died on day 1 from start were excluded from the study*.)

The lowest concentration giving significant differences from a control (LOEC) in mean *time elapsed from the first nauplius stage to development of the first copepodite stage* was 6.6 μg/L observed by the two *Amphiascus* laboratories NOAA and BEI. However, for NOAA there was a lack of increasing dose-response as only 6.6 μg/L and 540 μg/L resulted in significant effects. Also for BEI there was a lack of clear dose-response effects in the sense that in 20 μg/L no significant difference from the control was observed (but significant effects were seen in the others). For ITM and IVL there was a clear dose-response but the LOECs were somewhat higher than for the two American *Amphiascus* laboratories; 60 and 20 μg/L, respectively. For Toxicon, 20 and 60 μg/L gave significant effects from the control. Overall, relatively similar results were seen, but it is likely that the declining 3,5-DCP concentrations have introduced artefacts in the anticipated dose-dependent trends.

The lowest concentration giving significant differences from a control (LOEC) in mean *time elapsed from the first copepodite stage to development of the adult stage* was 6.6 μg/L observed by ITM. There was however a lack of clear dose-response effects in the sense that in 60 μg/L no significant difference from the control was observed. (It should be noted, however, that at this concentration there was a significant effect for the earlier developmental stage (i.e. from nauplius to copepodite), which indicates that development overall was significantly affected for this laboratory). For NOAA and Toxicon, the LOEC was as high as 540 μg/L. It is noteworthy that for NOAA there was no significant difference from the control in any developmental phase in 20, 60 and 180 μg/L. For BEI there was a dose-related trend and the LOEC was 60 μg/L. For IVL there were significant differences from the control in 60 and 180 μg/L but not in 540 μg/L.

Table 4. Time in days to copepodite, 25th and 50th percentile

P-value of Log-rank test. Significant differences from the controls are marked by a light gray background.

Lab and conc.	p25	p50	p-value	
NOAA				*A. tenuiremis*
0	10	11		
6.6	8	9	.0425	
20	9	11	.6	
60	9	10	.155	
180	10	10	.258	
540	11	12	.000137	
BEI				*A. tenuiremis*
0	8	10		
6.6	9	10	.0265	
20	9	11	.103	
60	12	13	1.21e-08	
180	11	15	4.30e-10	
540			2.75e-08	
ITM				*N. spinipes*
0	9	9		
6.6	9	10	.432	
20	9	9	.967	
60	9	11	.00437	
180	9	10	.0146	
540	10	12	.000208	
IVL				*N. spinipes*
0	8	9		
6.6	8	9	.0834	
20	8	9	.0469	
60	8	9	.00503	
180	12	12	.0272	
540	12	14	2.08e-06	
Toxicon				*N. spinipes*
0	7	8		
6.6	8	8	.601	
20	6	7	.0000705	
60	6	7	.000456	
180	8	9	.539	
540	7	8	.431	

Table 5. Time in days between copepodite and adult, 25th and 50th percentile.

P-value of Log-rank test. Significant differences from the controls are marked by a light gray background.

Lab and conc.	p25	p50	p-value	
NOAA				*A. tenuiremis*
0	8	8		
6.6	6	7	.136	
20	6	7	.19	
60	7	8	.825	
180	7	8	.622	
540	9	10	.000234	
BEI				*A. tenuiremis*
0	8	8		
6.6	7	9	.155	
20	6	8	.597	
60	6	8	.0325	
180	9	9	.0273	
540	-	-	-	
ITM				*N. spinipes*
0	6	7		
6.6	7	8	.0121	
20	7	8	.00129	
60	6	8	.231	
180	7	8	.0114	
540	8	9	6.15e-06	
IVL				*N. spinipes*
0	4	5		
6.6	4	7	.125	
20	4	6	.145	
60	4	7	.0353	
180	4	7	.0265	
540	4	5	.174	
Toxicon				*N. spinipes*
0	10	13		
6.6	11	12	.58	
20	11	13	.533	
60	10	14	.272	
180	10	11	.241	
540	13	15	.0214	

Overall the LOECs are relatively similar among the five laboratories related to decrease in developmental rate. For NOAA, BEI and ITM the LOEC was μg/L and for IVL and Toxicon the LOEC was 20 μg/L. This indicates that *Amphiascus* may be somewhat more sensitive than *Nitocra* but too far-reaching conclusions should not be made due to the problematic decline in DCP-levels (see section 3.4 below), which was evident (but not necessarily identical) for all laboratories.

It should be noted that the declining 3,5-DCP concentrations are likely to have introduced artifacts in the concentration-response relationship. Consequently, it is difficult to provide a valid interpretation of these data. Hence the derivation of LOEC and NOEC values and statements made about species sensitivity must be treated with caution.

Sex ratio

For a test to be valid, the proportion of females (or males) in the control should be within 40% to 60% (OECD 2005b).

Table 6 shows the female proportion for each laboratory and concentration, and for the four laboratories that have made sex determinations, the female proportion in the control is within the guideline limits.

Table 6. Female proportion in adults

Laboratory					
Conc.	NOAA	BEI	ITM	IVL	Toxicon
0	0.53	0.56	0.57		0.43
6.6	0.49	0.50	0.43		0.49
20	0.52	0.53	0.60		0.45
60	0.58	0.58	0.55		0.46
180	0.44	0.78	0.53		0.47
540	0.29	-	0.53		0.31

The following observations were made regarding development time ranges from copepodite to adult for the control, among individuals that became adults (this is not a survival analysis estimate):

- For NOAA and BEI there were no significant differences between sexes.
- For ITM there was a significant difference (mean time: female 7.9 days, male 6.4 days). For Toxicon there was a large significant difference (mean time: female 10.4 days, male 14.3 days).
- For ITM this difference was due to difference from nauplii to copepodite but for BEI there was no difference from nauplii to copepodite.

Sex ratios are often difficult to analyse in the sense that in these tests there is seldom information available about the sex of dead immature animals. Conclusions of sex-specific effects should as a consequence always be carefully made. In this context, one could regard the male:female ratio of 40:60 or 60:40 as a prerequisite to be able to construct statistically reasonable n-size for male:female pairs for reproduction studies. This prerequisite has been met for all laboratories participating in the ring test.

3.1.2 Reproduction

Three laboratories performed the reproduction test (NOAA, ITM and Toxicon) and the outcomes analysed were fertilisation success, offspring production and time to first and second clutch. No survival analysis was made since information about the censored pairs was missing. Therefore we cannot exclude that an observed effect for an exposed group might be a consequence of differences in mortality for that group compared to the control.

Fertilisation success

Fertilization success was measured as the proportion of pairs able to produce two successive clutches within the follow up time, of all mating pairs at start[2]. The follow up time differs between laboratories (NOAA 10 days, ITM 19 days and Toxicon 20 days), which possibly make the comparisons between species and laboratories more difficult. The results are shown in Tables 7 to 9; number of mating pairs at start, the proportion of pairs with two clutches together with exact 95 % confidence limits and a p-value for testing the hypothesis of "similar proportions in the exposed and the control" (Fisher's exact test). The lowest LOEC was observed in 6.6 μg/L for ITM but there was a lack of dose-response in the sense that there were no significant differences from the control at both 20 and 60 μg/L. It is obvious that for ITM the 6.6 μg/L treatment has been negative for most endpoints, which makes it doubtful that this LOEC is realistic. This also becomes evident when looking at the other two laboratories, having LOECs for this endpoint as high as 180 μg/L (Toxicon) and 540 μg/L (NOAA). So, if we consider the very low LOEC for ITM to be in error (i.e. an artefact), we would end up with an overall LOEC of 180 μg/L for this endpoint.

Table 7. Proportion of pairs with two clutches, for NOAA (*A. tenuiremis*).

Significant differences from the controls are marked by a light gray background.

Conc	Freq	Prop	Lower_95	Upper_95	p-value
0	23	.74	.52	.9	.
6.6	23	.65	.43	.84	.75
20	20	.50	.27	.73	.13
60	18	.50	.26	.74	.19
180	21	.52	.3	.74	.21
540	7	.14	.0036	.58	.0086

[2] ITM observed that several of their pairs consisted of two females. These pairs were excluded.

Table 8. Proportion of pairs with two clutches, for ITM (*N. spinipes*).

Significant differences from the controls are marked by a light gray background.

Conc	Freq	Prop	Lower_95	Upper_95	p-value
0	18	.67	.41	.87	.
6.6	20	.25	.087	.49	.021
20	18	.56	.31	.78	.73
60	10	.50	.19	.81	.44
180	15	.13	.017	.4	.004
540	11	.091	.0023	.41	.0057

Table 9. Proportion of pairs with two clutches, for Toxicon (*N. spinipes*).

Significant differences from the controls are marked by a light gray background.

Conc	Freq	Prop	Lower_95	Upper_95	p-value
0	17	.82	.57	.96	.
6.6	21	.76	.53	.92	.71
20	22	.59	.36	.79	.17
60	22	.59	.36	.79	.17
180	18	.39	.17	.64	.015
540	16	.063	.0016	.3	.000013

The observed proportions of pairs able to produce two clutches are lower for the exposed compared to the control, for all concentrations in all three laboratories. If we use significance level[3] 5%, every one of the laboratories shows a significant difference from the control for 540 μg/L, and ITM along with Toxicon also show significant differences from the control for 180 μg/L. In addition, ITM shows a significant difference from the control for 6.6 μg/L, which makes the data not following a dose related trend. This concentration did not respond as expected (i.e. being the lowest concentration it was expected to give little or no effects) for this laboratory. Care should thus be taken when setting a LOEC for this laboratory and endpoint.

The probability that a statistical test of this kind will fall out as significant (the power of the test) depends on the size of the true effect and the sample size. Obviously, it was not possible to conclude, for a non-significant test, that there were no effects. The power might just have been too low to reveal it. For that reason, no attempt was made to estimate NOEC or LOEC with these data.

An alternative approach is to assume a model for the dose response relationship. For these data, there seems to be a monotonic dose response trend (with an exception at 6.6 μg/L for ITM) so a logistic

[3] not adjusted for multiple test

regression model was fitted for each laboratory separately. The models are illustrated in Figure 10 in Appendix A, and except for the observation from ITM at 6.6 μg/L, the fit is quite good.

In order to estimate an EC_{50} -value from these data, a non-linear model, like the logistic function, would have to be fitted. This is not a straightforward task (model choice, programming, model check etc.) and since there was limited time available for a major revision of the statistical analyses, no attempt was made to produce EC estimates for the ITM study.

Offspring production

For a test to be valid the average number of viable offspring per clutch in the controls should not be lower than 5 (OECD 2005b). This criterion was met for all three laboratories (average number was higher than 8 for all three, data not shown). For pairs that were able to produce two clutches, the total number of viable offspring was calculated. The results are displayed separately for each laboratory in Table 10 to 12. The tables show the frequency of mating pairs able to produce two clutches, and the mean value of total offspring production together with 95 % confidence limits (based on the t-distribution). A one-way ANOVA for each laboratory separately shows no significant differences (p-values[4]; 0.10, 0.60 and 0.74 for laboratory NOAA, ITM and Toxicon, respectively). The results are also illustrated with box plots in Figure 11 in Appendix A. It can be noted that in 540 μg/L (for all laboratories) almost no pairs were able to produce two clutches.

Table 10. Total viable offspring production, for NOAA (*A. tenuiremis*).

Conc	Freq	Mean	Lower_95	Upper_95
0	17	17.4	15.3	19.4
6.6	15	14.7	12.7	16.8
20	10	15.7	13.4	18
60	9	12.7	8.77	16.6
180	11	14.1	10.2	18
540	1	5	.	.

Table 11. Total viable offspring production, for ITM (*N. spinipes*).

Conc	Freq	Mean	Lower_95	Upper_95
0	12	22.3	16.3	28.2
6.6	5	18.6	11.5	25.7
20	10	23.7	18.2	29.2
60	5	20	9.32	30.7
180	2	28.5	9.44	47.6
540	1	29	.	.

[4] Concentration 540 with only one observation was excluded in this test.

Table 12. Total viable offspring production, for Toxicon (*N. spinipes*).

Conc	Freq	Mean	Lower_95	Upper_95
0	14	16.9	13.3	20.4
6.6	16	19	15.7	22.3
20	13	19.8	16.1	23.4
60	13	17.6	13.3	21.9
180	7	20.3	10.6	30
540	1	25	.	.

Time to first and second clutch

Tables 13 to 15 show the median time to the first clutch, estimated for the pairs that had produced a first clutch (not a survival analysis estimate). A Mann Whitney U-test was performed to test the hypothesis of equal time distributions for the exposed and the control test populations. P-values on these tests are shown in the table. Since we have ignored the censored observations, the interpretation of these results should be cautious. The results are also illustrated with box plots in Figure 12 in Appendix A. Observe that the lack of dose-response in Tables 13-15 (i.e., no significant difference from the control in the highest concentration) is likely due to the low number of observations at this concentration. The LOEC was 60 μg/L for all laboratories.

Table 13. Time to production of first clutch, for NOAA (*A. tenuiremis*).

Significant differences from the controls are marked by a light gray background.

Conc	Freq	Median	p_value
0	17	3	
6.6	17	3	.79
20	11	3	.48
60	11	4	.029
180	13	4	.00027
540	1	7	+

Table 14. Time to production of first clutch, for ITM (*N. spinipes*).

Significant differences from the controls are marked by a light gray background.

Conc	Freq	Median	p_value
0	15	7	
6.6	6	8.5	.5
20	13	10	.082
60	7	9	.016
180	4	10	.049
540	2	6.5	.55

Table 15. Time to production of first clutch, for Toxicon (*N. spinipes*).

Significant differences from the controls are marked by a light gray background.

Conc	Freq	Median	p_value
0	15	4	
6.6	18	6	.18
20	17	6	.21
60	18	7	.0041
180	12	9.5	.00056
540	4	5.5	.61

Tables 16 to 18 show the median time between the first and second clutches, estimated for the pairs that had produced a second clutch (not a survival analysis estimate). A Mann-Whitney U-test was performed to test the hypothesis of equal time distributions for the exposed and control populations. P-values on these tests are shown in the table. Since we have ignored the censored observations, interpretation of the results should be cautious.

Table 16. Time between first and second clutch, for NOAA (*A. tenuiremis*).

Conc	Freq	Median	p_value
0	17	1	
6.6	15	1	.65
20	10	1	.19
60	9	2	.59
180	11	2	.18
540	1	3	

Table 17. Time between first and second clutch, for ITM (*N. spinipes*).

Conc	Freq	Median	p_value
0	12	3	
6.6	5	5	.67
20	10	4	.17
60	5	3	.91
180	2	3	.84
540	1	3	

Table 18. Time between first and second clutch, for Toxicon (*N. spinipes*).

Conc	Freq	Median	p_value
0	14	3.5	
6.6	16	3	.35
20	13	3	.27
60	13	3	.42
180	7	4	.26
540	1	4	

Overall, the reproductive data are relatively consistent between the three harpacticoid laboratories, with lowest LOECs ranging from 20 to 60 μg/L. Also, the three endpoints related to reproduction follow the same trends for all three laboratories (with some small exceptions that may rather be credited to artefacts), with time to production of the first clutch being the most sensitive endpoint, and offspring production the least sensitive endpoint. In a population model, all of these three endpoints would lower e.g. population growth, which illustrates the importance to study other reproductive endpoints than solely production. Since the 3,5-DCP decline was probably most severely manifested in the end of the test and since the three laboratories' test times were very different, the results still indicate a similar cross-species sensitivity to the reference substance. The total offspring production data were calculated for pairs that were able to produce (a minimum of) two clutches. This approach ignores potential effects on reproduction whereby some pairs may only produce a single brood or egg sacs that fail to develop (female may be fertilized but eggs do not develop!). Consequently, it may be argued that the data analysis is 'biased' towards the most fertile females in the 'population' of animals within each treatment.

3.2 Calanoid

3.2.1 Development

Mortality of F_0

For a test to be valid the average mortality[5] of animals in the controls shall not exceed 30% of the hatched animals (OECD 2005a). Table 18 presents the number of replicates used, and the mortality proportion, for each concentration and laboratory. The average mortality in the controls (Conc.=0) is lower than 30% for every one of the laboratories.

Table 18. Number of replicates and mortality proportion.

	Conc.						
Lab	0	6.6	20	60	180	540	1620
DHI	9			4	4	4	4
	.21			.18	.17	.20	.40
DTU	12	6	5	6	6	6	
	.15	.21	.11	.15	.12	.17	
OPUS	12	5	5	5	5	5	
	.071	.12	.17	.091	.11	.53	

LDR of F_0 and F_1

LDR was measured as the proportionate copepodite stage of the total viable animals (number of copepodites/(number of copepodites + number of nauplii)) at the end of exposure. Because of over-dispersion in the data (i.e., more variation between replicates within treatments than expected under the assumption of equal probability) the analysis of data was performed with a random effect logit model. It should be noted that the validity criterion (i.e. control LDR should be within 50 ± 20 %) for the LDR endpoint was not met for two of the laboratories (i.e. DHI and OPUS, in both generations), which weakens the lab-to-lab comparability of these results (Table 19).

Tables 19 and 20 show the estimated LDR value from the logit model, together with a p-value for testing the hypothesis of equal LDR in the exposed and the control. No clear conclusion about the relationship between LDR and exposure can be made. For DHI an effect appears only for the highest concentration. For DHI an effect appears only for the highest concentration (the "extra" concentration 1620 μg/L not included). For DTU the plot looks strange; LDR in the controls are surprisingly low, and the

[5] Eggs at the end of exposure were excluded in the mortality calculation.

observations for concentration 20 are clustered into two distinct groups. For OPUS there is a monotonically decreasing relationship with concentration (with some extreme outliers).

Table 19. Estimated LDR of F_0, p-value.

Significant differences from the controls are marked by a light gray background.

	conc						
lab	0	6.6	20	60	180	540	1620
DHI	.71			.61	.65	.33	.00
	-			>.05	>.05	<.001	---
DTU	.47	.58	.55	.45	.41	.04	
	-	.0186	.0674	.613	.15	<.0001	
OPUS	.89	.65	.5	.22	.22	.03	
	-	.0001	<.0001	<.0001	<.0001	<.0001	

The lowest LOEC for the F_0-generation was observed at 6.6 μg/L for DTU and OPUS, but for DTU there was a lack of concentration-response in the sense that there were no significant differences from the control at 20, 60 and 180 μg/L. The LOEC for DHI was 540 μg/L for this endpoint.

Table 20. Estimated LDR of F_1, p-value.

Significant differences from the controls are marked by a light gray background.

	conc					
lab	0	6.6	20	60	180	540
DHI	.79	.75	.75	.74	.72	.31
	-	.239	.101	.120	.0268	<.0001
DTU	.46	.51	.46	.5	.35	.21
	-	.502	.99	.6	.135	.0003
OPUS	.89	.9	.76	.63	0	0
	-	.921	.00208	<.0001	-	-

The data show that for DHI, the LOEC for development in the F_1-generation was 180 μg/L (Table 20). For OPUS, the F_1-generation was less sensitive with a LOEC of 20 μg/L, as compared to 6.6 μg/L for the F_0-generation. However, the LDR validity criterion for OPUS was not met for either of the generations, so care should be taken when analysing these results.

Overall, the lowest "concentration-related" LOEC for development was 6.6 μg/L observed for OPUS in the F_0-generation. The validity criterion that LDR should be within 50 ± 20 % was not met for two of the three laboratories (DHI & OPUS) for both generations (71 & 89% and 79 & 89% as shown in Table 19 and 20, respectively). Considering the limited number of data sets that met the validity criterion (i.e. non valid tests), it is difficult to provide a robust interpretation of the data.

Length of male and females in F_0

The mean length of males and females at the termination of the development of the F_0-generation in F_0 is presented in Table 21. There were no obvious dose response trends in these data. To test for differences in mean length for each of the exposures compared to the control, t-tests with pooled variance estimates were performed. To control for compounded alpha error from multiple comparisons, Bonferroni corrections were made within each laboratory and sex dataset. With that correction a difference was deemed significant at the 0.05 level if the p-value was lower than 0.01. The p-values are shown in Table 21. For DHI, the mean length at concentration 180 μg/L were significantly *higher* than the control, for both sexes. For DTU there were no significant differences in mean length between the exposed and the control.

Overall, these data indicate that there are no obvious dose-related effects on length in *Acartia tonsa* exposed to 3,5-DCP up to 540 μg/L. Considering the time and effort required to produce the length data, and the accuracy of length measurements, it may be worth reviewing whether this endpoint should be included in the test method.

Table 21. Number of observations, mean length (mm) of male and females in the development of the F_0 generation and p-value.

	lab and sex			
	----- DHI ----		----- DTU ----	
conc	Female	Male	Female	Male
0	50	50	27	24
	.92	.75	.9	.76
	-	-	-	-
6.6	25	25	38	25
	.91	.75	.88	.74
	.11	.13	.065	.11
20	25	25	25	37
	.92	.75	.88	.77
	.38	.27	.083	.15
60	25	25	25	40
	.93	.75	.88	.74
	.21	.27	.11	.06
180	25	25	25	51
	.95	.76	.9	.76
	.00071	.0065	.46	.46
540	25	25	31	25
	.93	.75	.89	.74
	.18	.33	.18	.045

Hatching success of F_0 and F_1

For a test to be valid, the hatching success in the control must exceed 80% (OECD 2005a). The hatching success was measured as the proportion of hatched eggs at the end of each exposure and calculated as (1– number of unhatched eggs at end /number of eggs at start). The hatching success for DTU in the control was lower than 80% in 2 replicates out of 12 (the average was 86%). For DHI and OPUS the hatching success exceeded 80% in every one of the replicates. The hatching success in the F1-generation in the controls was within the limit for every one of the replicates.

3.2.2 Reproduction

Egg production

For a test to be valid, the average control egg production should be higher than 30 eggs/female/day at each of the three observation days from day 15 to day 18 (OECD 2005a). Table 22 shows that this criterion was met for DHI and DTU, but not for OPUS (average 27, 28, and 29). There was no mortality in the controls (the mortality should be lower than 20%; OECD 2005a). In order to examine exposure effects on egg production the average egg production over the 3 days was calculated for each female. Figure 13 in Appendix A shows a scatter plot of average egg production against concentration. There are no obvious dose response trends in the data, except possibly for DTU. To test for differences in mean egg production for each of the exposures compared to the control, t-tests with pooled variance estimates were performed. To control for multiple tests, Bonferroni corrections were made within each laboratory. With that correction, a difference was significant at the 0.05 level if the p-value is lower than 0.01. For DTU there is a significantly lower mean egg production for all exposures compared to the control. DHI shows a significantly higher mean egg production at 20 μg/L, and OPUS shows a significantly lower mean egg production at 540 μg/L.

Table 22. Number of females, mean and standard deviation of egg production, in controls over 15, 16 or 17 days.

	day		
lab	15	16	17
DHI	10	10	10
	40	38	42
	16	16	19
DTU	12	12	12
	45	40	39
	7.6	8.6	14
OPUS	7	7	7
	27	28	29
	6.7	9.5	6.4

Table 23. Number of females, mean egg production and p-value. Significant differences from the controls are marked by a light gray background.

	conc					
lab	0	6.6	20	60	180	540
DHI	10	12	12	12	11	12
	39.6	50.8	53.5	41.2	49	40.9
	-	.019	.005	.381	.041	.405
DTU	12	12	12	12	12	11
	41.6	33.3	32.3	24.3	24	25.9
	-	.004	.002	<.001	<.001	<.001
OPUS	7	5	6	7	6	7
	28.2	32.5	29.1	30.8	32.7	12
	-	.241	.437	.321	.22	.003

3.3 Compilation of NOEC and LOEC for both groups of species

Table 24 shows a compilation of all LOECs from the various tests conducted in the present ring test activity. For the three harpacticoid laboratories that completed full life cycle tests the lowest LOEC was observed for development (6.6 μg/L) whereas for the calanoids the picture was somewhat divided; for DTU reproduction was the most sensitive endpoint (6.6 μg/L), for DTU development and reproduction was equally sensitive (6.6 μg/L), and for OPUS development was far more sensitive than reproduction, 6.6 μg/L compared to 540 μg/L.

Table 24. Overview of LOEC-values (in μg/L) for all participating harpacticoid and calanoid laboratories.

Laboratory	Development	Reproduction		
		Time to 1st clutch	Proportion	Production
NOAA	6.6	60	540	[a]
BEI	6.6	-	-	-
ITM	6.6	60	180[b]	[a]
IVL	20	-	-	-
Toxicon	20	60	180	[a]
USC[c]	18	[d]	200	[d]
		Body length		Production
DHI	180	180[e]		6.6[f]
DTU	6.6[g]	540		6.6
OPUS	6.6	-		540

a No LOEC achieved owing to too few pairs in highest treatment.
b Lowest observed LOEC was 6.6 μg/L but this is likely an artefact in the sense that all higher treatments resulted in no significant difference from the control.
c Data from Pre-validation using Amphiascus tenuiremis (see OECD 2005c)
d No significant difference from the control up to and including 400 μg/L 3,5-DCP.
e Lack of concentration response since at 540 μg/L there were no response.
f Lowest observed LOEC was 6.6 μg/L but there was a lack of concentration response. Also, this response was associated to an increased production of offspring.
g Lowest observed LOEC was 6.6 μg/L but there was a lack of concentration response.

3.4 Chemical analysis of 3,5-DCP – potential reasons for declining levels

Measured 3,5-DCP concentrations follow a similar declining trend for all participating laboratories, but the trend is less steep for the *Acartia* labs (possibly a result of the larger test beakers). ITM has therefore conducted some follow-up tests where we have studied a) how fast the 3,5-DCP decline occurs and b) the fate of the parent compound. In brackish water at a pH of about 8-8.5 (which is the pH range most laboratories have reported) the disappearance from the water phase is significant within a couple of days (data not presented here). A first suggestion was that this problem was due to hydrolysis. After several experts' comments on this, it seems likely that this is not the case. Since the *pKa*-value for 3,5-DCP is about 8.2, it is obvious that the parent compound likely "oscillated" between a "protonated" and an ionized form during the test. It is common that pH decreases (may also increase) in the test medium as animals grow and/or debris accumulates in test beakers. A plausible hypothesis for the DCP-decline could be that pH-levels around the pKa-value may lead to ionization/protonation phenomena, which may explain the declining 3,5-DCP levels. Whether the ionized or the protonated form possibly "adheres" to test apparatus walls and becomes less bioavailable is currently not known.

We have in two consequent experiments (with a relatively high pH = 8.5) shown that there is much more 3,5-DCP in the water than on the walls. In brief, the tests were made as follows: Brackish water was adjusted with 0.1 M NaOH to a pH of 8.5. A stock solution of 80 µg/L 3,5-DCP was prepared and added to 96-well glass coated microplates (300 µL total volume) in replicates of two for each sampling occasion. On days 3, 7, 10 and 14, water and acetone samples were collected (acetone was filled to the top of each well). The same method as described above was used for analyses. A calibration curve (5-40 μg/L) was used for all analyses. No animals and no algae were added in these experiments. The results from this analysis are presented in Table 25 below. Given some mild evaporation problems, these results give no indication that a 3,5-DCP decline due to chemical association with test vessel walls would have occurred and caused the "poor dose-response" problems observed in the ring tests. On the other hand, there were neither animals nor algae added in this analysis, and these "phases" may have affected 3,5 DCP fate and bioavailability. Further studies are therefore needed to investigate the cause(s) of declining DCP concentrations with time.

Table 25. Two preliminary experiments showing the fate of 3,5-DCP in the test system used for the harpacticoid copepods.

Food and algae were not added. "Acetone" denotes the fraction of the substances sorbed to the wall and "Water" the fraction that was freely dissolved in the water phase.

	Test 1[2]			Test 2[3]		
Day	Water	Acetone	Total[1]	Water	Acetone	Total[1]
3	89	11	99	88	12	100.3
7	79	1.3	80	74	1.4	75.5
10	79	1.5	80	53	2.2	55.6
14	72	2.4	74	87	4.7	91.6

[1] Should add up to 80 µg/L. Values higher than that likely owes to up-concentration due to evaporation

[2] A stock solution was made in brackish water and was analysed day 3. Measured concentration was then 72 µg/L.

[3] A stock solution was made in brackish and was analysed day 3. Measured concentration was then 66 µg/L.

Figure 1 to 3 show that there was a steady decline of 3,5-DCP in the test medium in the harpacticoid tests. Interestingly, NOAA reported that they made new stock solutions from the crystalline powder *each* time the medium was renewed. As Figure 1a shows they still had declining levels, which may be explained by the fact that at most occasions the recovery of the test substance in the newly prepared medium was relatively low (on average 50%) (Figure 1b). This finding suggests that 3,5DCP is not stable in high ionic strength, basic pH seawater, and it should not be used as a reference compound in the future.

For ITM, which had some difficulties dissolving the substance in distilled water, a distilled water stock solution was prepared in the beginning of the test and used to prepare new brackish water solutions at each test medium renewal until day 10, when a new batch was made. This 3,5-DCP volumetric cylinder was analysed (after it had been held cold, sealed and in darkness) after 2½ months. The measured concentration was still approximately 50% of the nominal concentration, indicating that basic pH high ionic strength seawater drives 3,5DCP from its parent structure.

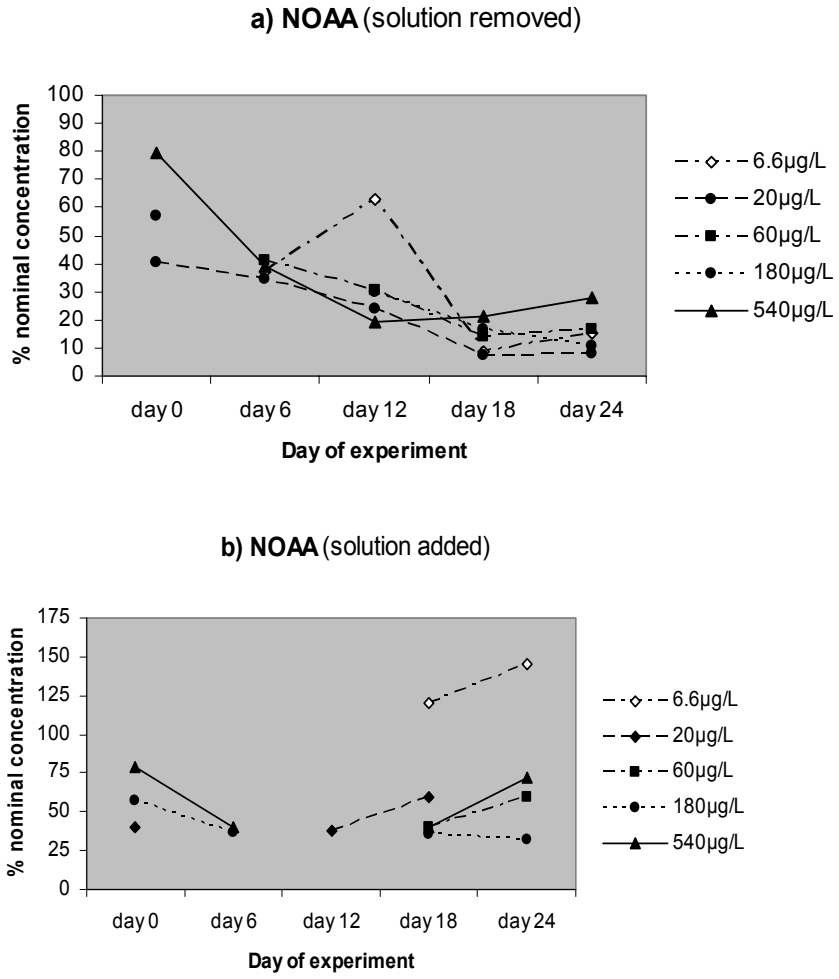


Figure 1 a & b. Chemical analysis of NOAA's test medium samples.

Massive decline between day 0 and day 24. Observe that also the freshly made test medium have lowered 3,5-DCP levels, on average around 50%.

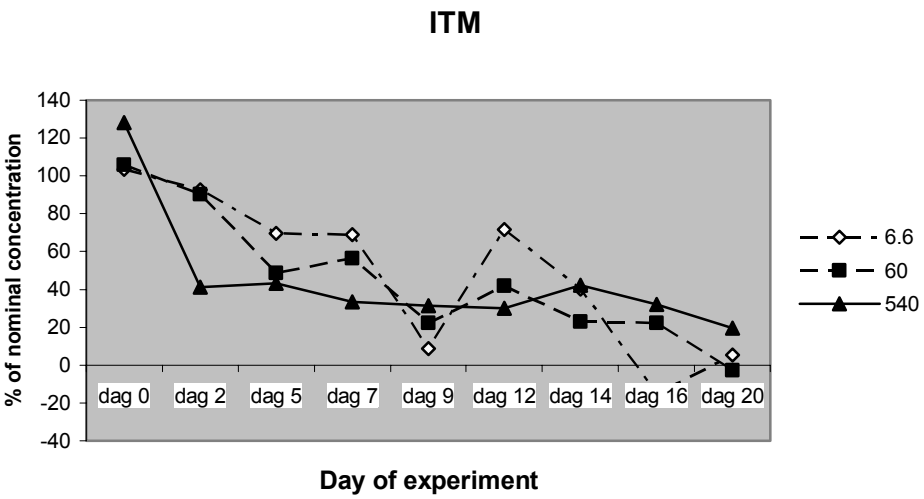


Figure 2. Chemical analysis of ITM's test medium samples.

Massive decline between day 0 and day 19.

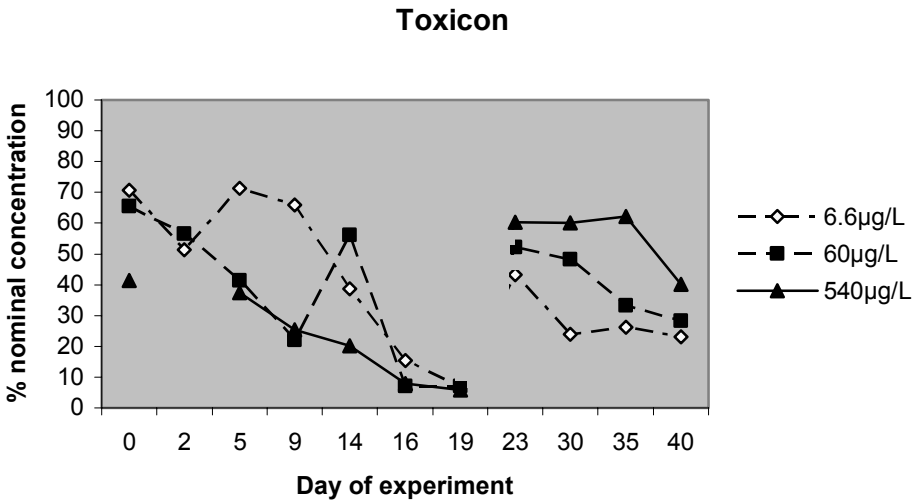


Figure 3. Chemical analysis of Toxicon's test medium samples.

Massive decline between day 0 and day 19. After that new and larger micro well plates were used. It should be noted that the same stock solution as was used during the first 19 days was used from day 10 and onwards.

Figures 4 to 6 show that there was relatively consistent variability in 3,5-DCP concentrations in the test medium in the calanoid tests, but no clear trend in decline with time. In Figure 5 it is clear that the "disappearance" of 3,5-DCP is much more dramatic for the lower than for the higher concentrations.

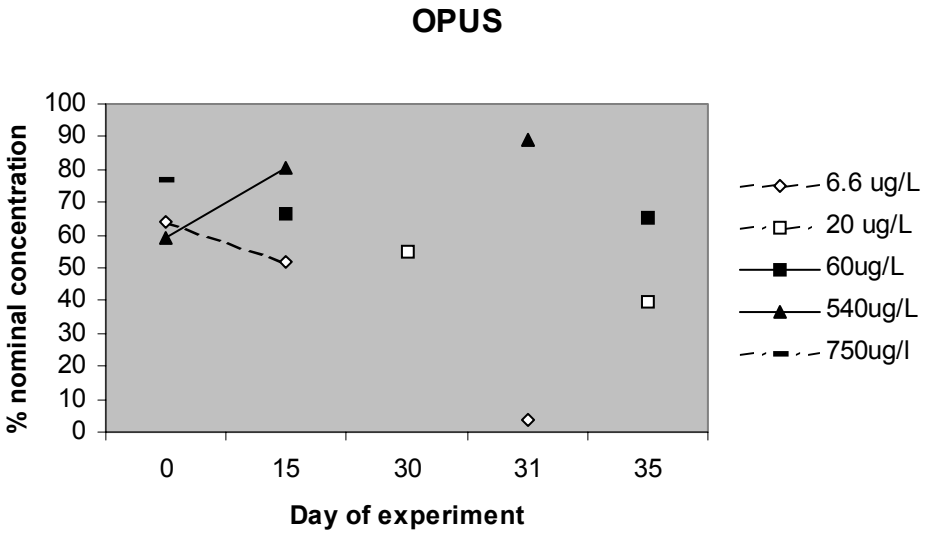


Figure 4. Chemical analysis of OPUS test medium samples.

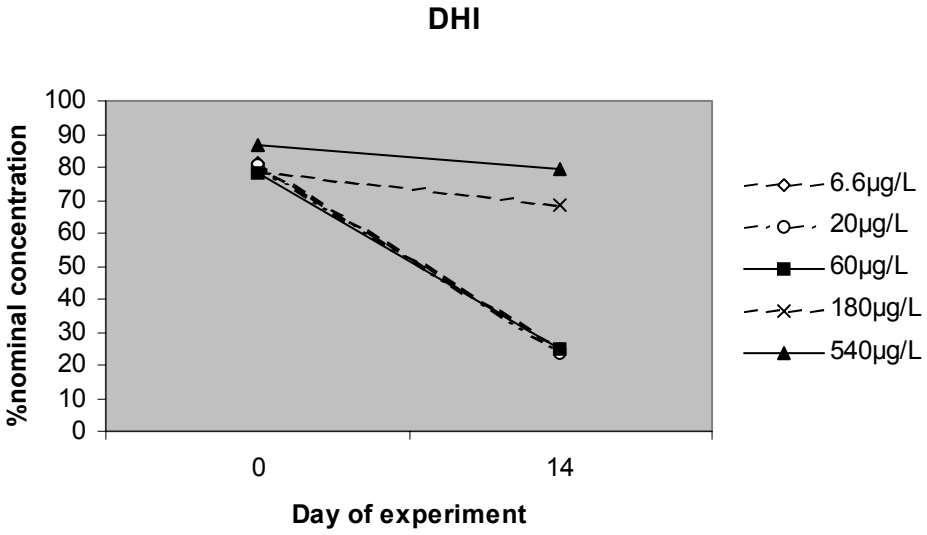


Figure 5. Chemical analysis of DHI's test medium samples at start and at day 14.

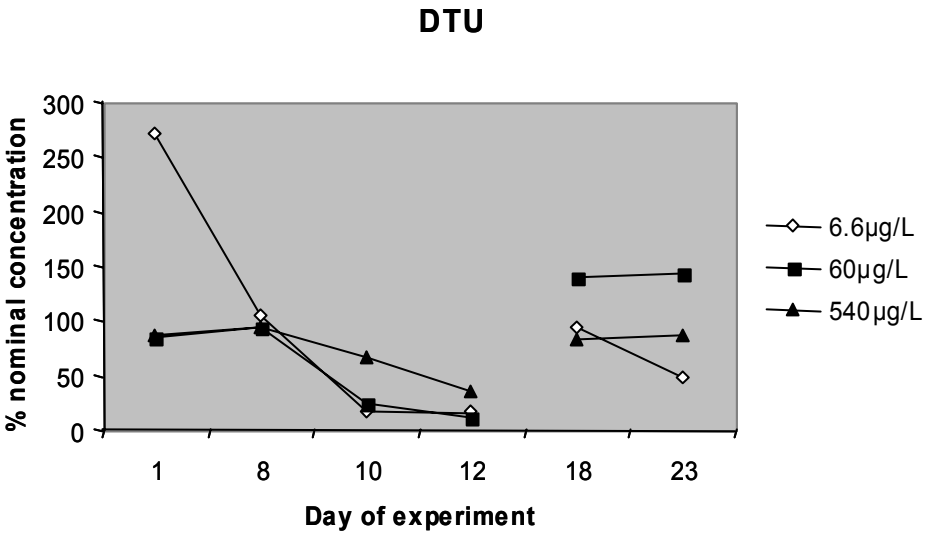


Figure 6. Chemical analysis of DTU's test medium samples.

For the lowest concentration there is likely an artefact day 0 since the measured levels were almost 3 times higher than the nominal. A plausible explanation could be that it is actually a 20 µg/L sample that has been denoted as a 6.6 µg/L sample.

Overall, the chemical analyses show that 3,5-DCP may not be the most suitable reference substance when conducting long-term experiments with marine species. Even if we today do not really know what caused the decline, it is likely that the problem is complex. Something happens both in the stock solution and in the test medium over time. Further, when animals and algae are not added, the problem is (as shown in Table 25 above) almost insignificant over a 14-day period. Whether this is an indication of metabolism in the animals, pH-changes and subsequent breakdown and/or accumulation in copepods and/or algae is not known. The importance to solve the 3,5-DCP problems as indicated in the present ring test should be of interest to many laboratories and organisations that use this substance as a reference molecule in standard ring tests.

3.5 Water quality within test vessels

Throughout the experiments the following water quality measurements were made in the control vessels:

ITM

- pH: 7.8 – 8.5
- Salinity: 6.2 – 6.9
- Dissolved O_2: 91 – 107 %
- Temperature: 21.1 – 23.1°C

NOAA
- pH: 8.0 – 8.3
- Salinity: 30 – 31‰
- Dissolved O_2: 88 -- 90 %
- Temperature: 24.8 -- 25.3°C

Toxicon
- pH: 7.73 – 8.37
- Salinity: 5.7 – 6.7 ‰
- Dissolved O_2: 97 - 105%
- Temperature: 19.4 – 24.3°C (extreme values, often around 21-22°C)

OPUS
- pH: 7.5 – 8.1
- Salinity: 34 - 36 ‰
- Dissolved O_2: 92 - 100%
- Temperature: 19.1 – 21.9°C

DHI
- pH: 7.8 - 8.4
- Salinity: 20 ‰
- Dissolved O_2: 97 - 99%
- Temperature: 20°C

DTU
- pH: 8.0 - 8.3
- Salinity: 20.1 – 20.4‰
- Dissolved O_2: 8.4 – 8.7 mg/L
- Temperature: 20 - 21°C

4. SUMMARY AND CONCLUSIONS

The present ring test activity shows that consultancy laboratories are able to run both the harpacticoid and calanoid full life cycle test, which both have relatively difficult and demanding experimental designs. However, the significant 3,5-DCP declines, as observed for all laboratories for which test medium was analyzed, obviously makes it difficult to draw thorough conclusions about exact effect levels. Nevertheless, similar declines were evident for all laboratories, which makes a comparison of LOECs possible on a nominal concentration basis. The developmental data were consistent among all the species used; normally the data also showed low variability. For the harpacticoids the LOEC range for "development" was between 6.6 and 20 μg/L whereas for the calanoids the range was somewhat larger and variable among the three participating laboratories (in the F_0-generation); i.e. 6.6 to 180 (or 540 if only considering dose-related effects) μg/L. It should also be noted that for the calanoid development data (i.e. LDR), the validity criteria was not fulfilled for one or more laboratories (in both generations), which of course limits the comparability of these data.

For the harpacticoids, the reproduction data was also consistent, at least for two out of three "reproductive" endpoints (i.e. "time to first clutch" and "offspring production"). The reproductive data for the calanoids varies more, with LOECs ranging from 6.6 to 540 μg/L. This might be due to a *within test variability* of biological tests which is well-known and normally corrected for via increased numbers of replicates. If we have a *within test variability* we consequently also have a *between test variability*. As complexity (and relevance) of the model increases, the reliability usually decreases since the large number of parameters interacting in the model will increase the rate of random errors and makes these tests less reproducible (Breitholtz et al. 2006). The judgement of the data from this ring test exercise should be made in light of the fact that there was no steady concentration level of the reference chemical for any treatment in any laboratory (which could be argued makes all tests invalid), and that there were perhaps too few participants per species to avoid possible negative impacts of random error and natural variation (on average three to four laboratories per species).

Another finding from this ring test activity is that the LOECs for the developmental data are most often lower than the LOECs for reproduction. This is a finding that has been observed in several studies using e.g. harpacticoid copepods (e.g. Breitholtz and Wollenberger, 2003; Breitholtz et al. 2003; Bejarano et al. 2006). Whether this is actually reflecting the sensitivity of the species exposed to 3,5-DCP in the present ring test is difficult to say since the levels were not stable over time. On the other hand, the chemical analyses of the test medium samples for e.g. Toxicon shows that the measured concentrations of 3,5-DCP seen on days 16-19 increased drastically as the reproductive phase was entered, but the LOEC for development was still lower for the developmental phase. Since especially for *N. spinipes* the total test time was over 40 days, these findings may call for application of partial instead of full life cycle tests. The only realistic full lifecycle alternative would otherwise be to continue with *A. tenuiremis,* for which the test time was more than 15 days shorter*,* as the only harpacticoid test species. Also, for this species there is already a standard test using the same experimental design available within ASTM (ASTM 2004) and numerous scientific articles has been published based on the current OECD draft TG experimental design (e.g. Chandler et al. 2004a, b).

For future validation work of the two draft test guidelines, it is recommend that:

- If all of the species used in the current ring test are to be included as a multi-species testing system in the future, then a partial life cycle test that focuses mainly on copepod development may be a better alternative if data comparability across all species is critical.

- If the regulatory need for a full life cycle test is greater, then *A. tenuiremis* should be used as a single harpacticoid species, owing to its much shorter test time and better reproductive output and success, which drastically increases the statistical power of the test.

- Other reference chemicals should be used until the reason(s) for the 3,5-DCP declines using the current brackish and marine species has been fully examined.

- More laboratories should participate in the ring test activities to avoid random error and natural variability that can drastically influence the overall judgement on validity of ring test results and future recommendations.

- In order to estimate EC_{50} values rather than traditional LOEC/NOEC's (which are more strongly-dependent on sample size and statistical power), non-linear regression-based models (e.g., Liber et al. 1992) should be fitted to the harpacticoid and calanoid development data because these models are more strongly dependent on numbers of test concentrations and organismal dose response along those test concentrations.

- The Phase 2 validation work should consider the difficulties encountered in the Phase 1 studies in the planning of the test protocol and overall testing program.

5. REFERENCES

American Society for Testing and Materials (ASTM). 1988. Standard guide for conducting acute toxicity tests with fishes, microinvertebrates and amphibians. ASTM Standard No. E 1192-88. ASTM, Philadelphia, PA, pp. 102-121.

American Society for Testing and Materials (ASTM). 2004. Standard Guide for Conducting Renewal Microplate-Based Life-Cycle Toxicity Tests with a Marine Meiobenthic Copepod. ASTM Standard No. E2317-04. ASTM, Philadelphia, pp. 1-16.

Bejarano, A.C., G.T. Chandler, L.He, T.L. Cary and J.L. Ferry. 2006. Risk assessment of the NIST petroleum crude oil standard water accommodated fraction (WAF): Further application of a copepod-based full life-cycle bioassay. Environ. Tox. Chem. 25,1953-1960.

Bjørnestad, E., Petersen, G.I., Robson, M., Reiersen, L.-O., Henriquez, L., Massie, L. and Blackman, R. 1993. Paris Commission ring test: testing of offshore chemicals and drilling mud on selected marine organisms. Science Total Environ Suppl 1993: 713-719.

Breitholtz M, Bengtsson B-E 2001. Oestrogens have no hormonal effect on the development and reproduction of the Harpacticoid Copepod *Nitocra spinipes*. *Mar. Pollut. Bull.* 42(10), 879-886.

Breitholtz M, Wollenberger L, Dinan L 2003. Effects of four synthetic musks on the life cycle of the harpacticoid copepod *Nitocra spinipes*. *Aquat. Toxicol.* 63, 103-118.

Breitholtz M, Wollenberger L 2003. Effects of three PBDEs on development, reproduction and population growth rate (r_m) of the harpacticoid copepod *Nitocra spinipes*. *Aquat. Toxicol.* 64, 85-96.

Breitholtz, M., Rudén, C., Hansson S.O. and Bengtsson, B-E. 2006. Ten challenges for improved environmental risk assessment. *Ecotoxicology and Environmental Safety* 63, 324-335.

Chandler, G.T., Green, A.S. 2001. Developmental stage-specific life-cycle bioassay for assessment of sediment-associated toxicant effects on benthic copepod production. Environmental Toxicology and Chemistry 20, 171-178.

Chandler, G.T., Cary, T.L., Volz, D.C., Walse, S.S., Ferry, J.L., Klosterhaus, S.L. 2004a. Fipronil effects on copepod development, fertility, and reproduction: A rapid life-cycle assay in 96-well microplate format. Environmental Toxicology and Chemistry 24, 177-124.

Chandler, T.G., Cary, T.L., Bejarano, A.C., Pender, J.L. and Ferry, J.L. 2004b. Population consequences of fipronil and degradates to copepods at field concentrations: An integration of life cycle testing with Leslie matrix population Modeling. Environmental Science and Technology 38 (23), 6407-6414.

Coull BC, Chandler GT. 1992 Pollution and meiofauna: field, laboratory and mesocosm studies. *Oceanog. Mar. Biol. Ann. Rev*. 30:191-271.

Hicks, G.R.F., Coull, B.C. 1983. The ecology of marine meiobenthic harpacticoid copepods. *Ocenogr. Mar. Biol. Ann. Rev*. 21, 67-125.

Hutchinson, T.H., Pounds, N.A., Hampel, M., Williams, T.D. 1999a. Impact of natural and synthetic steroids on the survival, development and reproduction of marine copepods (*Tisbe battagliai*). *Sci. Tot. Environ*. 233, 167-179.

Hutchinson, T.H., Pounds, N.A., Hampel, M., Williams, T.D. 1999b. Life-cycle studies with marine copepods (*Tisbe battagliai*) exposed to 20-hydroxyecdysone and diethylstilbestrol. *Environ. Toxicol. Chem*. 18(12), 2914-2920.

Huys, R., Gee, J.M., Moore, C.G., Hamond, R. 1996. *Marine and brackish water harpacticoid copepods (Part 1).* (Huys, R., Gee, J. M., Moore, C. G. and Hamond, R. Eds.), Synopsis of the British Fauna (New Series), The Dorset Press, Great Britain, ISBN: 1-85153-256-0.

Huyskens, P., Mullens, J. and Gomez, A. 1975. Solubility of alcohols, phenols and anilines in water. *Bull Soc Chimiq Belges* 84 (3), 253-262

Ingersoll, C.G., Hutchinson, T.H., Crane, M., Dodson, S., DeWitt, T., Geis, A., Huet, M.C., McKenney, C.L., Oberdörster, E., Pascoe, D., Versteeg, D.J., Warwick, O. 1999. Laboratory tests for evaluating potential effects of endocrine disrupting Componds. In: *Endocrine disruption in invertebrates: SETAC Technical Publication* (de Fur, P.J., Crane, M., Ingersoll, C.G., Tattersfield, L.J. Eds.), Brussels, Belgium, ISBN 1-880611-27-9, pp.107-109.

ISO 1997. International Organization for Standardization. Water Quality - Determination of acute lethal toxicity to marine copepods (Copepoda, Crustacea). Draft International Standard ISO/DIS 14669. Geneve, Switzerland.

Knuckey RM, Semmens GL, Mayer RJ, Rimmer MA. 2005. Development of an optimal microalgal diet for the culture of the calanoid copepod *Acartia sinjiensis*: Effect of algal species and feed concentration on copepod development. Aquaculture *In press*.

Kuch H.M. and K. Ballschmiter 2001. Determination of endocrine-disrupting phenolic compounds and estrogens in surface and drinking water by HRGC-(NCI)-MS in the picogram per liter range *Environ Sci Technol* 35, 3201.

Liber, K., N.K. Kaushik, K.R. Solomon and J.H. Carey. 1992. Experimental designs for aquatic mesocosm studies: A comparison of the "ANOVA" and "Regression" design for assessing the impact of tetrachlorophenol on zooplankton populations in limnocoralls. Environ. Tox. Chem. 11, 61-77.

Mu X, LeBlanc GA. 2002. Environmental antiecdysteroids alter embryo development in *Daphnia magna. J. Exp. Zool*., 292, 287-292.

Mu X, LeBlanc GA. 2004. Synergistic interaction of endocrine-disrupting chemicals: Model development using an ecdysone receptor antagonist and a hormone synthesis inhibitor. *Environ. Toxicol Chem.* 23, 1085-1091.

OECD (Organization for Economic Cooperation and Development). 2005a. *Proposal for a New Guideline: Calanoid Copepod Development and Reproduction Test with* Acartia tonsa. OECD Draft Guidelines for Testing of Chemicals. 39 pp.

OECD (Organization for Economic Cooperation and Development). 2005b. *Proposal for a New Guideline: Harpacticoid Copepod Development and Reproduction Test*. OECD Draft Guidelines for Testing of Chemicals. 34 pp.

OECD (Organization for Economic Cooperation and Development). 2005c. Pre-validation of Full Life-Cycle test with the Harpacticoid Copepods *Nitocra spinipes, Tisbe battagliai* and *Amphiascus tenuiremis* (Authors: *M. Breitholtz, T.D. Williams and G.T. Chandler*). September 2005.

Renberg, L. 1981. Gas chromatographic determination of phenolic compounds in wastewater as their pentafluorobenzoyl derivatives. *Chemosphere* 10(7), 767-73.

Shiu, W.Y., Ma K.C., and Varhanickova, D. 1994. Chlorophenols and alkylphenols - a review and correlation of environmentally relevant properties and fate in an evaluative environment. *Chemosphere* 29 (6): 1155-1224.

SIS 1991. Determination of acute lethal toxicity of chemical substances and effluents to *Nitocra spinipes* Boeck - Static procedure (in Swedish) Swedish Standard SS 02 81 06. SIS - standardiseringskommissionen i Sverige, Stockholm, Sweden. 17 pp.

SPSF 2002. Standard Project Submission Form for a New Test Guideline on Copepod Development and Reproduction. Submitted by Sweden.

Wesenberg-Lund, C. 1937. *Ferskvandsfaunen, Biologiskt belyst, Invertebrata.* Gyldendanske Boghandel - Nordiskt Forlag, Copenhagen, Denmark.

Williams DR, Fisher MJ, and Rees HH. 2000. Characterization of ecdysteroid 26-hydroxylase: An enzyme involved in molting hormone inactivation. *Arch. Biochem. Bipphys.* 376, 389-398.

Zarorc-Konĉan, J., Gotvajn, A Z and Tisler, T. 2002. Hazard idntification for 3,5-dichlorophenol in the aquatic environment. Cellular & Microbiol Lett 7 (2): 381-382.

APPENDIX A

Figure 1

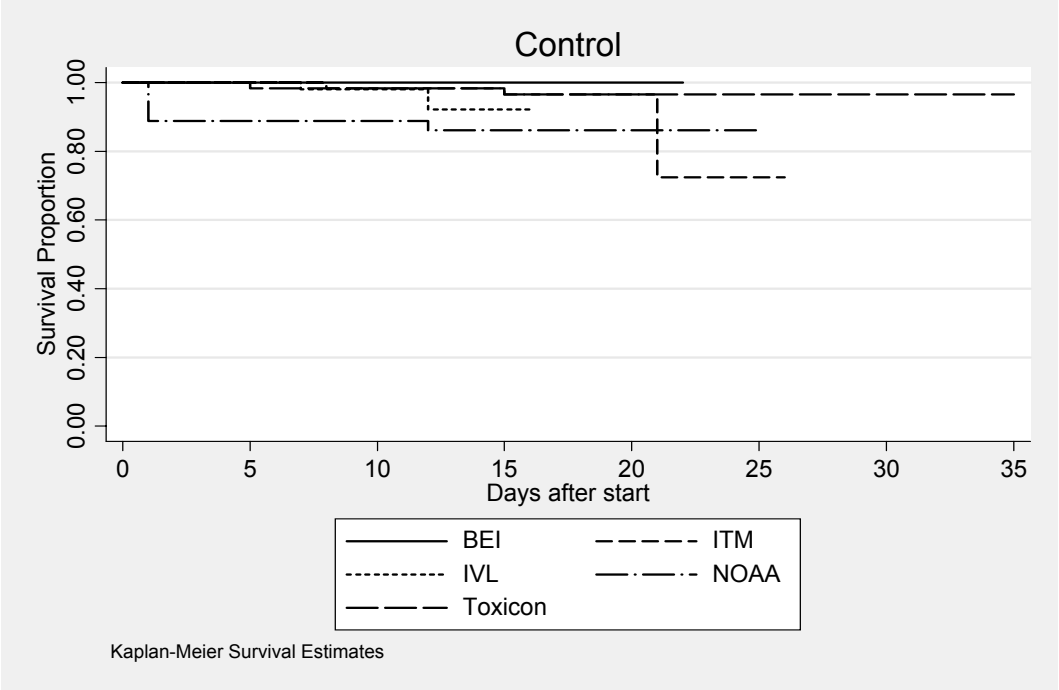


Figure 2

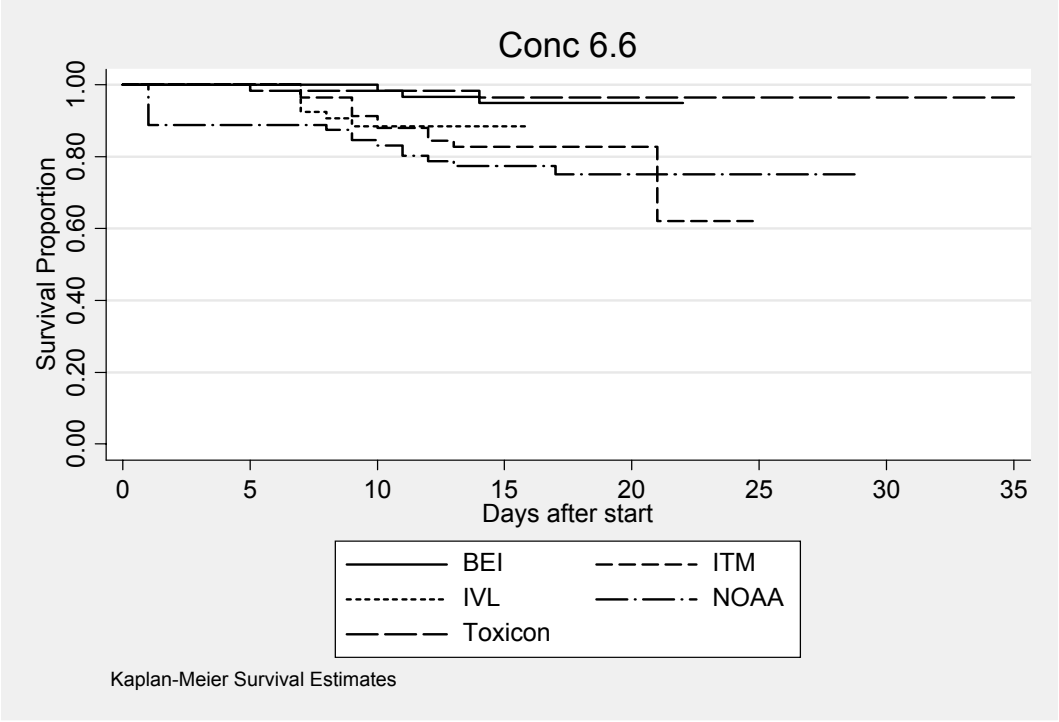


Figure 3

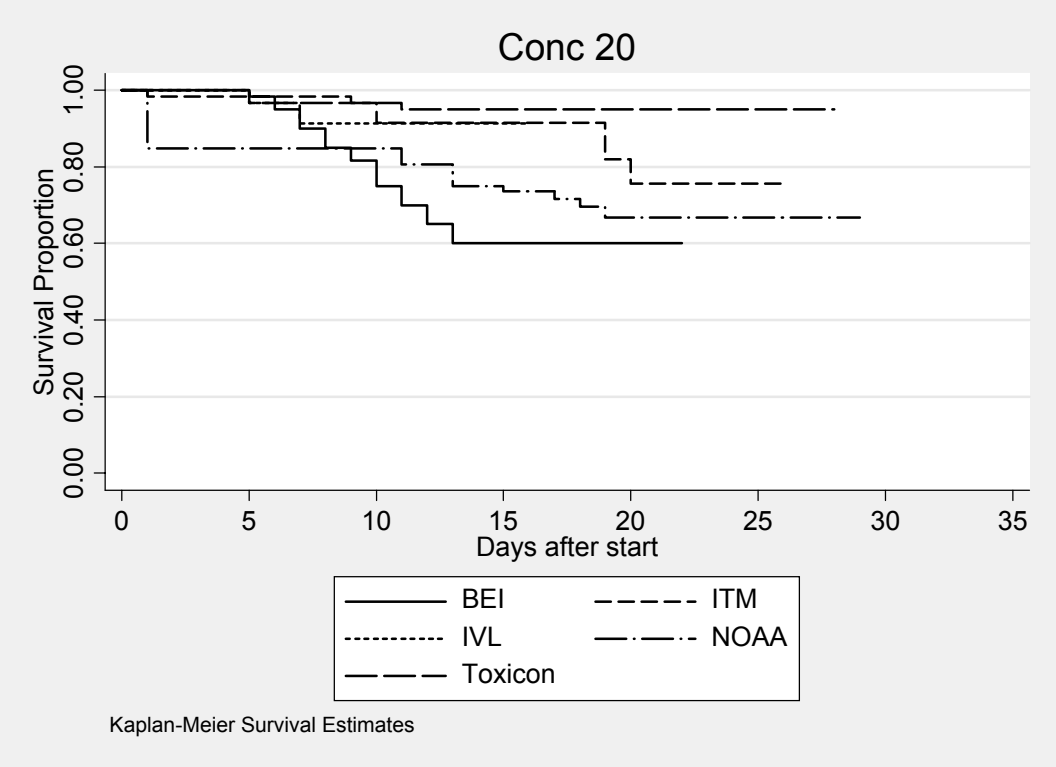


Figure 4

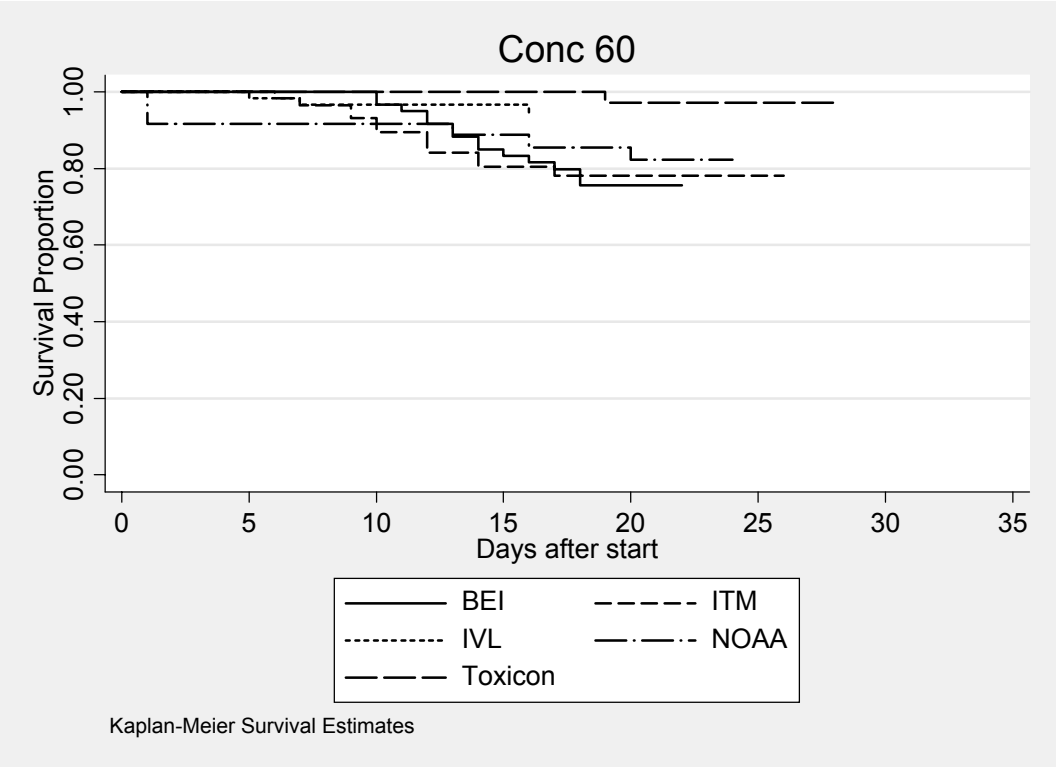


Figure 5

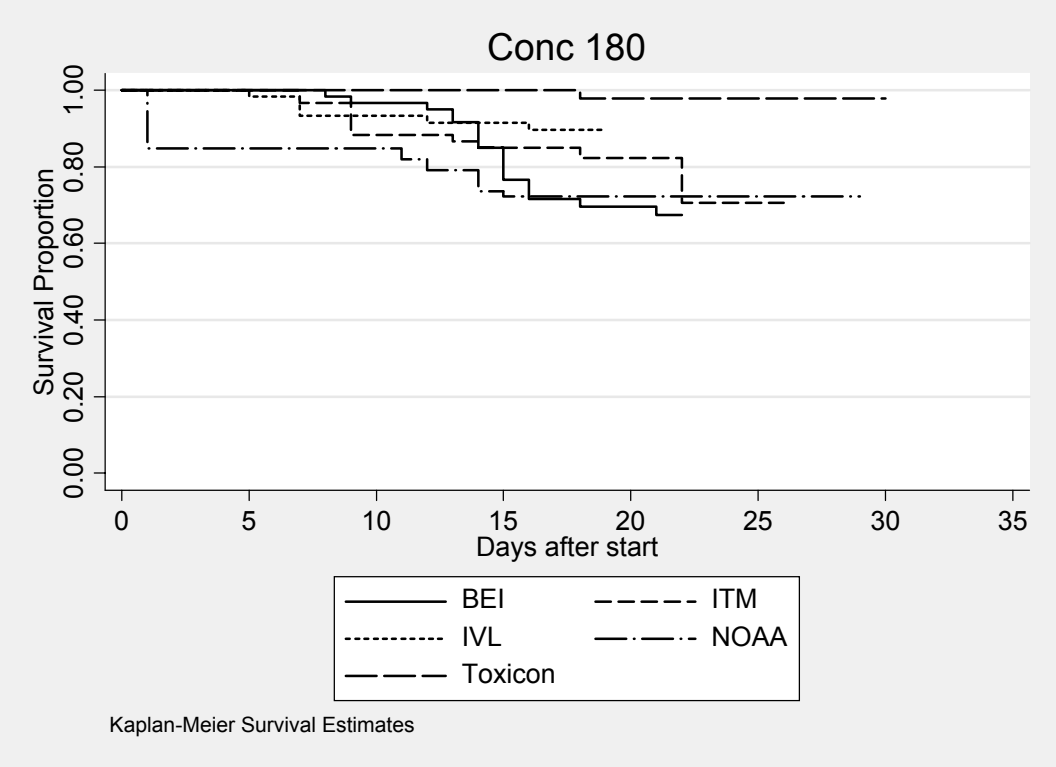


Figure 6

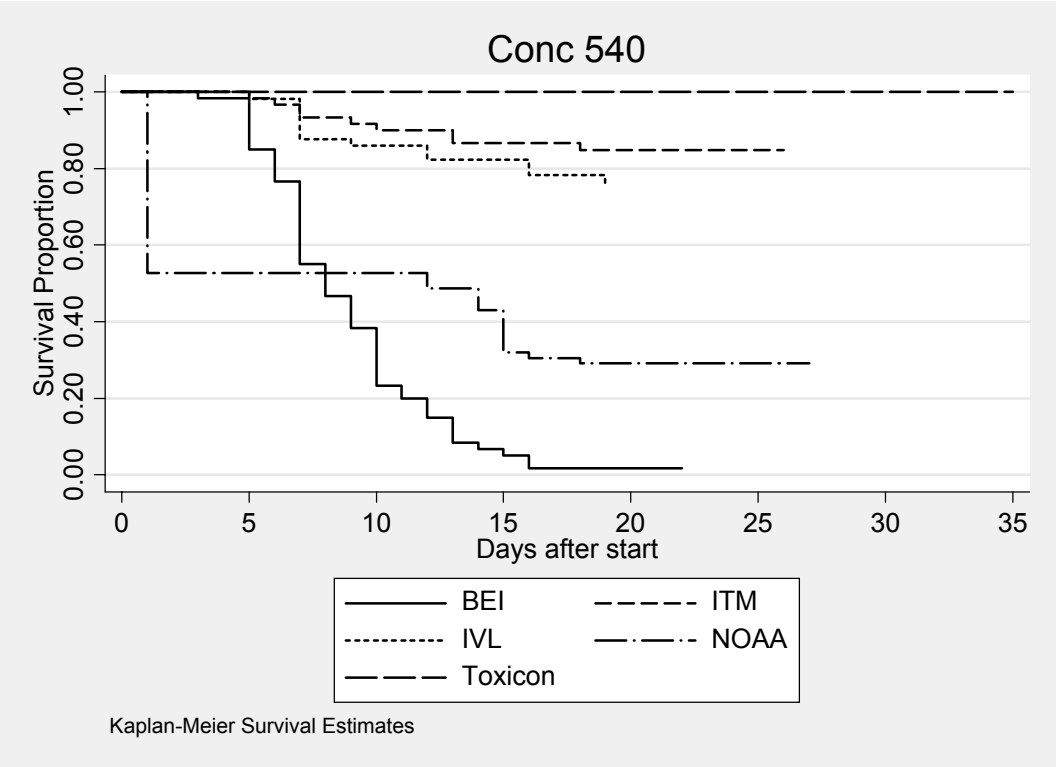


Figure 7

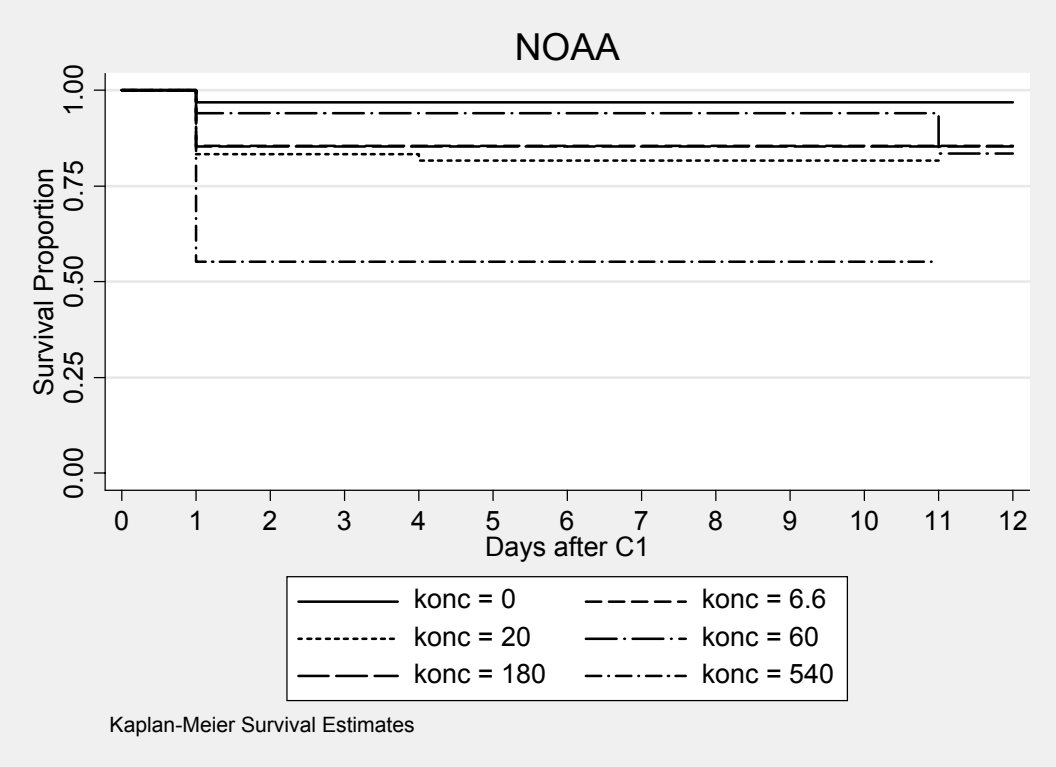


Figure 8

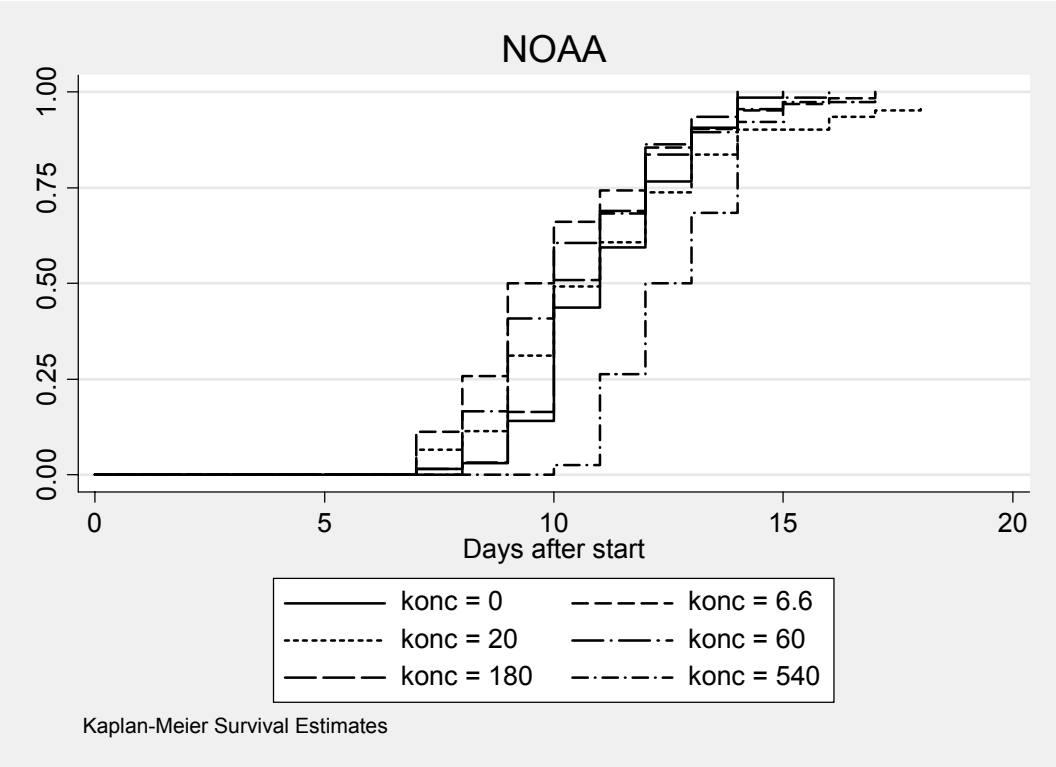


Figure 9

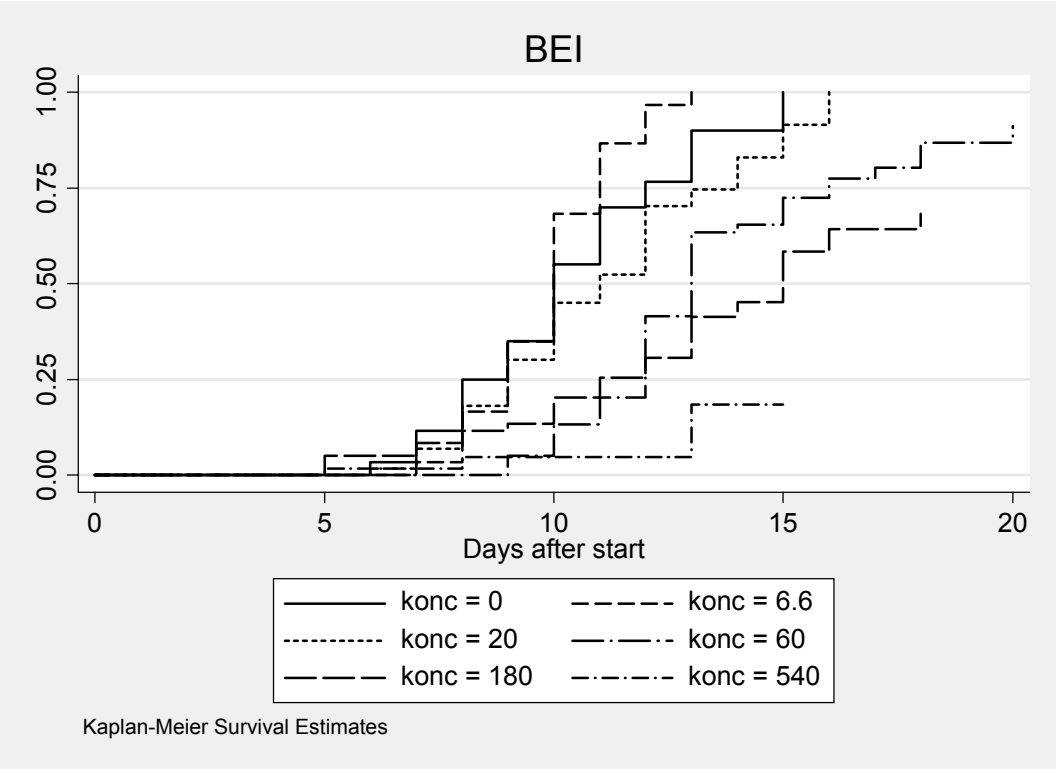


Figure 10

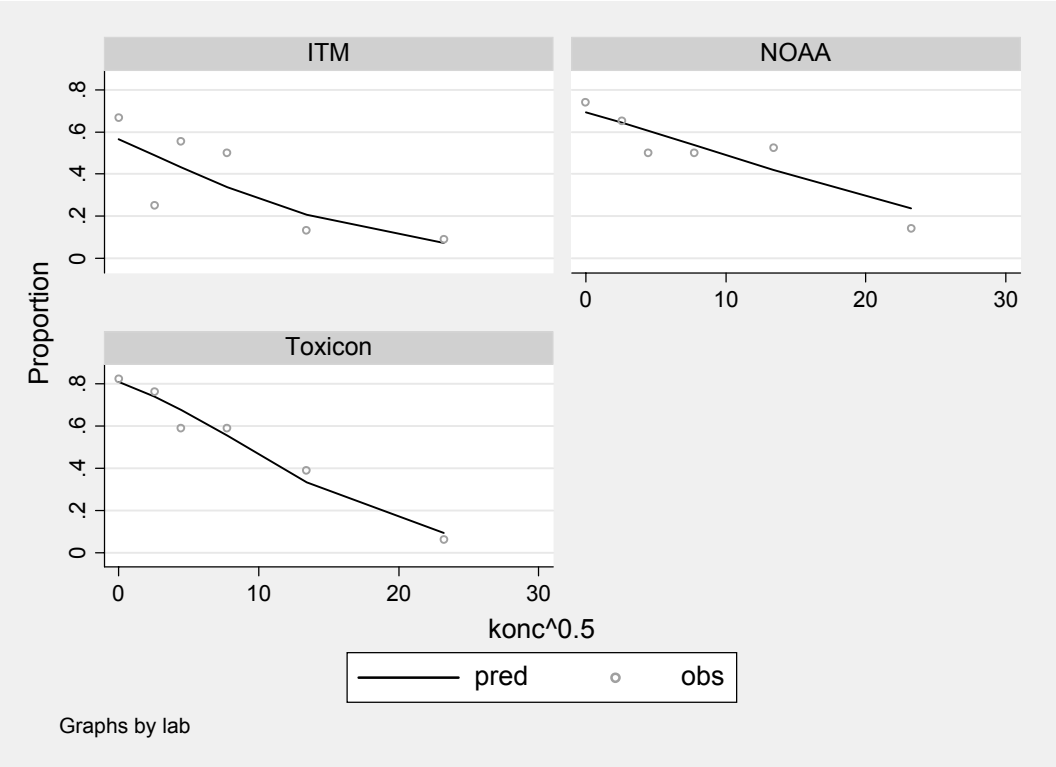


Figure 11

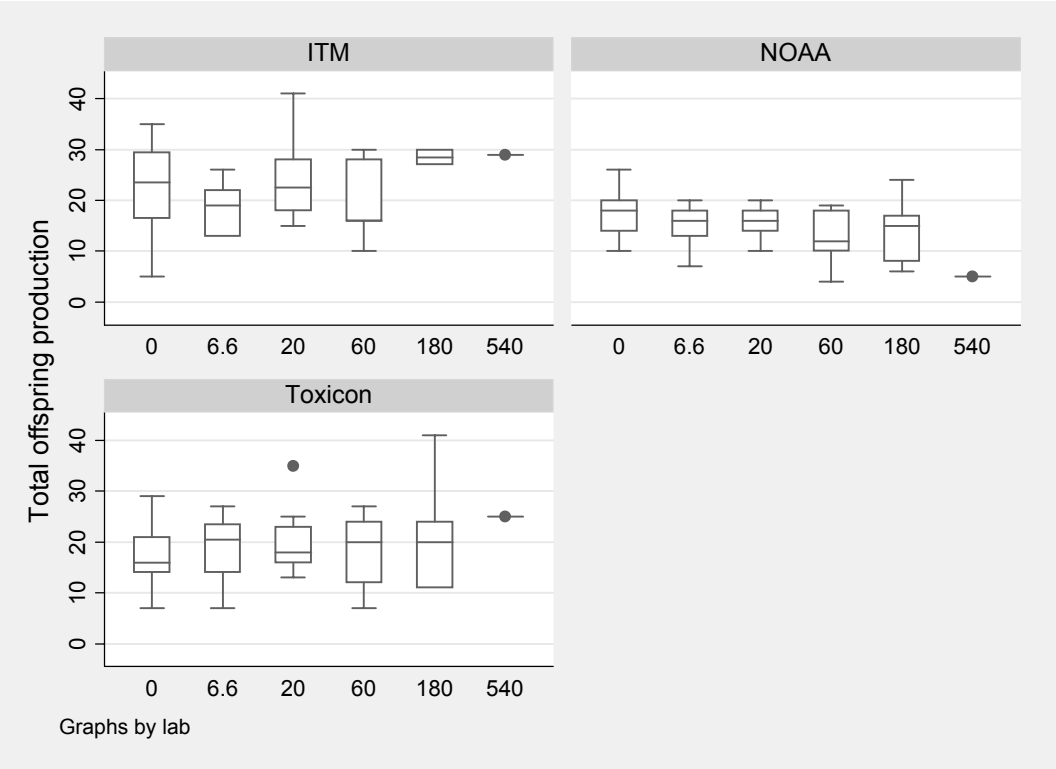


Figure 12

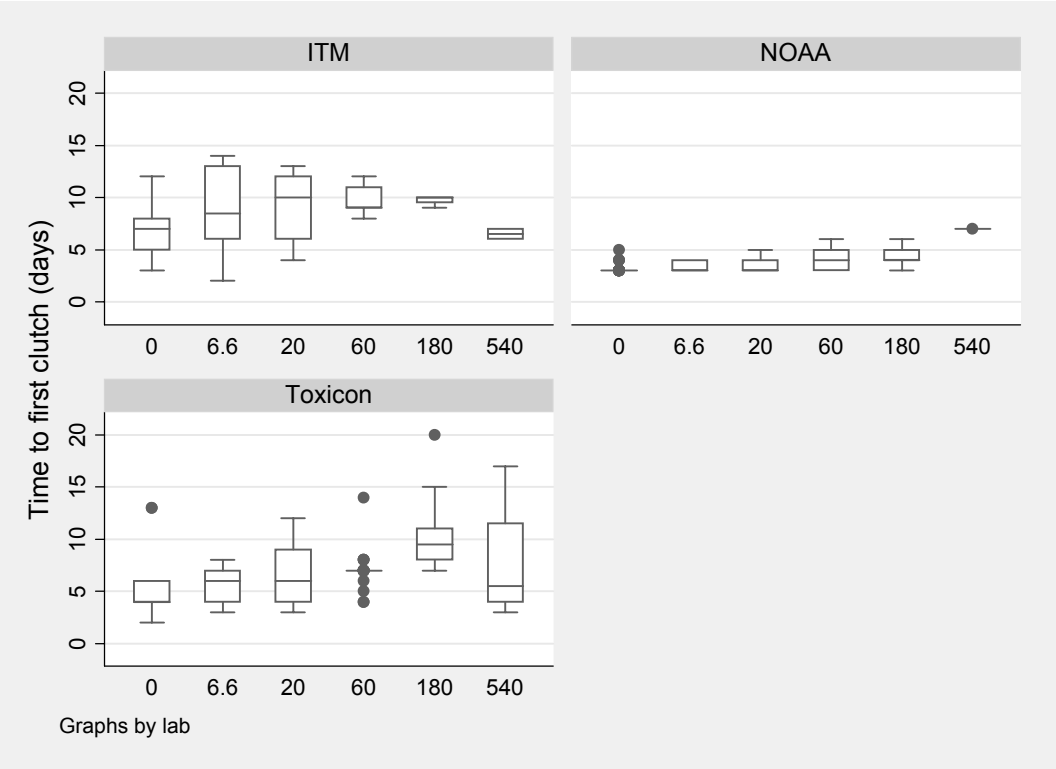
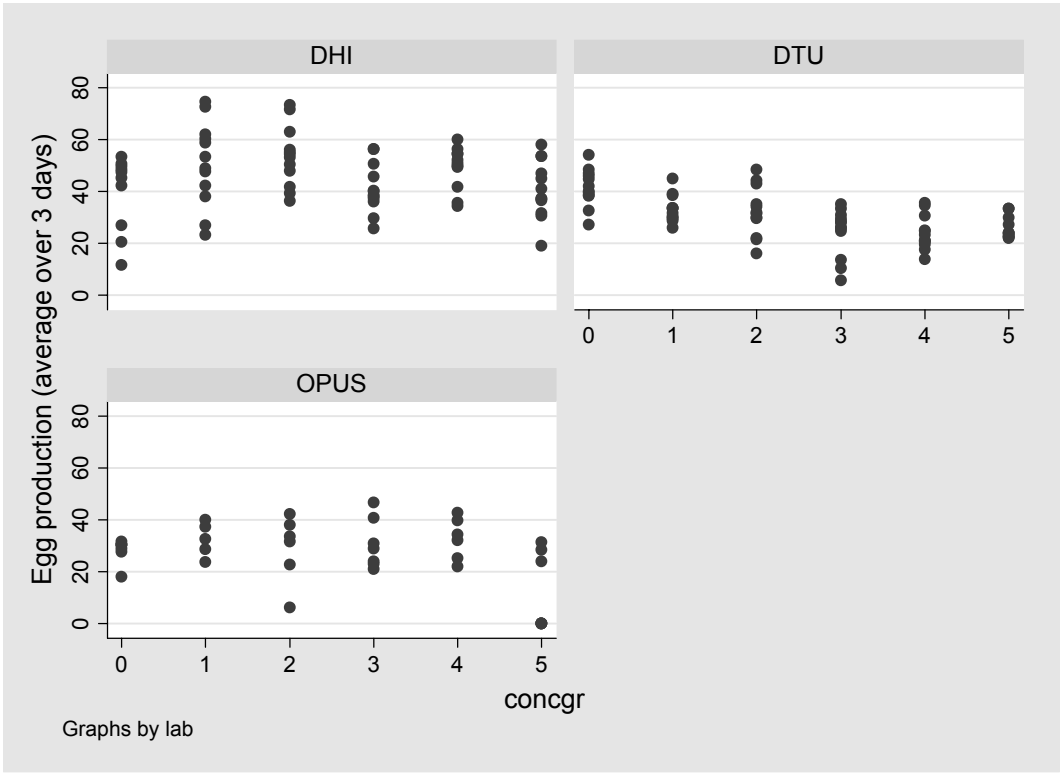


Figure 13. Average egg production (over 3 days) against concentration (Control=0).



APPENDIX B

VALIDATION OF THE COPEPOD DEVELOPMENT AND REPRODUCTION TEST, PHASE 1

DATA ANALYSIS

Jan-Olov Persson
Mathematical Statistics
Stockholm University

1) Harpacticoid Copepod Development and Reproduction Test

DATA MANAGEMENT

All data was delivered in Excel format from ITM and then transferred to, and analysed with the statistical software Stata version 8 (http://www.stata.com).

There were development data from all five laboratories; Adriana[6], Chandler[7], ITM, IVL and Toxicon, and reproduction data from Adriana, ITM, and Toxicon.

Development test

In the development test, the following outcome was analysed: mortality, development rate and sex ratio.

Mortality

a) Mortality in controls

For a test to be valid, the mortality in the control must not exceed 20% at the end of the exposure period (Test protocol Appendix A). Whether these requirements have been fulfilled was examined by the use of the non-parametric Kaplan Meier estimate of the survival function (1). In this analysis, an individual who turned out to be adult or missing was handled as censored at that specific time point. Also individuals still alive at the end of experiment were treated as censored.

The length of the experiment period differed between laboratories (Adriana 29 days, Chandler 22 days, ITM 26 days, IVL 16 or 19 days[8], Toxicon 35 days) and the mortality was examined both the last day (over all mortality) and at day 16.

Table 1 shows the estimated survival proportions, over all, with 95% confidence intervals. Table 2 shows the estimated survival proportions, at day 16, with 95% confidence intervals. The Kaplan Meier survival curves are shown in Figure 1 in appendix.

---

[6] In this report the name Adriana is used. The correct name is The NOAA Center for Coastal Environmental Health and Biomolecular Research

[7] In this report the name Chandler is used. The correct name is Block Environmental Inc.

[8] 19 days for concentration 180 and 540

Table 1. Over all survival

Lab	Survival	Lower_95	Upper_95
Adriana	0.861	0.757	0.923
Chandler	1.000	.	.
ITM	0.724	0.154	0.946
IVL	0.922	0.805	0.970
Toxicon	0.965	0.865	0.991

Table 2. Survival until day 16

Lab	Survival	Lower_95	Upper_95
Adriana	0.861	0.757	0.923
Chandler	1.000	.	.
ITM	0.965	0.867	0.991
IVL	0.922	0.805	0.970
Toxicon	0.965	0.865	0.991

For ITM the over all mortality exceeds 20 % (27.6%), but the mortality until day 16 is tolerable. The other laboratories show a tolerable mortality for this particular experiment.

An examination of the Kaplan Meier survival curve (Figure 1) shows a massive mortality at day one (8 individuals out of 72).

For Adriana we cannot be statistically certain that the "population mortality" is lower than 20 % (since the lower 95% confidence limit is below 80%).

b) Mortality in exposed / comparison with control

The mortality among exposed was investigated in the same manner as the control (estimation of the Kaplan Meier survival function). Figure 2 -4 shows the survival curves for three of the concentrations; 6.6, 20 and 540. For some of the laboratories there is a significantly higher mortality for the exposed compared to the control (details to be reported by Magnus Breitholtz).

The mortality was also investigated separately for the different development stages (nauplii to copepodite and copepodite to adult). One peculiar example is shown in Figure 5; the survival in the copepodite to adult stage, for Adriana. Almost all deaths occur one day after reaching copepodite stage, which was surprising and difficult to explain. There are probably some explanations other than randomness for that.

Magnus Breitholtz will report more details and graphs on the stage specific mortality.

Development

The development rates have to be estimated for the different development stages (nauplii to copepodite and copepodite to adult) separately. The developments data should also be treated as survival data were the outcome variable is copepodite/adult (yes or no), and were censored observations are present since the time to copepodite/adult is not observed for every individual. A first approach to the analysis was to fit a parametric survival function to data (lognormal or gamma), but it did not fall out well. The reason for that could be that this particular data is not pure survival data; there may perhaps be individuals that never become copepods or adults even if we follow them to infinity. In order to take this in account a more sophisticated model would have been needed. As a simpler alternative the analysis was based on the nonparametric Kaplan Meier survival estimate.

In the analysis of development rate from nauplii to copepodite, an individual who turn out to be dead or is missing, was handled as censored and a nauplii that did not develop to copepodite during the follow up time was also censored.

The analysis of development rate from copepodite to adult was performed on individuals that had developed to copepodites during the follow up time. If an individual was dead, missing or did not become an adult within the follow up time, it was handled as censored at that specific time point.

There are different measures of development rate which can be used, e.g. mean time, median time, proportion developed at a particular time point, and it is not straightforward how to choose the most suitable measure for this data. The first candidate is the median value, which is an ordinary measure for survival data. For this particular data, the median will be a quite rough measure. The reason for that is the low precision of the time measurements compared to the quite small exposure effect. The time measurements are rounded up to one whole day and consequently the median have a value in a whole day. The second measure candidate is the mean value. The mean value is calculated as the area under the Kaplan Meier curve, and that measure should be less rough and more appropriate, if we had pure survival data and if the follow up time was longer. For data in this study, the mean value may be sensitive to the choice of endpoint time. For that reason the median (50th percentile) complemented with the 25th percentile was chosen. In addition to that, a Log rank test was performed to test equality of the survival curve between an exposure and the control. The results are presented in Table 3 and Table 4 and the corresponding Kaplan Meier failure curves are shown in Figure 6 to 10 in appendix (only for the development from nauplii to copepodite).

Note: Individuals that died day 1 from start were excluded from the study.

Table 3. Time in days to copepodite, 25th and 50th percentile.

P-value of Log-rank test

Lab and conc.	p25	p50	p-value
Adriana			
0	10	11	
6.6	8	9	.0425
20	9	11	.6
60	9	10	.155
180	10	10	.258
540	11	12	.000137
Chandler			
0	8	10	
6.6	9	10	.0265
20	9	11	.103
60	12	13	1.21e-08
180	11	15	4.30e-10
540			2.75e-08
ITM			
0	9	9	
6.6	9	10	.432
20	9	9	.967
60	9	11	.00437
180	9	10	.0146
540	10	12	.000208
IVL			
0	8	9	
6.6	8	9	.0834
20	8	9	.0469
60	8	9	.00503
180	12	12	.0272
540	12	14	2.08e-06
Toxicon			
0	7	8	
6.6	8	8	.601
20	6	7	.0000705
60	6	7	.000456
180	8	9	.539
540	7	8	.431

Table 4. Time in days between copepodite and adult, 25th and 50th percentile. P-value of Log-rank test

Lab and conc.	p25	p50	p-value
Adriana			
0	8	8	
6.6	6	7	.136
20	6	7	.19
60	7	8	.825
180	7	8	.622
540	9	10	.000234
Chandler			
0	8	8	
6.6	7	9	.155
20	6	8	.597
60	6	8	.0325
180	9	9	.0273
540			
ITM			
0	6	7	
6.6	7	8	.0121
20	7	8	.00129
60	6	8	.231
180	7	8	.0114
540	8	9	6.15e-06
IVL			
0	4	5	
6.6	4	7	.125
20	4	6	.145
60	4	7	.0353
180	4	7	.0265
540	4	5	.174
Toxicon			
0	10	13	
6.6	11	12	.58
20	11	13	.533
60	10	14	.272
180	10	11	.241
540	13	15	.0214

Sex ratio

For a test to be valid, the proportion of females (or males) in the control should be within 40% to 60% (Test protocol Appendix A).

Table 5 shows the female proportion for each laboratory and concentration, and for the four laboratories that have made sex determination, the female proportion in the control is within the limits.

Table 5. Female proportion in adults

	Lab.				
conc.	Adriana	Chandler	ITM	IVL	Toxicon
0	0.53	0.56	0.57		0.43
6.6	0.49	0.50	0.43		0.49
20	0.52	0.53	0.60		0.45
60	0.58	0.58	0.55		0.46
180	0.44	0.78	0.53		0.47
540	0.29		0.53		0.31

A notification was made about the time range from copepodite to adult for the control, among individuals that had become adults (not a survival analysis estimate):

For Adriana and Chandler there were no significant differences between sexes. For ITM there was a significant difference (mean time: female 7.9 days, male 6.4 days). For Toxicon there was a large significant difference (mean time: female 10.4 days, male 14.3 days).

For ITM this difference was due to difference from nauplii to copepodite but for Chandler there was no difference from nauplii to copepodite.

Reproduction test

Three laboratories performed the reproduction test (Adriana, ITM and Toxicon) and the outcomes analysed was fertilisation success, offspring production and time to first and second clutch. No survival analysis was made since information about the censored pairs was missing. We cannot therefore exclude that an observed effect for an exposed group might be a consequence of a differences in mortality for that group compared to the control.

Fertilisation success

The fertilization success was measured as the proportion of pairs able to produce two successive clutches within the follow up time, of all mating pairs at start[9]. The follow up time differs between

[9] ITM observed that several of their pairs consisted of two females. These pairs were excluded.

laboratories (Adriana 10 days, ITM 19 days and Toxicon 20 days), which possibly make the comparisons between laboratories invalid. The results are shown in Table 6 to 8; number of mating pairs at start, the proportion of pairs with two clutches together with exact 95 % confidence limits and a p-value of testing the hypothesis of similar proportions in the exposed and the control (Fisher's exact test) .

Table 6. Proportion of pairs with two clutches, for Adriana.

Conc	Freq	Prop	Lower_95	Upper_95	p-value
0	23	.74	.52	.9	.
6.6	23	.65	.43	.84	.75
20	20	.50	.27	.73	.13
60	18	.50	.26	.74	.19
180	21	.52	.3	.74	.21
540	7	.14	.0036	.58	.0086

Table 7. Proportion of pairs with two clutches, for ITM.

Conc	Freq	Prop	Lower_95	Upper_95	p-value
0	18	.67	.41	.87	.
6.6	20	.25	.087	.49	.021
20	18	.56	.31	.78	.73
60	10	.50	.19	.81	.44
180	15	.13	.017	.4	.004
540	11	.091	.0023	.41	.0057

Table 8. Proportion of pairs with two clutches, for Toxicon.

Conc	Freq	Prop	Lower_95	Upper_95	p-value
0	17	.82	.57	.96	.
6.6	21	.76	.53	.92	.71
20	22	.59	.36	.79	.17
60	22	.59	.36	.79	.17
180	18	.39	.17	.64	.015
540	16	.063	.0016	.3	.000013

The observed proportions of pairs able to produce two clutches are lower for the exposed compared to the control, for all concentrations in all three laboratories. If we use significance level[10] 5%, every one of the laboratories show significant effects of concentration 540, and ITM along with Toxicon

[10] not adjusted for multiple test

also show significant effects of concentration 180. In addition, ITM show significant effect for concentration 6.6.

The probability that a statistical test of this kind will fall out significant (the power of the test) depends on the size of the true effect and the sample size. Obviously, we cannot conclude, for a non-significant test, that there are no effects. The power might just have been to low to reveal it. Of that reason, no attempt was made to estimate NOEC or LOEC.

An alternative approach is to assume a model for the dose response relationship. For this data, there seems to be a monotonic dose response trend (with exception on concentration 6.6 for ITM) and a logistic regression model was fitted for each laboratory separately. The models are illustrated in Figure 10 in Appendix, and except for the observation from ITM on concentration 6.6, the fit is quite good.

In order to estimate a EC_{50} -value a non-linear model, would have to be fitted. This is not a straightforward task (model choice, programming, model check etc.) and because of the rather low quality of data (we cannot exclude that the effects seen, depends on mortality) and shortage of time, no such attempt was made.

Offspring production

For a test to be valid the average number of viable offspring per clutch in the controls should not be lower than 5 (Appendix A). This criterion was met for all three laboratories (average number was higher than 8 for all three, data not shown).

For pairs that were able to produce two clutches, the total number of viable offspring was calculated. The results are displayed separately for each laboratory in Table 9 to 11. The tables show the frequency of mating pairs able to produce two clutches, and the mean value of total offspring production together with 95 % confidence limits (based on the t-distribution). A one-way ANOVA for each laboratory separately shows no significant differences (p-values[11]; 0.10, 0.60 and 0.74 for laboratory Adriana, ITM and Toxicon respectively).

The results are also illustrated with box plots in Figure 11 in Appendix.

Table 9. Total viable offspring production, for Adriana.

Conc	Freq	Mean	Lower_95	Upper_95
0	17	17.4	15.3	19.4
6.6	15	14.7	12.7	16.8
20	10	15.7	13.4	18
60	9	12.7	8.77	16.6
180	11	14.1	10.2	18
540	1	5	.	.

[11] Concentration 540 with only one observation was excluded in this test.

Table 10. Total viable offspring production, for ITM.

Conc	Freq	Mean	Lower_95	Upper_95
0	12	22.3	16.3	28.2
6.6	5	18.6	11.5	25.7
20	10	23.7	18.2	29.2
60	5	20	9.32	30.7
180	2	28.5	9.44	47.6
540	1	29	.	.

Table 11. Total viable offspring production, for Toxicon.

Conc	Freq	Mean	Lower_95	Upper_95
0	14	16.9	13.3	20.4
6.6	16	19	15.7	22.3
20	13	19.8	16.1	23.4
60	13	17.6	13.3	21.9
180	7	20.3	10.6	30
540	1	25	.	.

Time to first and second clutch

Table 12 to 14 shows the median time to the first clutch, estimated for the pairs that had produced a first clutch (not a survival analysis estimate). A Mann Whitney U-test was performed to test the hypothesis of equal time distribution in the exposed and the control. P-values on these tests are shown in the table. Since we have ignored the censored observations, the interpretation of the result should be careful. The results are also illustrated with box plots in Figure 12 in Appendix.

Table 12. Time to production of first clutch, for Adriana.

Conc	Freq	Median	p_value
0	17	3	
6.6	17	3	.79
20	11	3	.48
60	11	4	.029
180	13	4	.00027
540	1	7	

Table 13. Time to production of first clutch, for ITM.

Conc	Freq	Median	p_value
0	15	7	
6.6	6	8.5	.5
20	13	10	.082
60	7	9	.016
180	4	10	.049
540	2	6.5	.55

Table 14. Time to production of first clutch, for Toxicon.

Conc	Freq	Median	p_value
0	15	4	
6.6	18	6	.18
20	17	6	.21
60	18	7	.0041
180	12	9.5	.00056
540	4	5.5	.61

Table 15 to 17 shows the median time between the first and second clutch, estimated for the pairs that had produced a second clutch (not a survival analysis estimate). A Mann-Whitney U-test was performed to test the hypothesis of equal time distribution in the exposed and the control. P-values on these tests are shown in the table. Since we have ignored the censored observations, the interpretation of the result should be careful

Table 15. Time between first and second clutch, for Adriana.

Conc	Freq	Median	p_value
0	17	1	
6.6	15	1	.65
20	10	1	.19
60	9	2	.59
180	11	2	.18
540	1	3	

Table 16. Time between first and second clutch, for ITM.

Conc	Freq	Median	p_value
0	12	3	
6.6	5	5	.67
20	10	4	.17
60	5	3	.91
180	2	3	.84
540	1	3	

Table 17. Time between first and second clutch, for Toxicon.

Conc	Freq	Median	p_value
0	14	3.5	
6.6	16	3	.35
20	13	3	.27
60	13	3	.42
180	7	4	.26
540	1	4	

2) Calanoid Copepod Development and Reproduction Test[12]

DATA MANAGEMENT

All data was delivered in Excel and Word format from ITM and then transferred to, and analysed with the statistical software Stata.

Data were from three laboratories: DHI, DTU and OPUS.

Development test

In the development test mortality and Larval development ratio (LDR) of F_0 was analysed. Analysis of LDR of F1 and Length of male and females will be reported by Magnus Breitholtz.

Mortality of F_0

For at test to be valid the average mortality of animals in the controls shall not exceed 30% of the hatched animals (Appendix B). Table 18 presents the number of replicates used, and the mortality proportion, for each concentration and laboratory. For the laboratories DTU and DHI, one replicate each was excluded due to disagreements in data (the number of units at the observation day was larger than at the start day). The average mortality in the controls (Conc.=0) is lower than 30% for every laboratory.

Table 18. Number of replicates and mortality proportion.

	Conc.					
Lab	0	6.6	20	60	180	540
DHI	9	4	4	3	4	4
	.18	.25	.19	.11	.22	.39
DTU	12	6	5	6	6	6
	.15	.21	.11	.15	.12	.17
OPUS	12	5	5	5	5	5
	.071	.12	.17	.091	.11	.53

[12] The analysis of the reproduction test will be reported by Magnus Breitholtz.

LDR of F_0

LDR was measured as the proportion copepodite of the viable animals (number of copepodites/(number of copepodites + number of nauplii)). Table 19 shows the mean value and standard deviation of LDR for each concentration and laboratory. Figure 13 in Appendix shows a scatter plot of LDR against concentration. No clear conclusion about the relationship between LDR and exposition can be made. For DHI an effect appear only for the highest concentration. For DTU the plot looks strange; LDR in the controls are surprisingly low, and the observations for concentration 20 are clustered into two groups. For OPUS there is a monotonous decreasing relationship with some extreme outliers.

Table 19. LDR of F_0, mean value and standard deviation.

	konc					
lab	0	6.6	20	60	180	540
DHI	.79	.74	.75	.76	.73	.29
	.037	.11	.057	.0049	.079	.11
DTU	.47	.58	.48	.45	.41	.043
	.049	.034	.18	.073	.037	.025
OPUS	.89	.65	.51	.24	.2	.033
	.071	.034	.21	.21	.14	.051

APPENDIX

Figure 1. Over all survival in the controls.

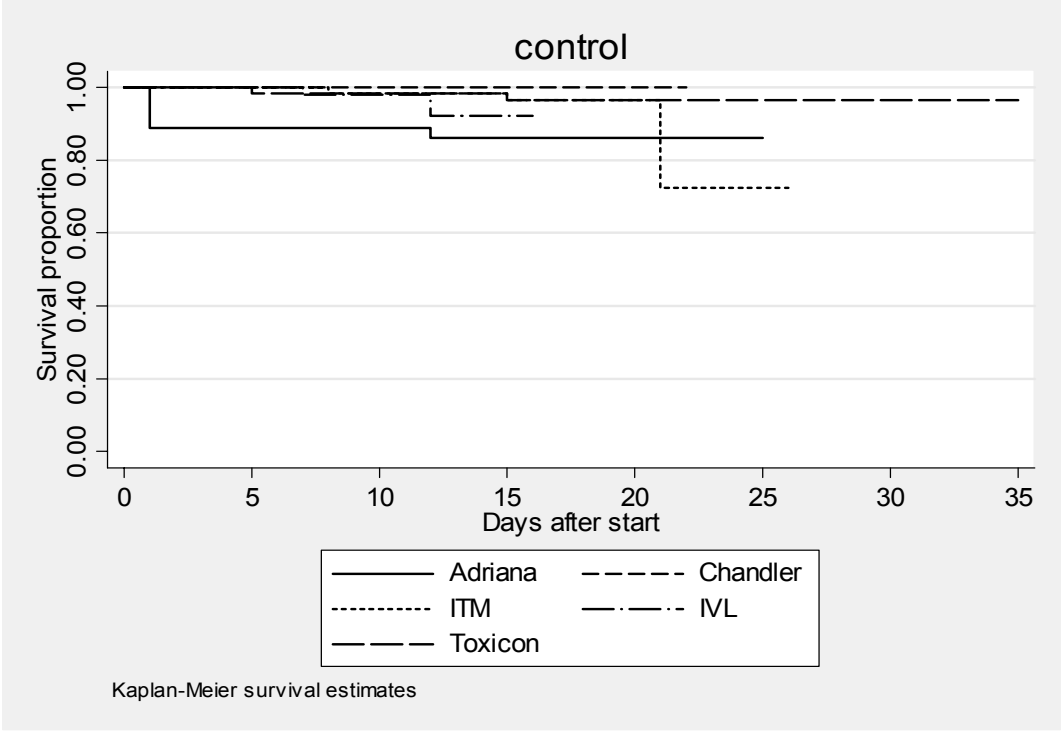


Figure 2. Over all survival in exposed, concentration 6.6.

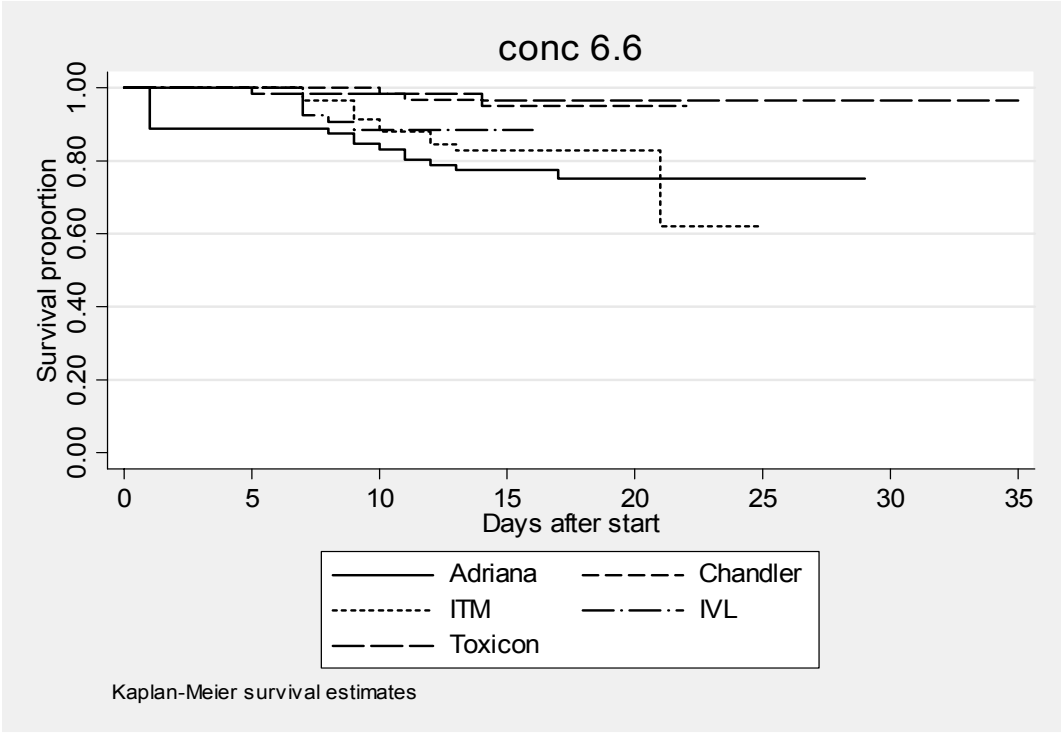


Figure 3. Overall survival in exposed, concentration 20.

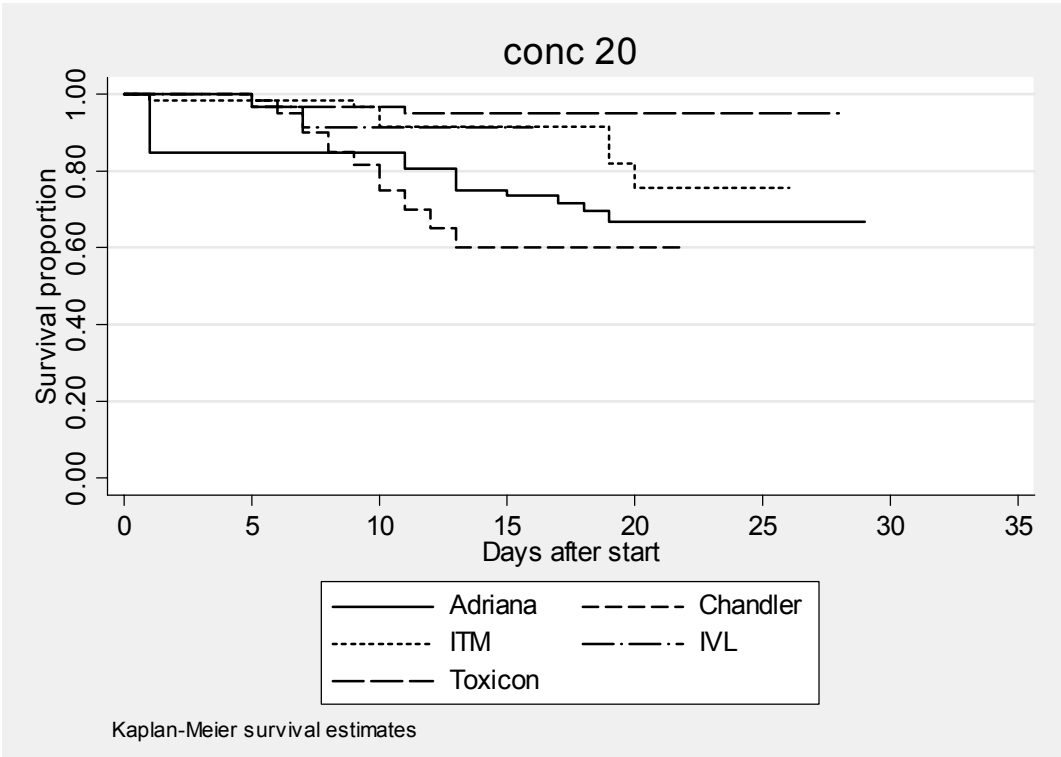


Figure 4. Over all survival in exposed, concentration 540.

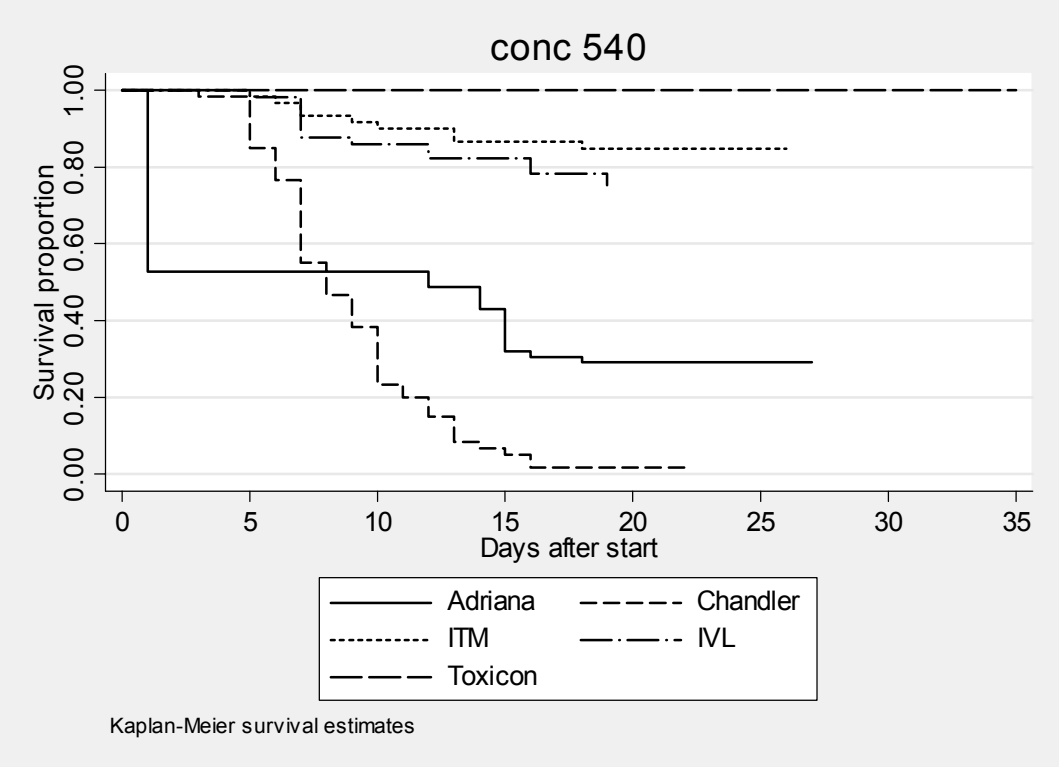


Figure 5. Survival in the copepodite to adult stage. Lab.: Adriana.

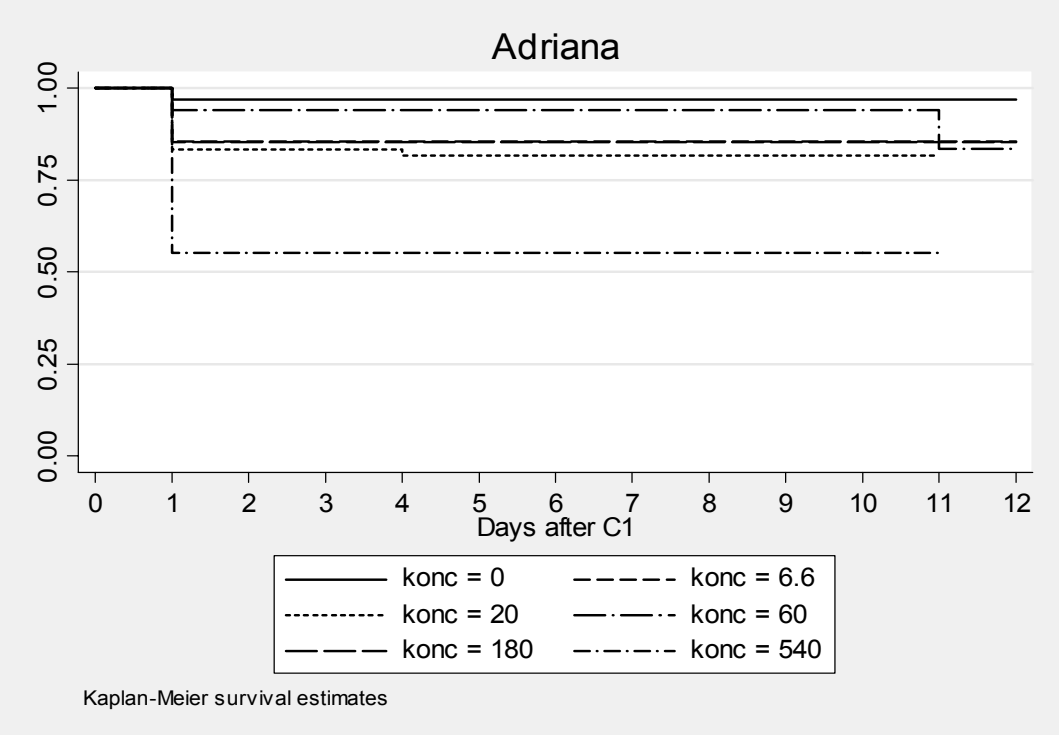


Figure 6. Development from nauplii to copepodite, Adriana.

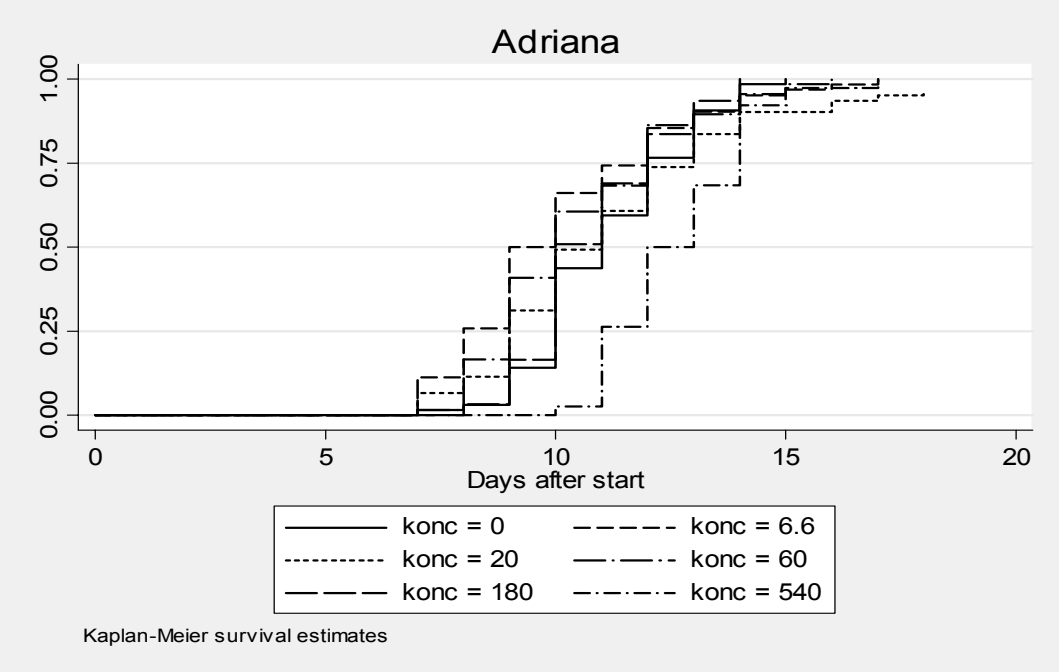


Figure 7. Development from nauplii to copepodite, Chandler.

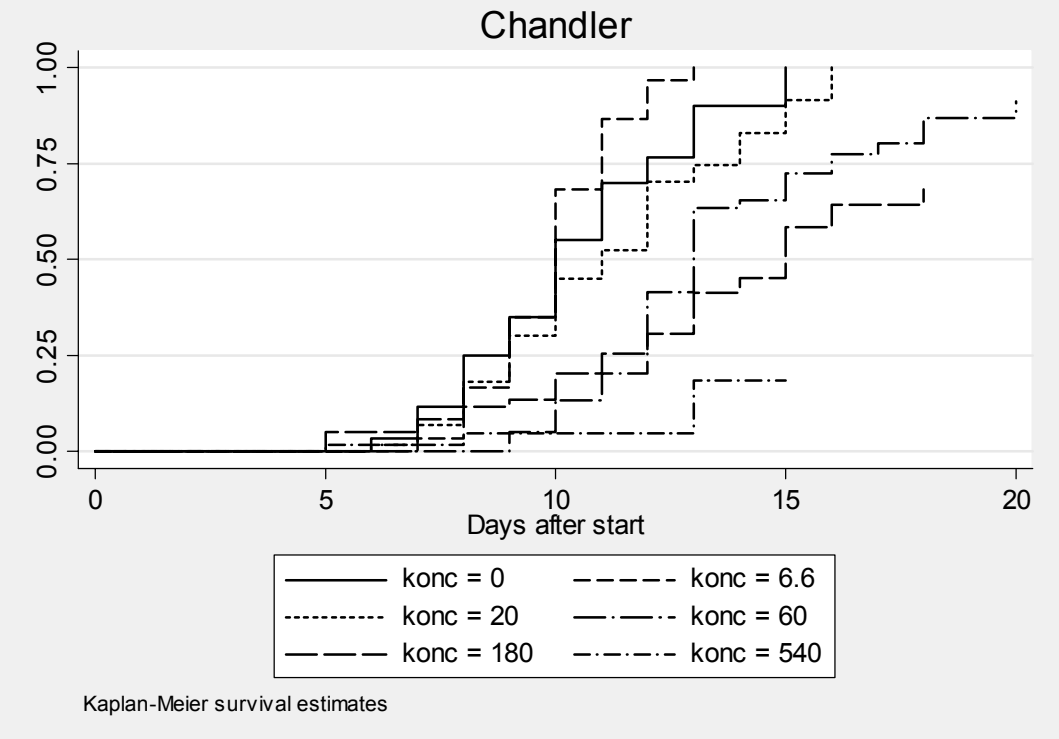


Figure 8. Development from nauplii to copepodite, ITM.

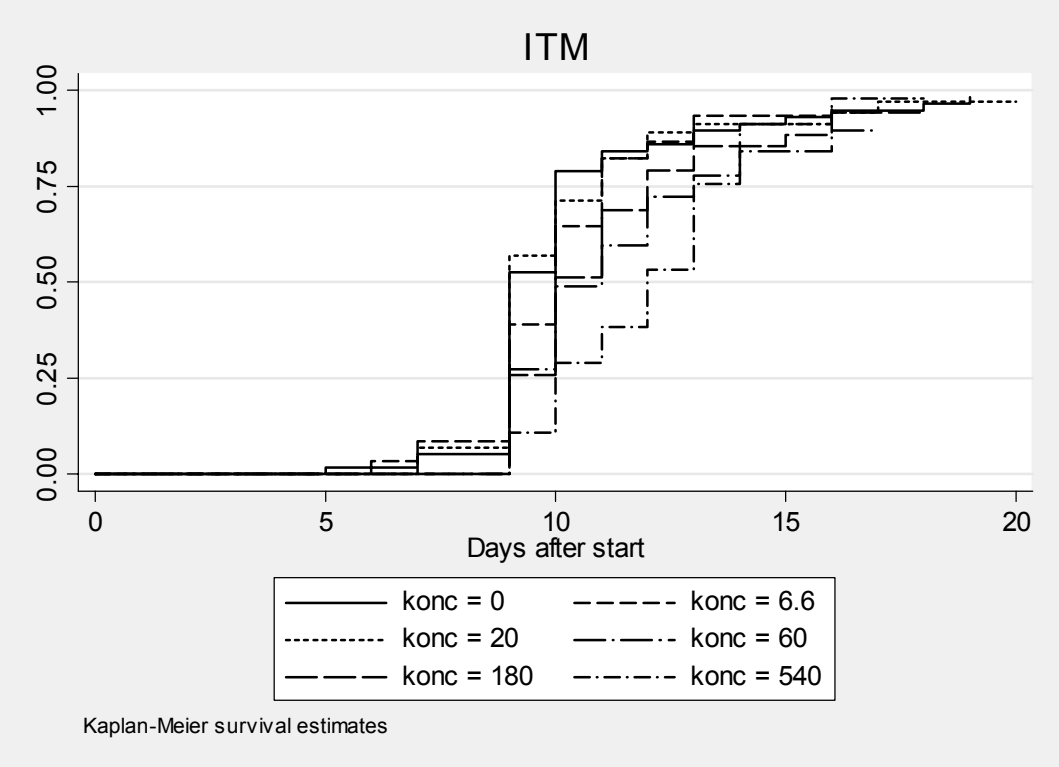


Figure 9. Development from nauplii to copepodite, IVL

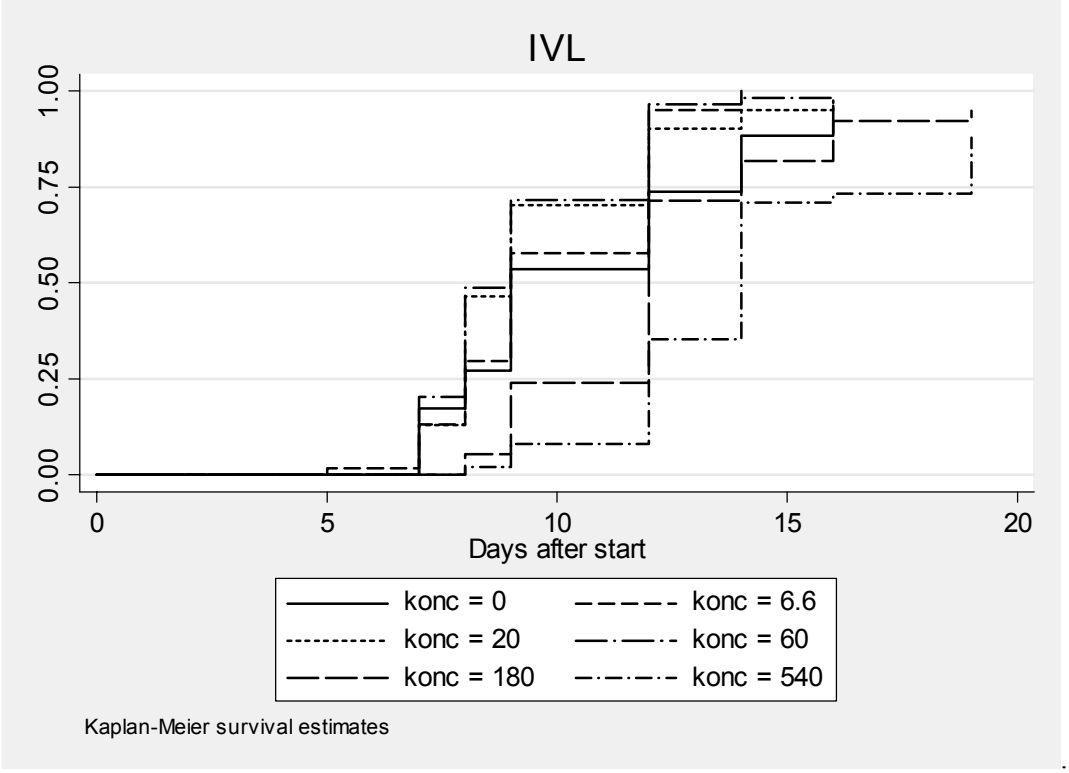


Figure 10. Proportion of mating pairs able to produce two clutches; observed and predicted values.

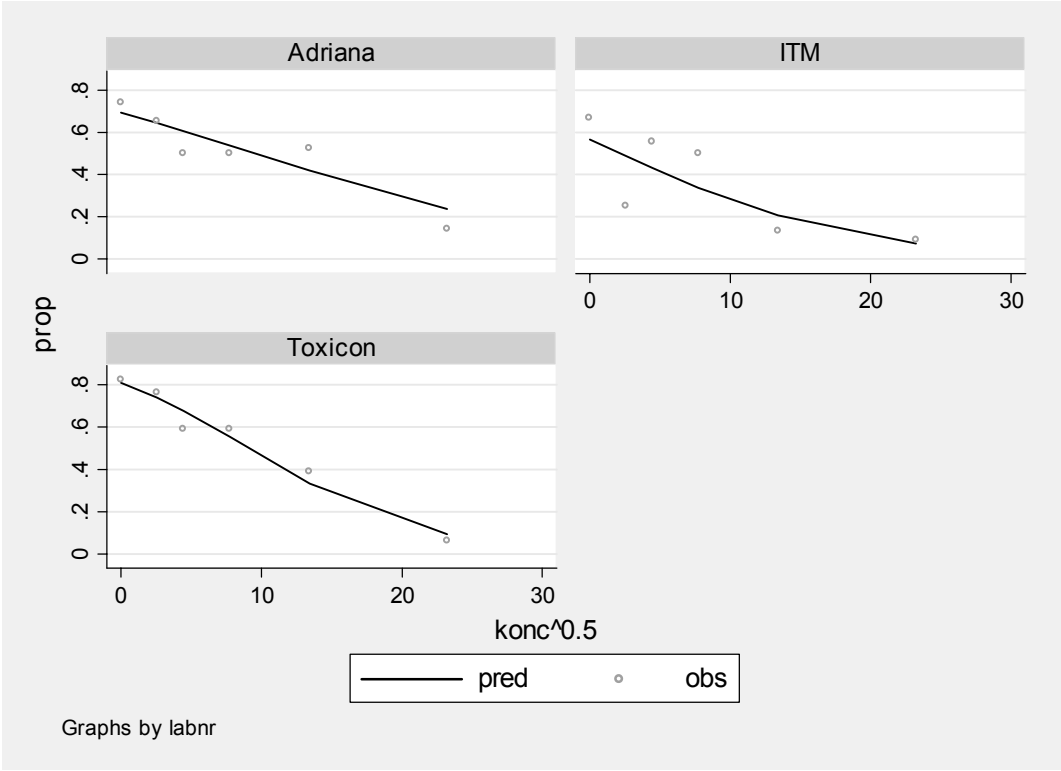


Figure 11.Total viable offspring production for pairs able to produce two clutches.

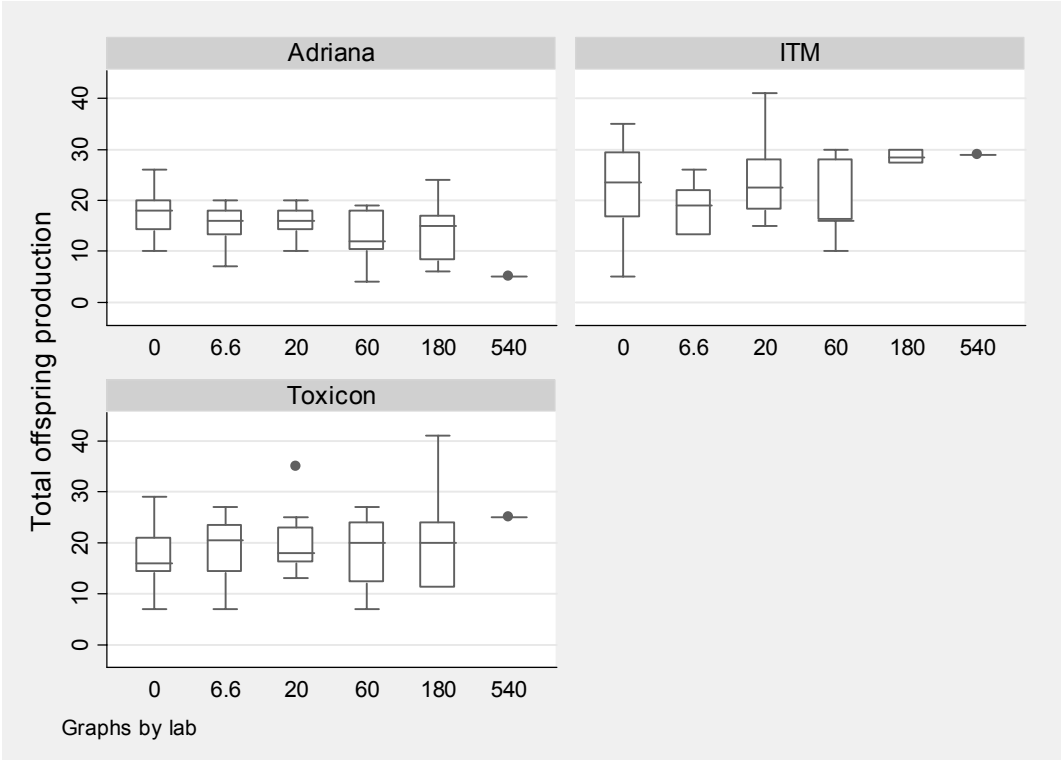


Figure 12. Time to production of first clutch.

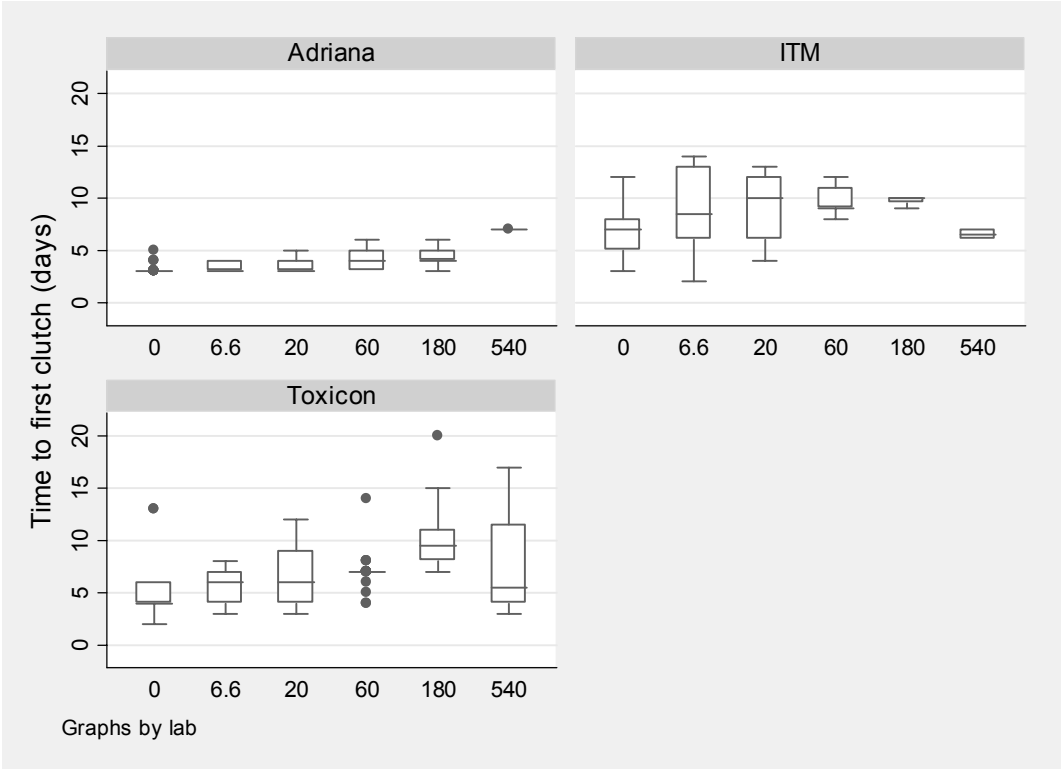
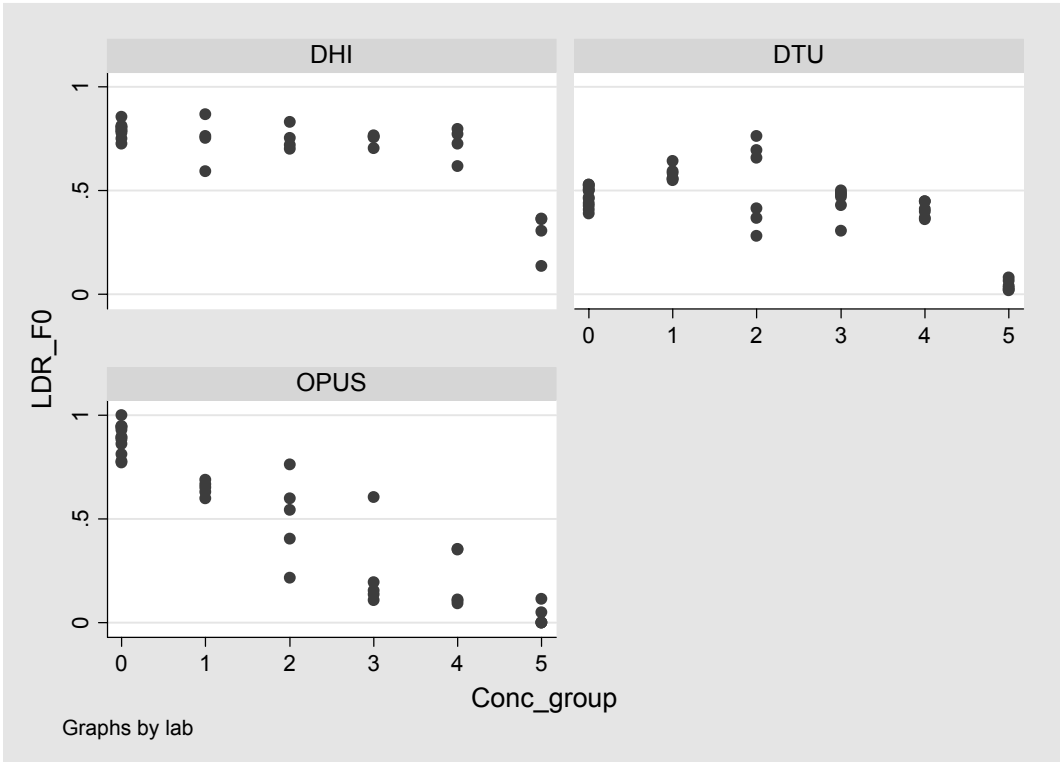


Figure 13. LDR of F_0 against increasing concentration (Control=0).



APPENDIX C (THIS VERSION OF THE *INVITATION LETTER* DOES NOT INCLUDE ANNEXES)

Invitation for participation in

Ring test of the draft OECD[13] Test Guideline

"Harpacticoid and Calanoid Copepod

Development and Reproduction Test"

ABOUT THE OECD TEST GUIDELINES PROGRAMME (TGP)

Test guidelines, including those produced by the OECD, are used by governments, industry and independent laboratories to determine the hazard or safety of chemicals. The use of test guidelines that are based on validated test methods promotes the generation of dependable data for human and animal health and environmental hazard assessment (OECD, 2005)[14].

In recognition of the advantages of internationally agreed test methods, OECD Member countries have developed the OECD Guidelines for the Testing of Chemicals in order to:

- enhance the validity and international acceptability of test data;
- make the best use of available resources in both governments and industry;
- avoid the unnecessary use of laboratory animals;
- minimise non-tariff trade barriers.

[13] OECD – Organisation for Economic Co-operation and Development

[14] OECD Series on Testing and Assessment No 34: Guidance Document on the Validation and International Acceptance of New and Updated Test Methods for Hazard Classification, 2005

In addition, OECD countries have adopted the Decision of the Council concerning the Mutual Acceptance of Data in the assessment of chemicals. This Decision, which is binding on Member countries, requires that data generated in an OECD Member country in accordance with OECD Test Guidelines and Principles of Good Laboratory Practice (GLP) will be accepted in other OECD countries for purposes of health, safety and environmental assessment (OECD web site).

Test method validation is a process based on scientifically sound principles by which the reliability and relevance of a particular test, approach, method, or process are established for a specific purpose. Reliability is defined as the extent of reproducibility of results from a test within and among laboratories over time, when performed using the same standardized protocol. The relevance of a test method describes the relationship between the test and the effect in the target species and whether the test method is meaningful and useful for a defined purpose, with the limitations identified. In brief, it is the extent to which the test method correctly measures or predicts the (biological) effect of interest, as appropriate. Regulatory need, usefulness and limitations of the test method are aspects of its relevance. New and updated test methods need to be both reliable and relevant, *i.e.*, validated (OECD, 2005)[2].

RING TEST OF THE DRAFT OECD TEST GUIDELINE "HARPACTICOID AND CALANOID COPEPOD DEVELOPMENT AND REPRODUCTION TEST"

Background

At the moment there are no internationally harmonised (i.e. OECD) chronic test methods for marine invertebrates, although they represent an ecologically important and large group of organisms that need to be protected to secure transfer of energy from the primary producers to higher levels in the ecosystems, because this group of animals serves as food for fish and marine mammals. Copepods, subdivided into seven groups, of which the calanoids and harpacticoids are two, has the largest number of species of any of the lower crustacean groups. They are enormously abundant as individuals and constitute perhaps the most important group in the plankton of the world's oceans. They are also perhaps the ecologically most important of the crustaceans in that they are herbivores, grazing on the phytoplankton and thus forming the basis of most food chains in the sea. Therefore, the development of a test guideline on development and reproduction of marine copepods has been conducted under the work plan of the **Error! Hyperlink reference not valid.**. The copepod reproduction test is needed in several regulatory systems and will be of wide applicability and as such will facilitate international harmonisation of risk and hazard assessments. It addresses important endpoints and environmental compartments (marine/estuarine/brackish) that are poorly covered at present. Four common and regularly used species in toxicity tests are presented as model test organisms, namely the harpacticoid copepods *Nitocra spinipes, Amphiascus tenuiremis* and *Tisbe battagliai* as well as the calanoid copepod *Acartia tonsa.*

The work was initially started by Swedish members at the Department of Applied Environmental Science **Error! Hyperlink reference not valid.**, Stockholm University. In 2002 a first unofficial draft including both harpacticoid and calanoid copepods (which were written by the Swedish members together with colleagues from England and Denmark) were sent to the members of the OECD expert group on invertebrates. Since then, the draft guideline has been divided into two separate documents, which have been circulated to the OECD expert group on invertebrates two times. In September 2005 the two drafts have been revised according to the results and experiences of the pre-validation round. These revised drafts are found in *Annex A* (harpacticoid draft) and *Annex B* (calanoid draft), respectively.

Sweden (i.e. ITM together with the Swedish Chemicals Inspectorate) will lead the validation of the copepod tests, which includes sending out reference chemicals, preparation of the protocols to be used in the ring tests, data treatment and writing of a final validation report (see *Table 1* below). The two test protocols are currently separate to reflect the differences in biology and thus in the testing conditions between calanoid species and harpacticoid species. However, protocols are intended to be merged into one OECD Test Guideline in the future, when the validation has demonstrated the reliability of the protocols, including common parts valid for both groups and two separate parts reflecting the difference in experimental set up between the calanoid and harpacticoid species (see *Test design*).

Test design

The proposed harpacticoid and calanoid tests focus on identifying effects on *development* and *reproduction*, with emphasis on potential endocrine disruptive effects. A brief overview of the respective test is given below. Harpacticoid and calanoid copepods are closely related phylogenetically, but they differ ecologically. Harpacticoids are almost always benthic or epibenthic and have strong thigmotaxes that make them want to be in contact with structure (e.g., sediments or the walls of microwell test chambers). Calanoids on the other hand are much stronger swimmers, usually negatively thigmotactic and seek suspension in larger volumes of test media. Owing to these differences the experimental setup of the two tests also differs.

Harpacticoida

The test is a full life cycle test, where newly hatched larvae (termed nauplia or metanauplia), aged less than 24 hours at the start of the test, are exposed individually in microwell (300-5000 µL total volume) test chambers to the test substance added to seawater at a range of concentrations. Mandatory test variables to investigate in the F_0 generation are *mean development rates* to the copepodite and adult stages, respectively, *fertilization success* (i.e. number of mating pairs able to produce two successive clutches), *total viable offspring production per mating pair* (i.e. through two clutches averaged over the number of mating pairs created and tested), *time to production of first clutch*, *time interval between successive clutches*, *sex ratio* and *survival* of the parent animals. The test duration of the full life cycle test will vary according to life cycle rate of the test species, but it should extend to hatching of $\geq$ 80% of the second clutches in control pairs at 20ºC (termed the F_1 generation). When the copepods reach adulthood, which normally takes 10-20 days for the present species, the sex of each copepod is determined by light microscopy. Individual male:female mating pairs are constructed and allowed to mate for 7-14 days in new isolated micro well test chambers containing seawater and the test substance added to seawater at the same range of concentrations as during maturation/development. The mating pairs are followed until the control females have been fertilized and have released at least two clutches of offspring. The reproductive output of the animals exposed to the test substance is compared to that of the control(s) over a fixed time period in order to determine the lowest observed effect concentration (LOEC) and hence the no observed effect concentration (NOEC). In addition, and as far as possible, the data are analyzed using a regression model in order to estimate the concentration that would cause an x% reduction in reproductive output, *i.e.* EC_x (*e.g.* EC_{10}).

Calanoida

The test is a full life-cycle test, where the organisms are exposed to various concentrations of the test substance from the egg stage to the juvenile stages of the next generation (F_1 generation). Mandatory test variables to investigate in the F_0 generation are *mean development of early life stages* (i.e. larval development ratio also called LDR*)*, *survival of early stages and egg producing females*, *egg production* and *length of animals*. The next generation (F_1) is exposed during its early life stages and *mean larval development ratio* and *survival* are investigated. The naupliar (larval) and copepodite (juvenile) stages are

morphologically distinct, therefore, the transition from the last naupliar to the first copepodite stage is easily observed. Larval development ratio is normally recorded after 5-7 days, when about 50% of the control animals have reached a copepodite stage, and is expressed as the ratio of copepodites to the total number of surviving early life stages (nauplii + copepodites) at the end of the LDR test. Hatching success of eggs and mortality of the early stages are presented along with the LDR. Egg production of females, which have been exposed through their whole life from egg to adult, is also studied. At the proposed test condition this takes about 12-14 days. Such eggs may bear a chemical burden transferred from the exposed females to the eggs and this can in some cases result in a more sensitive response of the next (F_1) generation LDR test (relative to the F_0 test), because the animals are also exposed in their embryonic stage before hatching. The observed variable of the animals exposed to the test substance is compared to that of the control(s). As far as possible, the data are analyzed using a regression model in order to estimate the concentration that would cause an x% reduction in reproductive output, *i.e.* EC_x (*e.g.* EC_{10}). Also the lowest observed effect concentration (LOEC) and hence the no observed effect concentration (NOEC) may be found using ANOVA statistics.

A short compilation of the harpacticoid and calanoid full life cycle test design (test duration, test solution renewal, feeding, etc.) is presented in *Annex C*.

Reference chemicals

At the OECD invertebrate expert meeting in Paris 2003 it was decided that two reference chemicals should be used in order to cover both general toxicity and specific endocrine activities. *3,5-dichlorophenol (DCP)* was chosen as a reference substance because of its non-specific toxic mode of action, since it is a regularly used reference substance in biological tests and is relatively easy to handle and analyze chemically. This substance has also been used in a pre-validation round in 2004-2005. As a reference substance with a specific mode of action we will use *bisphenol-A*, which has been shown to delay development of young stages in calanoids, harpacticoids and cladocerans. A disrupting mode of action has been proposed for all three groups of organisms; e.g. in cladocerans bisphenol-A was supposed to exert its activity via an enhancement of the activity of endogenous juvenile hormone analogs that decrease larval and juvenile development.

Both chemicals are relatively stable in water over short time periods. DCP will partly dissociate at the test pH, because of its pK_a of 8.2. Bisphenol A has a pKa of 10.8 and is not supposed to dissociate at the test pH. Further information about the reference substances is presented in *Annex D*.

Both reference substances will be available free of charge from ITM to ensure that all participants are using the same batch of test chemicals.

Chemical analysis

The project has received about 10.000 euro to fund parts of the chemical analysis. Efforts will be made to analyse the respective test substances from all participating laboratories at the same laboratory in order to avoid what may become a confounding factor of unreasonable lab-to-lab analytical variability, and to minimize the costs for test analyses.

Laboratories involved in the development of the copepod guideline

The laboratories that have been participating in the development and pre-validation of the copepod guideline are listed below (main species in brackets). Cultures of both copepods and microalge will be available from these laboratories on request. It is desirable that the participating laboratories have former experience from working with copepods in controlled laboratory experiments. Information on how to get into contact with the responsible scientists at each lab is presented at the end of this invitation.

Harpacticoid copepods:

- Department of Applied Environmental Science (ITM), Sweden (*Nitocra spinipes*)
- Department of Environmental Health Sciences, University of South Carolina, USA (*Amphiascus tenuiremis*)
- AstraZeneca, Brixham, UK (*Tisbe battagliai*)

Calanoid copepod

- Environment & Resources DTU, Technical University of Denmark (*Acartia tonsa*)
- DHI Water & Environment, Hørsholm, Denmark (*Acartia tonsa*)

Planned exercise for 2005-2006

Each participating laboratory should use the existing drafts (annex A and B, respectively) to run full life-cycle experiments on both 3,5-DCP and bisphenol-A. However, a more thorough experimental validation plan will be sent out as soon as a laboratory has shown interest in participating in the validation.

The experiments should at the most take 2 months each, including pre-treatment, range finding and after-treatment. This means that each laboratory should be able to finalise tests on both reference chemicals and report effect values and other findings to ITM in April 2006. This allows ITM sufficient time to compile the results and write a validation report, which should be finished and sent to OECD no later than 31st of June 2006.

Table 1. Planned validation activities for 2005 and 2006.

	2005				2006					
Activity	Sept	Oct	Nov	Dec	Jan	Febr	March	April	May	June
Send out invitation	x									
Answer on invitation		x								
Delivery of cultures		x	x							
Order/deliver ref subst.			x							
Optimize cultures			x	x						
3,5-DCP test					x	x				
Bisphenol-A test							x	x		
Send data to ITM								x		
Data treatment ITM								x	x	
Write report									x	x
Report to OECD										x

Laboratories interested in participating in the validation should contact Dr Magnus Breitholtz at ITM no later than 15th of October 2005, preferably via e-mail. Please, indicate whether the laboratory has earlier experience working with copepods life cycle tests. Also, to be able to direct further inquiries about cultures of the different species, the following information should be included in the letter of interest:

- Which species the laboratory is interested in working with;
- Need for cultures of both copepods as well as micro algae for feed.

Yours sincerely,

Dr. Magnus Breitholtz (*Nitocra spinipes*)
Stockholm University
Department of Applied Environmental Science (ITM)
SE-106 91 Stockholm
Sweden
Tel. +46-8-674 7241
Fax +46-8-674 7636
e-mail: magnus.breitholtz@itm.su.se

Professor G. Thomas Chandler (*Amphiascus tenuiremis*)
Dept. of Environmental Health Sciences
Arnold School of Public Health
University of South Carolina
Columbia, SC 29208
e-mail: tchandler@sc.edu

Dr Tim D. Williams (*Tisbe battagliai*)
AstraZeneca
Brixham Environmental Laboratory
Freshwater Quarry
Brixham
Devon
TQ5 8BA
Tel : +44 (0) 1803 882882
Fax : +44 (0) 1803 882974

Professor Kresten Ole Kusk (*Acartia tonsa*)
Technical University of Denmark
Institute of Environment & Resources DTU
Building 115, DK-2800 Lyngby
Denmark
Tel +45-4525 1569
Fax +45-4593 2850
e-mail: *kok@er.dtu.dk*

Estelle Bjørnestad *(Acartia tonsa)*
DHI Water & Environment
Hørsholm
Denmark
esb@dhi.dk

APPENDIX D

Harpacticoid Copepod Life-cycle

Ring-test Protocol

1. Background

A draft test guideline for a harpacticoid (three species) life cycle protocol has been developed. Pre-validation has been conducted and reported and the draft has been revised accordingly.

OECD has sent out an invitation letter to potential participants in the coming ring test. So far, labs from three countries are involved in the ring test of the harpacticoid copepod Test Guideline. However, owing to a low number of participating laboratories for *Tisbe battagliai*, the ring test will only be conducted with *Nitocra spinipes* and *Amphiascus tenuiremis*.

The actual ring-test will start in early 2006. However, it was decided at 2nd OECD Invertebrate Expert Group meeting in Paris to only use 3,5-DCP as a reference chemical, since it was concluded that the aim of the copepod test is not to detect a specific mode of action, but to evaluate test procedures and adverse consequences of chemicals on development and reproduction. This means that the ring-test follows the scheduled outlined in the invitation letter (see Table 1).

For more thorough background information, we refer to the *'Invitation Letter'* that was distributed by OECD in September 2005 to national coordinators and interested laboratories.

2. Ring test activities

2.1. Laboratories participating in the ring test

The laboratories that are presently involved in the ring test are presented below.

Nitocra spinipes

- Department of Applied Environmental Science (ITM), Stockholm University, SE-106 91 Stockholm, Sweden. Contact: magnus.breitholtz@itm.su.se.
- IVL Swedish Environmental Research Institute, Box 210 60, SE-100 31 Stockholm, Sweden. Contact: ann-sofie.allard@ivl.se.
- Toxicon AB, SE-261 92 Landskrona, Sweden. Contact: anders.sjolin@toxicon.com.

Amphiascus tenuiremis

- Department of Environmental Health Sciences, Arnold School of public Health, University of South Carolina, SC-29208, USA. Contact: Chandlgt@gwm.sc.edu
- Block Environmental Inc., Pleasant Hill, California, USA. Contact: dblock@blockenviron.com
- The NOAA Center for Coastal Environmental Health and Biomolecular Research (CCEHBR), Charleston, South Carolina, USA. Contact: geoff.scott@noaa.gov

2.2. Schedule

The general schedule for the ring test of the harpacticoid copepod guideline is outlined in Table 1 below.

Each participating laboratory should use the existing draft guideline to run full life-cycle experiments on the reference chemical 3,5-DCP.

The experiments should at the most take 2 months, including pre-treatment, range finding and after-treatment. This means that each laboratory should be able to finalize the test and report raw data and other findings to ITM in late April 2006. This allows ITM sufficient time to compile the results and write a validation report, which should be finished and sent to OECD no later than 31st of June 2006.

Table 1. Ring test activities for 2005 and 2006.

2.3.

	2005		2006					
Activity	Nov	Dec	Jan	Febr	March	April	May	June
-Delivery of cultures	x	x						
Order/deliver ref subst.		x						
Optimize cultures		x	x	x				
3,5-DCP test					x	x	x	
Send data to ITM						x	x	
Send test medium to ITM						x	x	
Chemical analysis ITM							x	x
Data treatment ITM							x	x
Write report								x
Report to OECD								x

Principle of the test

The test is a full life cycle test, where newly hatched larvae (termed nauplia or metanauplia), aged less than 24 hours at the start of the test, are exposed individually in microwell (300 µL total volume) test chambers to the test substance added to seawater at a range of concentrations. Mandatory test variables to investigate in the F_0 generation are *mean development rates* to the copepodite and adult stages, respectively, *fertilization success* (*i.e.* number of mating pairs able to produce two successive clutches), *total viable offspring production per mating pair* (*i.e.* through two clutches averaged over the number of mating pairs created and tested), *time to production of first clutch*, *time interval between successive clutches*, *sex ratio* and *survival* of the parent animals. The test duration of the full life cycle test will vary

according to life cycle rate of the test species, but it should extend to hatching of ≥ 80% of the second clutches in control pairs (termed the F_1 generation). When the copepods reach adulthood, which normally takes 10-20 days for the present species, the sex of each copepod is determined by light microscopy. Individual male:female mating pairs are constructed and allowed to mate for 7-14 days in new isolated micro well test chambers containing seawater and the test substance added to seawater at the same range of concentrations as during maturation/development. The mating pairs are followed until the control females have been fertilized and have released at least two clutches of offspring. The reproductive output of the animals exposed to the test substance is compared to that of the control(s) over a fixed time period in order to determine the lowest observed effect concentration (LOEC) and hence the no observed effect concentration (NOEC). In addition, and as far as possible, the data are analyzed using a regression model in order to estimate the concentration that would cause an x% reduction in reproductive output, *i.e.* EC_x (*e.g.* EC_{10}).

2.4. Reference chemical

3,5-dichlorophenol (3,5-DCP) has been chosen as a reference chemical because of its non-specific toxic mode of action, since it is a regularly used reference substance in biological tests and is relatively easy to handle and analyze chemically. This substance has also been used in a pre-validation round in 2004-2005. 3,5-DCP is relatively stable in water over short time periods but may partly dissociate at the test pH, because of its pK_a of 8.2.

The reference substance will be available free of charge from ITM to ensure that all participants are using the same batch of test chemical.

2.5. Chemical analysis

ITM has received about 10.000 euro to fund parts of the chemical analysis, which will be made by ITM on all participating laboratories test medium samples in order to avoid what may become a confounding factor of unreasonable lab-to-lab analytical variability, and to minimize the costs for test analyses.

Test medium should be sampled and pooled across microplates within treatments from each well on each test medium renewal occasion (≤ 15 ml). Fresh test solution from the full concentration range should also be sampled each time new stock solution are prepared including start of the test (if a stock solution of 200 mL is prepared this volume should be sufficient throughout the test). All test medium samples should be stored at -20°C until chemical analyses.

The general procedure for the chemical analyses is based on a dervatization method for phenolic substances. In brief, after addition of NaOH and fluorine-reagents directly to the test medium water, it is possible to extract the fluorine esters that are formed to hexane and then use GC/ECD and MS/NICI for quantification.

Owing to a high cost (both analyses and shipping) as well as difficult handling procedures, the chemical analyses will only be made on specific treatments, *i.e.* in the lower end of the dose-range (*e.g.* without an effect) and in the high end (*e.g.* were significant mortality has occurred).Test medium samples will at least be analyzed at the beginning, in the middle and in the end of the exposure. This means that participating laboratories should not send all test medium samples to ITM but instead those that will be identified after all laboratories have reported their raw data.

Old test medium should therefore be sampled and pooled (replicates within each treatment) for the following concentrations: 6.6, 60 and 540 µg/L.

2.6. Procedure

General

2.6.1 Cultures

The culture procedures for *A. tenuiremis* and *N. spinipes* are described in Annex 2 and 3 of the harpacticoid draft guideline.

2.6.2 Conditions of exposure

Test vessels – microplates

Microwell test chambers and other test vessels and apparatus, which will come into contact with the test solutions, should, if possible, be made entirely of glass or other materials chemically inert to the test chemical. Microwell plates are typically made of polystyrene plastic. Polystyrene is not well suited to testing with hydrophobic organic chemicals because of non-specific plastic:organic binding, which can reduce chemical bioavailability to copepods. Therefore, microplates with microwells pre-coated with glass or polyacrylamide gel are better suited for copepod testing of hydrophobic chemicals than microplates with naked polystyrene wells. For *N. spinipes* it is recommended to use glass coated microwell chambers or to silanize the walls (during exposure of individuals) because of problematic surface tension.

- For *A. tenuiremis,* 96 well microplates (each well containing 250 µL seawater test solution) are used during the development phase. After sexing and creation of sexed pairs, adult pairs are exposed in new, unused microwells of each 96-well microplate.

- For *N. spinipes*, 96 well microplates (each containing 270 µL) are used during the development phase. After sexing and creation of sexed pairs, adults are exposed in 24 well microplates (each containing 2 mL).

Duration

The test duration of the full life cycle test will vary according to life cycle rate of the test species, but it should extend to hatching of $\geq$ 80% of the second clutches in control pairs (termed the F_1 generation).

Animal loading

Newly hatched larvae (termed nauplia or metanauplia), aged less than 24 hours at the start of the test, are exposed individually (> 60 replicate nauplii per treatment) in microwell (300 µL total volume; see above for details) test chambers to seawater (i.e. control) and to the test substance added to seawater at a range of concentrations. Treatments should be dispersed randomly among separate replicate (n=3) microplates, as should all subsequent handling, loading, observation and feeding. Failure to do this may result in bias that could be construed as being a chemical-dependent effect. In particular, if experimental units are handled in treatment or concentration order, then some time-related effect, such as operator fatigue or other error, could lead to greater or lesser effects at the higher or lower concentrations. Only one test chemical concentration should be tested per microplate to reduce chances for microwell cross contamination. Twenty or more nauplii should be tested per microplate. Where possible, nauplii should be placed within the microwell array away from any end or edge microwells. Microwells located on ends and edges of the microplate are most susceptible to temperature and evaporation. Furthermore, if the test results are likely to be affected by an initial or environmental condition of the test, such as position in the

laboratory or environmental chamber, then consideration should be given to blocking the test, or using a Latin-square experimental design.

Feeding

Species-specific information on feeding of individual copepods in test chambers during maturation/development and reproduction is presented below:

- Amphiascus tenuiremis

Each copepod is fed 2µl of a 1:1:1 volume mixed-algal cell suspension (10^7 cells/mL) of a chlorophyte (*e.g.*, *Dunaliella tertiolecta*), a chrysophyte (e.g., *Isochrysis galbana*) and a cryptophyte (e.g., *Rhodomonas salina*) every six days during both maturation/development and reproduction. The concentration added to each test chamber at each feeding is ~ 2 x 10^4 cells/microwell; i.e., ~2 x 10^5 cells/mL test solution in a 250 uL microwell.

- Nitocra spinipes

Each copepod is fed a cell suspension (stock suspension 5 x 10^7 cells/mL) of a cryptophyte, preferably *Rhodomonas salina* or *Rhodomonas baltica*, every third day during both maturation/development and reproduction. Copepods are fed on test solution renewal occasions hence algae is added to the bulk volume of prepared test solution prior to dispensing into test chambers. The concentration added to each test chamber at each feeding is ~ 5 x 10^5 cells/microwell; e.g., ~ 2.5 x 10^5 cells/mL test solution in a 2 mL microwell

Light and temperature

During testing, microplates are covered and placed in a clear/opaque humidity chamber (a covered plastic tub holding several sponges soaked in distilled water works well for this) in a temperature-regulated incubator, or a room set at 22°C. Amphiascus requires a 12-h light and 12-h dark test photoperiod. *N. spinipes* should be reared/tested in darkness throughout the test duration.

Aeration

The test vessels must not be aerated during the test, but dilution seawater should be aerated for 24 h prior to filtration at 0.45 µm and subsequent mixing with test chemical stocks.

Test concentrations and test medium renewal

The following concentration range should be used for the testing of 3,5-DCP:

6.6, 20, 60, 180 and 540 µg l^{-1} 3,5-DCP

Since 3,5-DCP is soluble in water at the given range, a carrier/solvent is not needed to prepare the stock solution(s) in water.

Natural or synthetic seawater may be used as the test medium. If natural seawater is used, it should be collected from a location distant from any known sources of pollution and filtered to remove indigenous organisms. If synthetic seawater is used, it should be prepared by dissolving reagents of recognised analytical grade, or a commercially available formulation, in distilled or deionised water. However, for these copepod species, there is insufficient information on the use of synthetic seawater to

allow a particular brand or type to be recommended. The suitability of synthetic seawater for culture and testing should, therefore, be evaluated in trials. Any water in which the copepods show suitable long-term survival with normal development and fecundity may be used as the test medium.

The salinity of the test medium should be the same as the culture medium for the respective species. Each species may be used at a range of salinities but in this ring test the following salinities should be used:

- ~6.5‰ for *N. spinipes*.
- 30 ‰ for *A. tenuiremis*.

Whichever salinity is employed, the test organisms must be cultured or maintained at the same salinity (±3‰) for at least 7 days before the start of the test. The dilution water must have a dissolved oxygen concentration above 90% of the air saturation value, and a pH of 8.0 ± 0.3 before being used to prepare the test solutions. It is recommended that all dilution water be filtered at 0.45 μm prior to 3,5 DCP test solution preparation to minimize bacterial contamination.

The dilution water used for testing should be from the same source as water that has been found to support culture of the organisms for at least two generations. If, however, the medium to be used in the test is different from that used for routine culture, it is a good practice to include an adequate pre-test acclimation period (*i.e.* sufficient time to culture the copepods through at least two generations) to avoid stressing the animals.

Potential evaporation losses of media from the test chambers should be observed daily and, if needed, be corrected by the addition of deionised or distilled water. When using microwell test chambers (300-2000 μL total volume), which are normally held in temperature-regulated incubators or rooms, care should be taken so that the evaporation losses do not exceed 7% - on a daily basis. To prevent such a scenario, microplates may be placed over a water bath in a humidity chamber in the incubator or room.

When the medium is renewed, a second series of microwell test chambers or other appropriate vessels are prepared and the animals transferred to them by, for example, a glass Pasteur pipette of suitable diameter. The volume of medium transferred with the copepods should be minimised. Alternatively, a fixed volume (preferable 70-90% of the total test medium volume) of the medium may be removed from each chamber using either automatic pipette or a syringe of suitable diameter. Both procedures should be performed under dissection stereomicroscopic observation, and care should be taken to use a pipette or a syringe of suitable diameter so as not to damage the copepods when they are aspirated into the pipette or syringe. If, after medium renewal, any copepods are observed to have been damaged during this procedure, the appropriate test chamber should be disqualified from the test.

Observations

As summarized in Table 2 below, observations include *sex ratio*, *larval development ratio (LDR)* of nauplius versus first copepodite stage on day 5-7 of the bioassay, *mean development rates of nauplius to copepodite, copepodite to adult, and nauplius to adult stages*, respectively, *fertilization success* (i.e. number of mating pairs able to produce two successive clutches), *total viable offspring production per mating pair* (i.e. through two clutches averaged over the number of mating pairs created and tested), *time to production of first and second clutches*, and *survival* of the F_O parent animals. Observations should also be made for any abnormal behaviour (relative to the controls), including uncoordinated swimming, loss of equilibrium, change in pigmentation and atypical quiescence. Sex ratio is determined by observations of

male and female secondary sexual characteristics (see draft guideline annexes), which in copepods normally can be distinguished in the 5th copepodite stage at the earliest.

- Development: The number of days it takes for development of each individually isolated copepod, from a first stage nauplius to a juvenile copepodite to a reproductively mature adult, is checked daily. For some harpacticoid species, morphological identification of the different stage durations is extremely difficult. In this case, precise determination of time to reach individual stage durations is best attempted by recording individual moults to adult (e.g. there are 5 copepodite stages [moults] to the adult). For all species within a given treatment or control microplate, once adulthood is achieved, all males should be pooled briefly in a 60-mm Petri dish, and all females within a separate 60-mm Petri dish. Random within-treatment individual pairs are then made by selecting one male and one female adult from each pooled collection of same treatment individuals. As many pairs as can be made per treatment (< 30 maximum possible pairs) are placed one each into new unused microwells loaded with the appropriate 3,5 DCP test concentration.

- Offspring production: After pairing of male:female mating pairs, observations are made daily for the presence of females carrying their first and second egg sacs, and offspring are counted as they hatch. Live offspring (nauplii) will usually be released from the egg sac after an incubation period of two to several days. If necessary, in order to count the number of offspring released from second broods, pairs may be placed in new unused microwells until hatch of the second brood. The development of egg sacs, the presence of abnormal eggs or aborted unhatching egg sacs, the release of offspring and the number of live and dead offspring are recorded at each observation point during the approximately 25 day test. Renewal of test medium should be continued as described above through hatching of brood two.

At the end of the exposure period, all remaining copepods (*i.e.* including unfertilised females and males) are either discarded or fixed in formaldehyde (4%) for determination of sex.

- Mortality: Survival of larvae (nauplii), juveniles (copepodites) and adults is recorded daily during maturation/development/mating up to and including the last day of the test. Mortality of mating pairs also should be checked and recorded daily.

Table 2. Summary of endpoints to be reported from the life cycle test.

Detail	Endpoints
Test variables to be reported	
Development	- Mean development rates for full life-cycle and for each life-stage window
	- Fertilization success
	- Total viable offspring production
	- Time to production of first clutch
Reproduction	- Time interval between successive clutches.
	- Sex ratio of F_0
	- Survival rate of F_0
Other	

Regular observations of F_0-generation nauplii and copepodites as well as counting of F_1-generation nauplii must be facilitated by the use of light microscopy. Upon termination of the test, staining in Lugol's solution may facilitate counting of released nauplii and aborted egg sacs before using light microscopy.

The results of the observations made during the test should be recorded on the data sheets provided below.

Physical-chemical parameters

During the test, dissolved oxygen, pH and temperature should be measured in the controls and all test concentrations each time test medium is renewed. As a minimum, these measurements shall be made in the control(s) and in the highest test concentration. Temperature should preferably be monitored continuously in at least one test vessel.

2.6.3. Reporting

Raw data should be compiled in the attached raw data sheets and sent to ITM as soon as the testing has been finalized for each participating laboratory.

APPENDIX E

Detailed Calanoid Copepod Life-cycle

Ring-test Protocol

1. Background

A draft test guideline for a calanoid life cycle protocol has been developed. Pre-validation has been conducted and reported and the draft guideline has been revised accordingly.

OECD has sent out an invitation letter to potential participants in the coming ring test. So far, four labs from three countries are involved in the ring test of the Calanoid Copepod Test Guideline. At the 2nd OECD Invertebrate Expert Group meeting in Paris, 3rd to 4th of November 2005, it was decided within the coming weeks, once again to take a direct contact with potential participants.

The actual ring-test will start in spring 2006. However, it was decided at the meeting in Paris to use 3,5-DCP as the only reference chemical, since it was concluded that the aim of the copepod test is not to detect a specific mode of action, but to evaluate test procedures and adverse consequences of chemicals on development and reproduction. This means that the ring-test follows the scheduled outlined in the invitation letter (see Table 1).

For more thorough background information, we refer to the *'Invitation Letter'* that was distributed by OECD in September 2005 to national coordinators and interested laboratories.

2. Ring test activities

2.1. Laboratories participating in the ring test with Acartia tonsa

The laboratories that are presently involved in the ring test are presented below (main species in brackets).

- Institute of Environment & Resources, Technical University of Denmark (E&R), Building 113, 2800 Lyngby, Denmark (Contact: K. Ole Kusk; kok@er.dtu.dk)

- DHI Water and Environment (DHI), Agern Allé 5, 2970 Hørsholm, Denmark (Contact Estelle Bjørnestad; esb@dhi.dk)

- Opus Plus Ltd, (OPUS), Flotta, Stromness, Orkney, KW16 3NP, United Kingdom (Contact Brenda Hudson; brenda.hudson@opusplus.co.uk)

- ECT Oekotoxikologie GmbH, Boettgerstr. 2-14 D-65439 Floersheim, Germany (Contact: Michael Meller; m-meller@ect.de)

2.2. Schedule

The general schedule for the ring test of the copepod guideline is outlined in Table 1 below.

Each participating laboratory should use the existing draft guideline to perform a full life-cycle experiment with the reference chemical 3,5-DCP.

The experiment should at the most take 2 months, but can be performed in about three-four weeks. This means that each laboratory should be able to finalise the test and report raw data, effect values and other findings to Department of Applied Environmental Science (ITM), Stockholm University, SE-106 91 Stockholm, Sweden (Contact: magnus.breitholtz@itm.su.se) in August 2006 at latest and preferably earlier.

Table 1. Ring test activities for 2006.

	2006							
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Activity								
-Delivery of cultures	x	x	x					
Order/delivery of reference substance				x				
Optimize cultures				x				
3,5-DCP test				x	x	x	x	
Send data to ITM						x	x	x
Send water samples to ITM						x	x	
Chemical analysis ITM							x	x
Data treatment ITM							x	x
Write report								x
Report to OECD								x

2.3. Principle of the test

The test is a full life cycle test, where the organisms are exposed to various concentrations of the test substance from the egg stage (F_0 generation) to the developing stages of the next generation (F_1) including adult egg production (F_0). Compulsory test variables to investigate in the F_0 generation are *hatching success*, *development of early life stages (mean larval development ratio* also called LDR*)*, *survival of early stages and egg producing females*, *egg production* and *length of animals*. The next generation (F_1) is exposed during its early life stages and *hatching success, mean larval development ratio* and *survival* are investigated. Optional/Recommended test variables to investigate are *acute toxicity* for range finding purposes, *sex ratio* in the F_0 generation and *egg diameter*. The total test duration at 20 °C and 20 ‰ salinity is about three to four weeks (5-7 days for F_0 generation LDR test + 17-18 days for F_0 generation life cycle exposure + 5-7 days F_1 generation LDR test), which is sufficient time to investigate the development of the first generation (F_0) - and in the life-cycle test for the control animals to reach adulthood and produce eggs, as well as for the second generation (F_1) to partly develop into copepodites. The nauplii (larval) and copepodite (juvenile) stages are morphologically distinct, therefore, the transition from the last naupliar to the first copepodite stage is easily observed. Larval development ratio is normally recorded after 5-7 days, when about 50% of the control animals have reached a copepodite stage, and is expressed as the ratio of copepodites to the total number of surviving early life stages (nauplii + copepodites) at the end of the LDR test. Hatching success of eggs and mortality of the early stages are presented along with the LDR. Egg production of females, which have been exposed through their whole

life from egg to adult, is also studied. At the proposed test condition it takes about 12-14 days for the animals to reach adulthood. Eggs produced by exposed females may bear a chemical burden, which might result in a more sensitive response of the next (F_1) generation in the LDR test (relative to the F_0 generation), because the animals are also exposed during their embryonic stage.. Observed variables of exposed groups are compared to that of the control(s) in order to determine the lowest observed effect concentration (LOEC) and hence the no observed effect concentration (NOEC) by using ANOVA statistics. As far as possible, the data are also analyzed using a regression model in order to estimate the concentration that would cause an x% reduction in observed variable, *i.e.* EC_x (*e.g.* EC_{10}).

2.4. Reference chemical

3,5-dichlorophenol (3,5-DCP) has been chosen as a reference chemical because of its non-specific toxic mode of action, its regularly use as reference substance in biological tests and because it is relatively easy to handle and analyze chemically. This substance has also been used in a pre-validation round in 2004-2005. 3,5-DCP is relatively stable in water over short time periods but may partly dissociate at the test pH because of its pK_a of 8.2.

The reference substance will be available free of charge from ITM to ensure that all participants are using the same batch.

2.5. Chemical analysis

ITM has received about 10.000 euro to fund parts of the chemical analysis, which will be made by ITM on all participating laboratories test medium samples in order to avoid what may become a confounding factor of unreasonable lab-to-lab analytical variability, and to minimize the costs for test analyses.

Test medium should be sampled and pooled across the replicates of three exposure concentrations on each test medium renewal occasion (10-15 ml in total of the composite sample is enough for analysis). Fresh test solution from the same three concentrations should also be sampled each time new stock solutions are prepared including start of the test. All test medium samples should be stored at -20°C until chemical analyses. Concentrations sampled should be 6.6, 60 and 540 µg/L. Samples from more concentrations may be collected, but may not necessary be analysed at the end (see below).

The general procedure for the chemical analyses is based on a derivatization method for phenolic substances. In brief, after addition of NaOH and fluorine-reagents directly to the test medium water, it is possible to extract the fluorine esters that are formed to hexane and then use GC/ECD and MS/NICI for quantification.

Owing to a high cost (both analyses and shipping) as well as difficult handling procedures, the chemical analyses will only be made on specific treatments, *i.e.* in the lower end of the dose-range (*e.g.* without an effect) and in the high end (*e.g.* were significant mortality has occurred). Test medium samples will at least be analyzed at the beginning, in the middle and in the end of the exposure. This means that participating laboratories should not send all test medium samples to ITM but instead those that will be identified after all laboratories have reported their raw data.

2.6. Procedure

2.6.1 Culture

The culture procedures for *Acartia tonsa* are described in Annex 7 of the calanoid draft guideline.

2.6.2. Conditions of exposure

Test vessels and other equipment

Test vessels, other equipment and apparatus which will come into contact with the test solutions should, if possible, be made entirely of glass or other materials chemically inert to the test chemical.

Duration

The duration of the full life cycle test with Acartia tonsa at the specified conditions will be about 4 weeks including LDR tests on F_0 and F_1.

Animal loading

Larval development ratio (LDR) tests (F_0 and F_1) are started with 50-80 eggs per test vessel. The life cycle test (basic or mass exposure) is started with 500-600 eggs per test vessel. Eggs are counted on a filter before addition. In Annex 3 of the Calanoid guideline is given an exemplary flow diagram of the test procedure.

Feeding

Rhodomonas salina (Chryptophyceae) is used as feed. Feeding should be done daily or just after media changes. In the flow diagram (Annex 3 of the Calanoid guideline) are given the recommended algal densities.

Light and temperature

Test vessels shall be covered and placed in a temperature-regulated incubator or room set at 20 ºC with a 16-h light and 8-h dark photoperiod at low light intensity.

Aeration

In the life cycle test, vessels are aerated from day 7. In the LDR test vessels are not aerated.

Test media and media renewal

Synthetic seawater with a salinity of 20 ‰ is used as medium. Synthetic seawater should be prepared by dissolving reagents of recognised analytical grade or a commercially available formulation in distilled or deionised water. Any synthetic seawater of known composition in which the copepods show suitable long-term survival with normal development and fecundity may be used as the test medium.

The same type of media should be use as culture media, test media and dilution water. The media must have a dissolved oxygen concentration above 70% of the air saturation value, and a pH variation of maximally ± 0.3 from the initial pH. The initial pH is typically 8.1-8.3 in the synthetic media used at DTU (see annex 7 in the Calanoid guide).

Potential evaporation losses of media from the test chambers should be observed regularly and, if needed, be compensated by the addition of deionised or distilled water.

Media is renewed three times a week. - In the calanoid guideline (paragraph 32 and footnote 2 and 3 and in the flow diagram of Annex 3) are given details on media renewal for the different parts of the test.

Test concentrations and test medium renewal

The following concentration range should be used for the testing of 3,5-DCP with *Acartia tonsa*:

6.6, 20, 60, 180 and 540 $\mu g\ l^{-1}$

Since 3,5-DCP is soluble in water at the given range, no carrier/solvent is needed to prepare the stock solution(s).

Media should be renewed as described in the draft guideline. An example of how to perform the test in practice is given in Annex 3 of the draft guideline for calanoid copepods.

Observations

As summarized in Table 2 below, observations include *hatching success, larval survival* and *larval development ratio* in F_0 and F_1 generation, *fecundity* (i.e. number of eggs per female per day), and length of males and females. Observations should also be made for any abnormal behaviour (relative to the controls), including uncoordinated swimming, loss of equilibrium, change in pigmentation and atypical quiescence. Sex ratio may be determined by observations of male and female secondary sexual characteristics (see draft guideline annexes), which in copepods normally can be distinguished in the 5th copepodite stage at the earliest.

- Hatching, survival and development of early stages: The exact number of eggs transferred to each vessel is counted. At the end of the exposure number of unhatched eggs, number of nauplii and number of copepodites are counted. These numbers are used for calculation of hatching success, survival and LDR.

The LDR tests are stopped when the fraction of copepodites in the controls is as close to 50 % as possible. To determine this point of time several additional control vessels shall be set up. On day 5 control vessels are sacrificed with time intervals by adding lugols solution and numbers of nauplii and copepodites are counted. If the copepodite fraction is close to 50 % the test is stopped. If the test is started with a time interval between the separate vessels, the same time interval should be observed when stopping the test.

One validity criteria for the LDR test is a copepodite fraction of 50 ± 20 %. If a 30 % copepodite fraction is not reached by the end of day 5 the test may be continued at a dark place at 15 °C and the examinations may then be resumed the following morning.

Fecundity: At the end of the basic exposure 12 females from each exposure group are placed individually in vessels with an inner chamber and a bottom of stainless steel net (180-200 µm mesh size). Eggs will pass through this net but not the female. The females are transferred to new test vessels every 24 hours. Eggs are not counted after the first 24 hours but the production over the successive three days is counted. (See Annex 3 in the calanoid guide for more details).

- Length of animals: The body length of 25 males and 25 females is measured under a microscope. The median axis from the front of the head to the end of the prosome is measured. The urosome

(or tail) is NOT considered. The urosome is often bended upwards and it is thus difficult to measure the total length. (See Annex 2 of the calanoid guide for more details).

Survival of 50 males and 50 females in the control of the basic exposure: To secure a sufficient number of animals for the egg production test and for measuring lengths this validity criterion has been defined.

Physical-chemical parameters

During the test, dissolved oxygen, pH and temperature should be measured in the controls and all test concentrations each time medium is renewed. As a minimum, these measurements shall be made in the control(s) and in the highest test concentration. Temperature should preferably be monitored continuously in at least one test vessel.

Table 2. Summary of endpoints to be reported from the life cycle test.

Detail	Endpoints
Test variables to be reported	
Development F_0	- Hatching success
	- Larval survival
	- Larval development ratio (LDR) (count number of eggs at start, number of unhatched, number of nauplii and number of copepodites at end)
	Length of 25 males and 25 females
	Total number of eggs per female per day (count number of eggs over three separate days and number of females)
Reproduction	- Hatching success
	- Larval survival
	- Larval development ratio (LDR) (count number of eggs at start, number of unhatched, number of nauplii and number of copepodites at end)
Development F_1	- Sex ratio of F_0
	- Survival rate of F_0
	(count all egg at start, count and determine sex on all surviving animals at end)
	- Egg diameter
Other	

2.6.3 Reporting

Raw data should be compiled in the attached raw data sheets and sent to ITM as soon as the testing has been finalized for each participating laboratory preferably as electronic files or by using the data sheets in Annex 5 of the draft guideline.