EUROPEAN COMMISSION
Community Research

ERICA

(Contract Number: FI6R-CT-2004-508847)

DELIVERABLE 5:
Annex A
Guidelines for the design and statistical analysis of experiments on chronic effects of radioactive substances

Editors: Garnier-Laplace J and Gilbin R (IRSN)

Authors: **A**dam C (IRSN), Garnier-Laplace J (IRSN), Gilbin R (IRSN), Jaworska A (NRPA) and Oughton D (UMB)

Reporting period: March 2004 – February 2006

Date of issue of this report: **28-02-2006**

Start date of project : **01/03/04** Duration : 36 Months

Project co-funded by the European Commission under the Euratom Research and Training Programme on Nuclear Energy within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	PU
RE	Restricted to a group specified by the partners of the [ERICA] project	
CO	Confidential, only for partners of the [ERICA] project	

DISTRIBUTION LIST

Name	Number of copies	Comments
EC, Henning von Maravic	1 2	electronically as pdf hard copy
ERICA Partners	1 2	electronically as pdf hard copy
www.erica-project.org	1	electronically as pdf on the Public area

ERICA will provide an integrated approach to scientific, managerial and societal issues concerned with the environmental effects of contaminants emitting ionising radiation, with emphasis on biota and ecosystems. The project started in March 2004 and is to end by February 2007.

Erica tetralix L.

Contract No: **FI6R-CT-2004-508847**
Project Coordinator: **Swedish Radiation Protection Authority**

Contractors:

Swedish Radiation Protection Authority	SSI
Swedish Nuclear Fuel and Waste Management Company	SKB
Facilia AB	Facilia
Södertörn University College	SUC
Norwegian Radiation Protection Authority	NRPA
Research Centre in Energy, Environment and Technology	CIEMAT
Environment Agency	EA
University of Liverpool	UNILIV
Natural Environment Research Council, Centre for Ecology and Hydrology	NERC
Westlakes Scientific Consulting Ltd	WSC
Radiation and Nuclear Safety Authority	STUK
Institute for Radiological Protection and Nuclear Safety	IRSN
GSF - National Research Center for Environment and Health, GmbH	GSF
Norwegian University of Life Sciences (previously NLH)	UMB
Electricité de France	EDF

Executive Summary

Background and objectives

This report is an annex of D5. It summarises the principles, guidelines and statistical analysis applied when designing and carrying out controlled laboratory experiments to investigate biological effects of radioactive stressors in non-human organisms. Its specific domain of application is the study of dose(rate)-effect relationships for chronic (long-term) exposure of organisms to low-level of radionuclides. These guidelines have been applied in specific and detailed protocols that allowed a focus on two extrapolation issues: (1) external *vs.* internal irradiation exposure and (2) individual *vs.* population effects. Rather than focusing on filling specific data gaps for specific organisms and specific endpoints, the planned studies have the more generic objective of demonstrating the types of methodology and modelling that can be applied to two fundamental extrapolation issues. The experiments planned in ERICA address both issues for two organisms (earthworm and daphnid), with a particular emphasis on chronic external and internal irradiation and a number of vital rates such as survival, growth and reproduction (which are basic parameters in modelling from individual to population).

Basically, the proposed studies aim to establish dose(rate)-effect relationships for external (^{60}Co or ^{137}Cs sources) and internal (^{241}Am) irradiation (dose rates up to ~40 mGy/h), looking at a variety of reproduction endpoints (*e.g.* according to the biological model tested, number of offspring, survival and growth of offspring, sexual maturation, DNA damage in sperm) over different life-stages and generations. The main output of these studies will be a demonstration of the way in which experimental testing and mathematical modelling can be applied with respect to adequate statistical analysis and provide a better estimate of the scientific uncertainty associated with data extrapolation. While the individual results will also have direct relevance to the organisms and biological endpoints in question, the main aim is that the methodology could be applied to a variety of organisms and applications.

Designing an experimental study

The study plan of any experiment will develop four main successive parts.

- Identification of the general field of the experiments with a brief title, a statement on the nature and purpose of the study, the selected umbrella endpoint to investigate and the wildlife group.
- Reference to tested methods or test guidelines used from the literature if any and justification for their selection – Description of the selected exposure pathway.
- Precise description of the experimental design including a chronological description of the procedure and of the records to be retained. This includes recommendations on a number of crucial aspects e.g. randomisation, control group(s), sample size and replication, number and spacing concentration or dose(rate)s, times for sampling.
- Reporting of results, statistics and modelling.

This report also identifies general principles for carrying out experiments to meet one of the two aims: (1) hypothesis testing and/or (2) concentration- or dose(rate)-effect modelling. According to the objectives of the experiment, one of these two experimental designs will be preferred:

- A control group plus a single concentration- or dose(rate) group;
- A multiple concentration- or dose(rate) design.

To be properly scaled, these experimental designs will be examined from the power analysis point of view. A significant effect is generally meant to be a statistically significant effect as resulting from a hypothesis test. Data collected in a research study is submitted to a significance test to assess the validity of the null hypothesis. The power provided by the significance test, and used to reject the null hypothesis, is a function of three factors: the larger the observed effect, the larger the sample size, and/or the more liberal the criterion required for significance (α), the more likely it is that the test will yield a significant *p*-value. Choice of a power level ranges generally from 80 to 95 %. For a prospective power analysis, it explores the relationships between the range of sample sizes that are deemed feasible, effect sizes thought to be biologically important, levels of variance that could exist in the population (usually taken from the literature or pilot data) and desired levels of α and power. The result is a decision about the sample size and α -level that will be used in the study and the target effect size that will be detectable with the given level of statistical power. A retrospective power analysis can also be useful if a statistically non-significant result is obtained.

General approach for data analysis – Overview of parametric and non-parametric methods

The main stages of experimental data analysis are reviewed, from raw data examination to justifiable conclusions and its level of confidence (see Fig. A as example for continuous data).

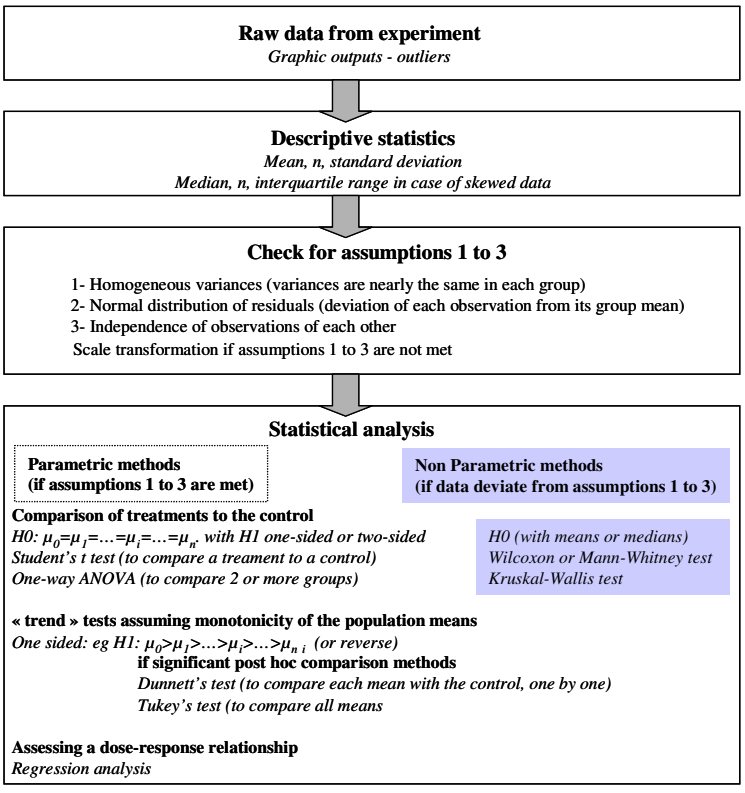


Figure A. Main stages to follow for experimental data analysis (case of continuous data). Only the most often used statistical methods are reported

In addition to traditional statistical analysis and concentration- or dose(rate)-effect modelling, biokinetics models may be useful to report and model change in the internal dose, and consequently to

estimate the internal doses. Such models have been largely published, from simple equilibrium relationships based on Concentration Ratios to complex dynamic models. Physiological models may be helpful to model change in a physiological endpoint such as growth or assimilation rate . When models from these categories will be used, assumptions, concepts, equations, limits and associated numerical recipes will be described.

Checklist of points to develop when designing and performing an experiment.

A pro forma is given in the Section 6 of this Annex. As a summary, this checklist will remind any experimenter to give description and justifications for the following points. This checklist consists of :

- Identification of the experiment.
- Overview of materials and methods.
- Description of the experimental design.
- Results and analysis.

Table of contents

Associated reports -

D5- Derivation of Predicted-No-Effect-Dose-Rate values for ecosystems (and their sub-organisational levels) exposed to radioactive substances. Garnier-Laplace J. and Gilbin R. (Eds). ERICA, European Commission, 6th framework, Contract N°FI6R-CT-2004-508847.

D5-Annex Part B. Experiments on chronic exposure to radionuclides and induced biological effects on two invertebrates (earthworm and daphnid). Results and discussion. Gilbin R. and Oughton D. (Eds). ERICA, European Commission, 6th framework, Contract N°FI6R-CT-2004-508847.

1 Introduction and general objectives

This report summarises the principles, guidelines and statistical analysis which should be applied when designing and carrying out controlled laboratory experiments to investigate biological effects of stressors in non-human organisms. Its specific domain of application is the study of dose(rate)-effect relationships for chronic (long-term) exposure of organisms to low-level of radionuclides. These guidelines have been applied to the specific protocols for dose-effect experiments planned within ERICA. They also have relevance for *a posteriori* statistical analysis of experimental data from the FRED and FREDERICA databases (see D5 Section 3.2 related to FRED data treatment to build dose(rate)-effect relationships).

A limited set of controlled experiments have been conducted within ERICA WP2. Rather than focusing on filling specific data gaps for specific organisms and specific endpoints, the planned studies have the more generic objective of demonstrating the types of methodology and modelling that can be applied to two fundamental extrapolation issues:

Weighting factors for internal (alpha or beta emitters) *vs.* external gamma irradiation. In risk assessment, weighting factors represent average, heuristic multipliers or "rules of thumb" that have in turn been derived from a range of experimental observations of the relative biological effectiveness (RBE) of different radiation types. It is widely acknowledged that RBEs can vary according to, amongst other things, biological endpoint, organism, life-stage, and dose (rate). However, published data on external irradiation dose-effect relationships greatly outnumbers that from internal exposure, and there are relatively few experiments that enable a direct comparison of RBE for external and internal exposure—and even fewer that address biological endpoints other than mortality. This is particularly relevant for internal alpha and beta emitters, invertebrates (both aquatic and terrestrial), reproductive endpoints and chronic low-level exposures (Daniel *et al.* , 2003). Consequently, the weighting factors applied in ecological risk assessments are likely to represent a highly contested area of scientific uncertainty.

Individual *vs.* population. For most laboratory experiments, the existing data only concerns effects on the individual level of organisation or sub-individual level. One of the major issues in ecological risk assessment is to be able to shift from individual to population and from population to community. Although important in themselves for evaluating the potential toxicological effects of pollutants, knowledge on individual effects is rarely sufficient for environmental risk assessment. Characterisation of risks also needs to include an evaluation of the possible population and ecological consequences. Recently a number of ecological and science based mathematical models have been proposed (Kammenga and Laskowski, 2000) that might aid in carrying out such extrapolations.

The experiments conducted within ERICA addresses both these issues for two organisms (earthworm and daphnid), with a particular emphasis on chronic exposure by external or internal irradiation and a number of vital parameters such as survival, growth and reproduction (which are basic parameters in modelling from individual to population). A series of experiments has been carried out to provide a robust set of data that will permit a focused evaluation of the extrapolation issues together with associated uncertainties and statistical analysis. The experimental set-up takes account of existing information on biological effects from ionising radiation (available from the FRED database), as well as ecological models for population effects and established ecotoxicological testing methods. In this respect, the main criteria for selection of the test organisms was that: i) there was some data on radiation effects to enable a preliminary statistical evaluation of data; ii) that the organism had a life-cycle short enough to allow experiments on chronic reproduction effects (*i.e.* at least first generation) to be completed in the time span; and iii) that there were established ecotoxicological models and

methods regarding reproduction and population effects. A more detailed overview of the actual set up of the experiments can be found in the specific protocols described in the pro forma, and examples of proper applications of statistical power analysis at the end of this report (Section 6).

Basically, the proposed studies aims to establish dose(rate)-effect relationships for external (^{60}Co or ^{137}Cs sources) and internal (^{241}Am) irradiation (dose rates up to ~40 mGy/h), looking at a variety of reproduction endpoints (*e.g.* according to the biological model tested, number of offspring, survival and growth of offspring, sexual maturation, DNA damage in sperm) over different life-stages and generations. The main output of these studies will be a demonstration of the way in which experimental testing and mathematical modelling can be applied with respect to adequate statistical analysis and provide a better estimate of the scientific uncertainty associated with data extrapolation. While the individual results will also have direct relevance to the organisms and biological endpoints in question, the main aim is that the methodology could be applied to a variety of organisms and applications.

Within this overall framework, the main objectives of the proposed guidelines are to provide a generic detailed description of the experimental design, statistical analysis and reporting of results. Having followed these guidelines, the studiesconducted within ERICA WP2 conform to up-to-date scientific knowledge and standards with respect to experimental protocols. The conception of experimental design is performed on the basis of OECD test guidelines for testing of chemicals referring to the selected biological models *i.e.* earthworms and daphnids respectively for soil and water quality (OECD, 1998a; , 2004), from ISO standards (ISO, 1998; , 2000) and from ASTM standard guides (ASTM, 2003; , 2004). This approach is carried out according to application of Good Laboratory Practices both from OECD (ISO, 1998; OECD, 1999) and from the guidance recently proposed by the Environment Agency in UK for developing experimental protocols for chronic irradiation studies on wildlife(Wood *et al.* , 2003).

2 Definitions and general principles for “controlled” experiments

2.1 What is a “controlled” experiment?

Festing and Altman (2002) give the following definition for a controlled experiment: An experiment is a procedure for collecting scientific data in a systematic way in order to maximise the chance of answering an hypothesis correctly (confirmatory research) or to provide material for the generation of new hypotheses (explanatory research). (Festing and Altman, 2002)

In the present report, guidelines are devoted to “controlled” experiments where exposure treatments (delivered radiological dose(rate)) are under the control of the experimenter to analyse the pattern of the response of the tested endpoints for the selected species. These controlled experiments are of two categories: confirmatory research for testing extrapolation from external to internal irradiation exposures and explanatory research for establishing the dose(rate)-effect relationship on individual and population endpoints.

2.2 Main definitions

The following definitions apply for these guidelines. The majority of them come from ecotoxicology and only a few of these terms are specific to the radioactive stressors. This list is supplemented by a number of other definitions in particular for statistical terms in the glossary at the end of this report (Section 7).

Experimental unit

The experimental unit (or replicate) is the smallest unit of experimental material to which a treatment can be applied independently of all other units. By definition, experimental units (*e.g.* aquariums, beakers, or plant pots) must be able to receive different treatments. Each experimental unit may contain multiple sampling units (*e.g.* fish, daphnia or plants) on which measurements are taken. Within each experimental unit, sampling units may not be independent. However, in some special case situations, individual organisms (housed in common units) can be treated as the experimental units: these special cases require some proof or strong argument to demonstrate independence of organisms.

Exposure concentration, dose or dose rate

The exposure concentration, dose or dose rate is the "amount" that an organism is exposed to. For a chemical element radioactive or stable, it can be expressed as a concentration (quantity of the substance per volume or mass of the exposure source, in *mol* or *g* or *Bq per L* or *per g)*. For a radionuclide, the absorbed dose (in *Gy)* is the total quantity of energy imparted by ionising radiation to unit mass of tissue of the organism (1 Gy=1 J/kg); the absorbed dose rate refers to the absorbed dose delivered over a specified unit of time (*e.g.* $\mu Gy/h$).

Endpoint

In toxicity testing and evaluation it is the biological response that is measured. Endpoints vary with the level of biological organization being examined and include responses at the subcellular level to the community level such as (i) biomarkers (subcellular level), (ii) survival, growth, reproduction (individual level), (iii) primary production, and structure (and abundance) and function in a community (population or community level). Endpoints are used in toxicity tests as measures of effect.

Effect

An effect is the change in an endpoint under consideration when it is compared to a control.

LOEC (or LOED or LOEDR) and NOEC (or NOED or NOEDR)

The Lowest Observed Effect-Concentration is the lowest Concentration out of the tested Concentration at which a statistically significant difference from the control group is observed.

The No Observed Effect -Concentration is the tested concentration just below the LOEC. They are obtained by hypothesis testing.

The same definitions apply for Dose and Dose Rate (in place of Concentration)

EC_x , $ED_{x\ and}$ EDR_x

In ecotoxicology, the term EC_x is defined as the concentration associated with an effect x where x is defined as the percent change in the (average) level of the endpoint considered $x\% = 100\left(\frac{y(EC_x)}{y(0)} - 1\right)\%$. The same definition can apply for the Dose (ED_x) or the dose rate (EDR_x). These parameters are estimated by modelling (concentration-effects, dose-effects or dose rate-effect modelling). EC_{10} is generally preferred to *NOEC* as the latter depends on the experimental design.

Figure 1 illustrates the signification of a number of terms used when applying concentration- or dose(rate)-effect relationships.

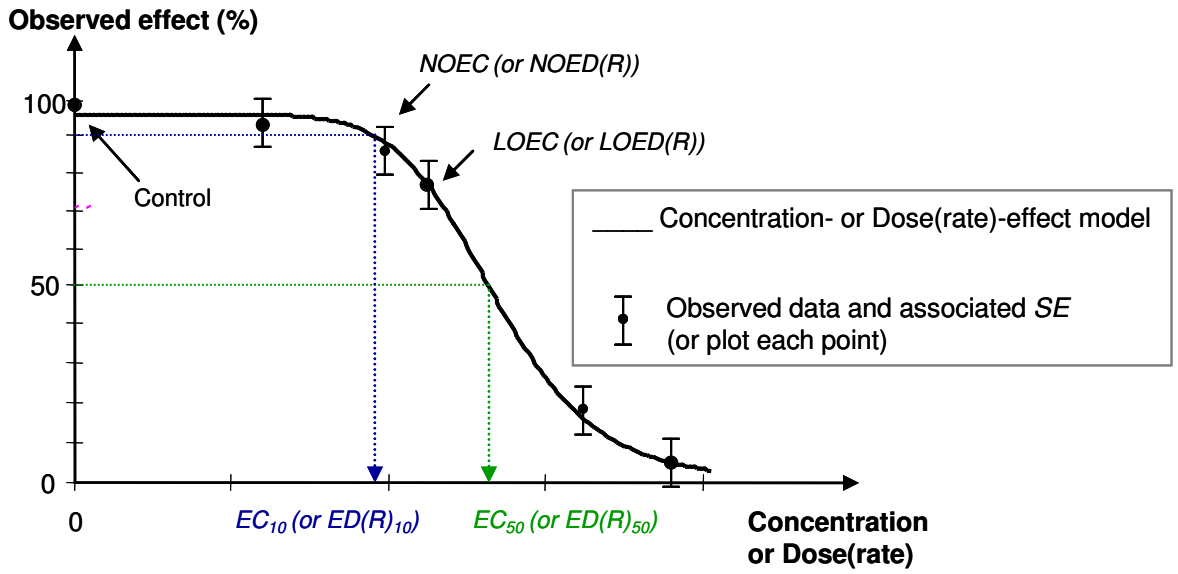


Figure 1. Illustration of the signification of a number of terms used in concentration- or dose(rate)-effect representation and interpretation.

2.3 General principles of statistics in toxicity testing and typical corresponding experimental design

The way statistics are applied may have considerable impact on the conclusions from any effect testing experiment. An extensive literature exists on the proper use of statistical methods, their domain of application and their limitations - *e.g.*(Box *et al.* , 1978; OECD, 2003; Sparks, 2000). This report highlights general principles to respect when carrying out experiments and applies statistical analysis to meet one of the two aims: (1) hypothesis testing and/or (2) concentration- or dose(rate)-effect modelling. According to the objectives of the experiment, two main experimental designs can be implemented:

- a control group plus a single concentration- or dose(rate) group– for instance, this design is recommended in OECD guidelines for testing of chemicals (OECD, 1999);
- a multiple concentration- or dose(rate) design – This design can be used to establish a *NOEC (or NOED(R))*, to build regression or mechanistic models for concentration- or dose(rate)-effect relationships and to calibrate population dynamic models.

To be properly scaled, these experimental designs need to be evaluated from the power analysis point of view. A significant effect is generally meant to be a statistically significant effect as resulting from a hypothesis test. The limit of detection for an effect depends on the quality and the size of the experiment and the statistical procedure used. The relationship between the detection limit and the quality of the experiment (including the sampling aspect) can be quantified by the concept of statistical power. Power calculations will use historical data to estimate the magnitude of within and among subgroup variation and correlation (see Section 2.3.3).

2.3.1 A control group plus a single concentration or dose(rate) for treatments comparison

The main goal is to compare responses among two (or more) test groups to a common control (or 0-dose(rate) group). One of the main uses in ecotoxicology is to estimate the difference in response

between two treatments. The usual analysis is to carry out a statistical test under the null hypothesis H0 (H0: *the two treatments give identical responses*). The hypothesis-testing methods can also be used to establish a *LOEC (or LOED(R))* and/or a *NOEC (or NOED(R))* with methods compatible or even recommended by OECD guidelines.

2.3.2 A multiple concentration or dose(rate) design

The main goal is to model concentration- or dose(rate)-effect relationship or to statistically accept or reject a "trend" hypothesis. Obviously, the corresponding design will be composed of more than two treatment groups and a common control. Data from this kind of design can be analysed by comparing means or by fitting regression models. This design can also be used according to the hypothesis-testing methods to establish a *LOEC* (or *LOED(R))* and a *NOEC (or NOED(R))*. The null hypothesis tested is H0: *all treatment groups give identical responses*. To build the concentration- or dose(rate)-effect relationship, the data analysis will include the use of regression models using stronger assumptions than the models used in the hypothesis testing approach. For the latter, the simplest statistical model generally used assumes only that the distributions of responses within the populations from treatment groups are identical.

2.3.3 Statistical power

A power analysis, executed when the study is being planned (prospective power analysis), can be used to anticipate the likelihood that the study will yield a significant effect and is based on the same factors as the significance test itself. Specifically, the larger the effect size used in the power analysis, the larger the sample size, and/or the more liberal the criterion required for significance (α), the higher the expectation that the study will yield a statistically significant effect. These three factors (effect size, sample size and significance criterion), together with power, form a closed system - once any three are established, the fourth is completely determined. The goal of a power analysis is to find an appropriate balance among these factors by taking into account the substantive goals of the study, and the resources available for the experiments.

The power of an experiment is then the probability to detect the specified effect for the given significance level and show it to be statistically significant. It corresponds to the probability that a false null hypothesis will be rejected by the statistical test in favour of a true alternative. Inversely, one minus the power is the chance of a false-negative result. Choice of a power level ranges generally from 80 to 95 %. For a prospective power analysis, it explores the relationships between the range of sample sizes that are deemed feasible, effect sizes thought to be biologically important, levels of variance for the effect variable (usually taken from the literature or pilot data) and desired levels of α and power. The result is a decision about the sample size and α -level that will be used in the study and the target effect size that will be detectable with the given level of statistical power. A retrospective power analysis can also be useful if a statistically non-significant result is obtained. In this case, the actual sample size is known, and the variance observed gives an estimation of the population variance. These values are used to calculate power at the minimum effect size thought to be of biological significance or the effect size detectable with the minimum desired level of power. Retrospective power analysis can also be used as a post-hoc power analysis to determine if the actual experiment is consistent with the criteria used at the design stage.

The important steps are to decide on what effect size should be considered to be large enough to be biologically important.

The most powerful statistical tests and the best experimental designs must be determined, rather than increasing the sample sizes.

The background variance can be taken as the pooled within experiment variance from a rather long period of historical control data (3 to 5 years). If more limited information on variance is provided, it would be appropriate to assume a slightly higher value than the one observed.

Section 5 gives more details and examples on statistical power analysis.

3 Experimental guidelines – Factors unique to consideration of radioactive stressors.

3.1 Scope

The experimental design will specify amongst other things the test concentrations of the substance (external and or internal dose(rate)s for radionuclides), the number of replicates, the number of subjects within each replicates as well as the times of observation. Usually, the independent variables are the concentration of the tested substances and the duration of exposure. For radionuclides and external exposure pathways, dose(rate)s will be calculated according to dosimetric models taking account of the exposure scenario, the geometry and the composition of the source and the target, the type of ionising radiation,…. For internal exposure pathway, one will need biokinetic knowledge at least at the whole organism level and for the radionuclide chemical speciation in the exposure source to allow reliable internal dose calculation. Within this global domain of application, the topics listed here after are recommended for inclusion in any specific experimental protocol to ensure a well-designed experiment.

3.2 Main outlines to follow when designing an experimental study

To plan any experiment properly, it is recommended that the approach suggested by the Environment Agency in the UK is followed (Wood *et al.* , 2003). Four main successive parts can be distinguished.

1. Identification of the general field of the experiments with a brief title, a statement on the nature and purpose of the study, the selected umbrella endpoint to investigate and the wildlife group.

- Clearly formulate the question and associated objectives.
- State the hypothesis to be tested.
- Justify the selected species and effects endpoints.
- Give the main outlines from an updated literature review on the investigated field. Formalise them in terms of lessons learnt for further experiments. Justify whether a pilot study is needed.

2. Reference to tested methods or test guidelines used from the literature if any and justification for their selection – Description of the selected exposure pathway.

- Fully characterise the tested species, such as the species, strain, origin of the supply, husbandry information, etc.
- Characterise the exposure pathway and the needed facilities used for the species exposure including type of radiation, irradiation pathway; select the radionuclide and the needed dosimetric models and measurements.

3. Precise description of the experimental design including a chronological description of the procedure and of the records to be retained.

- Select the independent variable, *i.e.* the variable the effect of which will be investigated and the dependent variable (*i.e.* the effect on an endpoint); give justification.
- Specify the constant factor (those that do not change) including the tested medium, the tested organism, its life stage, its feeding rate and diet; specify how these factors will be monitored.
- Plan the various treatments (exposure pathway, duration, range, spacing and values for the applied concentration and/or dose(rates)…), with a special attention to the control (group that serves as a standard of comparison).
- Decide the number of replicates, the number of individuals in each replicate and describe the statistical methods to be used; justify the choices.
- Describe all other pertinent points for understanding the materials and methods: materials, reagents, analytical techniques, type and frequency of analysis and samples preparation.

4. Reporting of results, statistics and modelling (according to the objectives of the experiments and to the experimental design, the following tips will be relevant or not – see Section 4 for further details).

- Application of group comparisons and associated statistical significance for difference.
- Estimation of *NOEC* (or *NOED* or *NOEDR*) if needed.
- Application of concentration- or dose(rate)-effect models.
- Application of population dynamics models if necessary.

These outlines are gathered at the end of this report in a checklist and a detailed pro forma modified from this proposed by Wood *et al.* (2003) for adaptation to ERICA experiments.

3.3 Randomisation

Randomisation has to be used at each stage of any experiments as far as possible (selection of materials, measurement, individuals allocation to treatment groups…). The aim is to eliminate bias in estimates of treatment effect (bias due to unknown sources of variation) and to ensure independence of error terms in statistical models. Each experiment will involve a number of Experimental Units (EU) assigned at random to a treatment. Randomisation and blinding for experiments are applied within the limits of radioprotection of experimenters (for instance with regards to the total dose acceptable for an occupational work on a radioprotection point of view). The EU should also be the unit of statistical analysis.

When possible, a completely randomised design should be used. That means that the EUs can be assigned at random to a treatment group; experiment is performed at one time in one location or it can be assumed that time and location have negligible effects on the experimental material. Ranges of formal experimental design are well described in many statistical textbooks *e.g.* (Box *et al.* , 1978; Festing, 2003; Festing and Altman, 2002). In any case, the experimenters will describe the used experimental design and the associated statistical methods.

3.4 Controls

For ecotoxicity testing, OECD guidelines very often recommend a range for acceptable values for a number of effects on the controls (*e.g.* minimum acceptable percent survival).

In some cases, experiments need multiple controls. For instance, when the tested substances cannot be administrated successfully without a carrier, two control groups will be needed: one with and one

without the vehicle (so-called negative control). Data from these two control groups are analysed to determine whether the vehicle influences the measured effect (*e.g.* comparison of controls with a Wilcoxon test and whether controls differ from each other, the decision will be to drop negative control or to pool controls). Each group must have the same vehicle concentration as the control and the assumption of no interaction between the vehicle and the substance under testing must be made (testing with the addition of a negative control). It is sometimes of interest to have positive controls to ensure that the experimental protocols were actually capable of detecting changes in the selected endpoint.

Since the control group is used in every comparison of treatment to control, consideration can be given to allocate more subjects to the control than to the dose-groups. Allocation rules are defined to optimise power for a statistical test to be used. For example, it has been shown for Dunnett's test, that the power was optimised for $n_0=n\sqrt{k}$, where n_0 is the number of subjects in the control group, n the number in the exposed groups and k, the number of treatments.

For radionuclides and external or internal irradiation pathways, the background level of the room where the experiment is located should be measured precisely at the beginning and all throughout the exposure duration. It is recommended that typical background rates (~ 1 µGy/h) are used for control treatment during low-level chronic exposure effects studies.

3.5 Sample size : number of experimental units (replication) and number of individuals per unit

Experiments need to have justifications for the choice of the sample size *i.e.* the number of experimental units (or true replicates), which must be large enough to be able to detect a significant effect, but not too large, for ethical reasons and/or constraints of cost, including the amount of radioactive waste produced, resources and time. Power analysis or resource equation methods give an estimation of the sample size required. These types of analyses can also help, when no significant effect could be highlighted, to show the size of biological effect that the experiment was probably capable of detecting. Resource equation may be used when there is no information about the standard deviation of the effect and/or because the effect size of interest is difficult to specify. According to this method of resource equation, an appropriate number of EU can be determined, based on the number of degrees of freedom for the error term in the analysis of variance or t test used to assess the effect. This number corresponds to : $E=N-T-B$, where E, N, T and B are the error, total, treatment and block degrees of freedom (*df)* in the ANOVA respectively. It is suggested that this number should be between 10 and 20. With less than 10 *df*, good returns can be expected from adding more experimental units. On the contrary, with more than 20 *df*, little additional information would be gained while adding EU.

In any case, the number of EU (true replicates) influences the power in hypothesis testing and the confidence limits of estimated parameters. Replicates are assumed to be independent. To ensure the statistical power to be significant, the treatment groups need a sufficient number of replicates. When the aim is to estimate an EC_x (or $ED(R)_x$), the study design must have a sufficient number of concentration or dose(rate) groups. Three concentration or dose(rate) groups and the control group is an absolute theoretical minimum. It is known that the precision of the estimated *NOEC* (or *NOED(R)*) depends more on the number of replicates per group.

Pseudoreplication (housing effects when individuals are housed together; non-independence of individual organisms's responses) should be avoided as far as possible. If pseudoreplication cannot be avoided, then the non-independence can be addressed by taking the variation between containers into account in the statistical model (nested ANOVA to estimate the components of variance associated with each level of nesting – *e.g.* (Festing and Altman, 2002)).

3.6 Number and spacing concentration or dose(rate)s – Nominal *vs.* measured concentration or dose(rate)s

Factors that must be considered when developing experimental designs include the number and spacing of concentration or dose(rate)s, the number of individuals per concentration or dose(rate) group, the nature and number of subgroups within the treatment groups. Decisions on those factors are related to adequate power to detect effects.

For instance, to estimate the *NOEC* (or *NOED(R)*), the aim is to bracket its value with concentration or dose(rate)s closely spaced. The later can be selected according to available information to cover a range of exposure levels expected to be encountered in the field and to include at least one concentration or dose(rate) with a no- or a very low- expected effect. This range can be refined and concentration or dose(rate)s closely spaced if more information is available. When effects are expected to increase in proportion to the log of concentration or dose(rate), the values selected can be equally spaced on some scales (*e.g.* $\log_2$ or $\log_{10}$ scale) to facilitate the statistical analysis. Three to seven concentration or dose(rate)s are suggested (plus the control). For a proper estimate of an EC_x (or $ED(R)_x$), this later should be bracketed and should be different from the effect value specific to the control group and the maximum effect value (if it is known).

Selected dose(rate)s should be both measured and calculated. Adequate dosimetric calculations should be performed, from very simple and large-scale models to 3-D micrometric scale models if needed. Models, assumptions, scenarios, results and uncertainties should be justified while designing the experiment. Means for dose(rate)s measurement should also be reported, with detection limit and measurement errors.

3.7 Times for sampling/measurements

It is known that EC_x values generally decrease for increasing exposure time, as long as the exposure concentration and the organisms' sensitivity remain constant. That means that actually dose-response relationship has to be represented by surface response instead of response curve. For internal exposure pathway, biokinetics of the radionuclide, taking account of the bioavailability of the various chemical forms, are therefore important to report and to model temporal change in the internal dose(rate) at various organisational level. Physiological models may be helpful to model temporal change in a physiological endpoint; and finally change in an effect endpoint. All models used in one field or another will be explained and justified. The selected times for sampling and associated measurements should be argued with regards to models and statistics used.

4 Guidelines for statistical analysis and reporting of results

4.1 General approach for data analysis – Overview of parametric and non parametric methods

A general scheme of the statistical approach for data analysis is given on Figure 2.

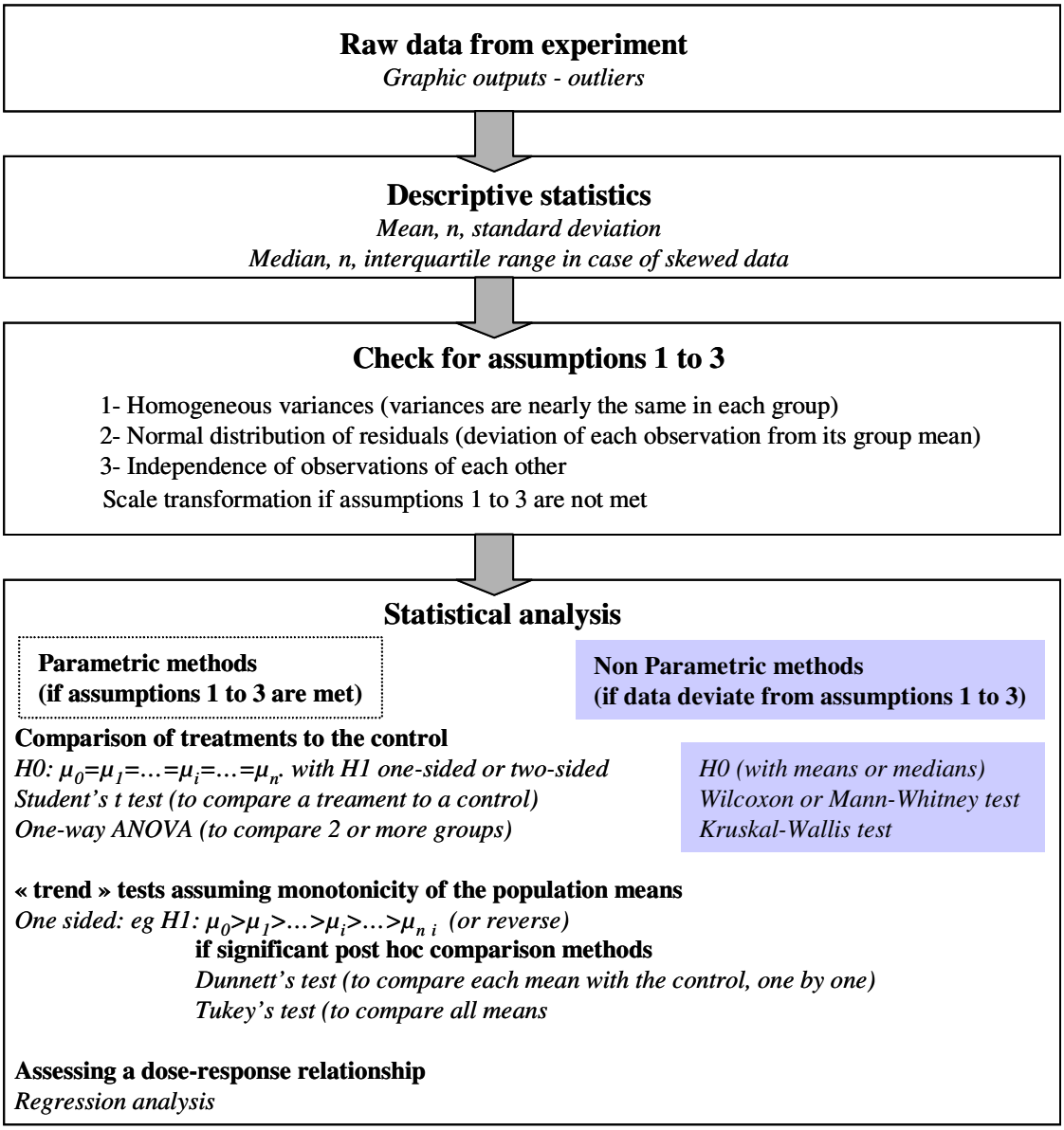


Figure 2. Main stages to follow for experimental data analysis (case of continuous data). Only the most often used statistical methods are reported.

At first, raw data (quantal and/or continuous) have to be analysed for their consistency with a focus on outliers. Those should be discarded only if reasons exist for their exclusion. Performed statistics with and without outliers could be useful at least to quantify whether or not they alter the conclusions. Then quantitative data will be expressed in terms of descriptive statistics with for example *mean*, *n* and Standard Deviation (*SD*). Statistic analysis can be performed to assess if the means or distributions of the different groups differ. Parametric methods are preferred but underlying assumptions need to be met: (1) variances are the same in each group, (2) residuals (deviation from the mean in each group) are normally distributed; (3) all observations are independent of each other. A scale transformation can be used if the assumptions listed before are not met (*e.g.* log to normalise the data). Many standard parametric methods (such as ANOVA, *t*-tests, linear regression analysis) assume normally distributed

data and homogenous variances and can be applied in case assumptions (1) to (3) are met. When data are not appropriate (*i.e.* if residuals – deviation from the group means – do not have a normal distribution, or if the variation is not roughly the same in each group), non-parametric methods can be used.

Among parametric methods, three main categories of analysis can be distinguished: those allowing comparison of treatments to the control; those corresponding to trend tests and those assessing dose(rate)-effect relationship. As general recommendation, Student's t-test should not be used when more than 2 groups are to be compared. Two groups or more will be treated with ANOVA. Usually, ANOVA is used to test the H0 that there is no difference among the treatment means. Where the ANOVA is significant (at $p<0.05$), *post hoc* comparisons can be used to study difference amongst individual means (Dunnett's test for comparing each mean with the control, others *e.g.* Tukey's test for comparing all means). Where there are several concentrations or dose(rate) levels, assessing concentration- or dose(rate)-effect relationship by using regression should be considered in preference to comparing each treatment with the control.

There are several non-parametric tests for equality of population means (*e.g.* Mann-Whitney test equivalent of the two-sample t-test; Kruskal-Wallis test equivalent of the one-way ANOVA).

Concerning specifically the *NOEC* (or *NOED(R)*) determination, pairwise comparisons or "trend-tests" could be used according to the selected experimental design. For n treatment groups and a control group (0-Dose(rate) group) and their associated mean μ_i with $i \in \{0, n\}$ to be compared, the most basic hypothesis can be stated as follows: *H0:* $\mu_0=\mu_1=\ldots=\mu_i=\ldots=\mu_n$. According to the trend of the expected effect, the alternative *H1* should be one-sided in a given direction (*e.g. H1:* $\mu_0>\mu_i$ for at least one i if an exposure was expected to induce an increase in the endpoint or *H1:* $\mu_0\neq\mu_i$ for at least one i and two-sided form). If no trend is assumed, the statistics will be based on comparing each treatment to the control, independently to the other treatments. To go further, if one suggests that as the level of exposure is increased, the magnitude of effect will be increased, a model assuming monotonocity of the population means ("trend" model) will be: $\mu_0>\mu_1>\ldots>\mu_i>\ldots>\mu_n$ (or with reversed inequalities). If $\mu_3\neq\mu_0$ and $\mu_2=\mu_0$ then the *NOEC* (or *NOED(R)*) is the test concentration associated with μ_2. Generally a test based on such monotonous model is consistent with a model for establishing concentration- or dose(rate)-effect relationships.

4.2 Concentration – or dose(rate) – effect modelling

Any statistical concentration-effect or dose(rate)- effect model serves to express the observed effect endpoint as a function of the concentration or dose(rate), to provide a tool to estimate the parameters of interest (EC_x or ED_x, or EDR_x) and assess their confidence intervals. Such models always consist of a deterministic part (the predicted concentration- or dose(rate)-effect relationship) and a stochastic part (describing the "noise"). Numerous concentration- or dose(rate)-effect models exist in the (eco)toxicological literature. Generally the most frequently applied models are based on the common form as follows: $y=(p_4-p_1)f(x)+p_1$ where y is the response, x is the concentration or dose(rate), p_1 and p_4 are the boundaries of the effect zone, p_1 is the known or calculated effect at zero concentration or dose(rate) and p_4 is the effect expected for a concentration or dose(rate) tending towards infinity. The curve is a monotonic function varying from p_1 to p_4. $f(x)$ is a probability function of the concentration or dose(rate) x varying from 0 to 1 with the dose. The Macro Excel REGTOX (Vindimian, 2003) allows the calculation of the parameters that define several traditionally used models: Hill equation, Log-Normal, Weibull. A model: GENTOX is also proposed for some genotoxicity tests. Two examples are given here after:

The Hill equation characterised by two parameters : Hill number and EC_{50} (or ED_{50} or EDR_{50}) with the probability function written as follows : $f(x) = \frac{x^{nH}}{x^{nH} + EC_{50}^{nH}}$. (Note that $f(x)=1/2$ when $x = EC_{50}$ and that this model is analytically equivalent to the logit model with EC_{50} that can be substituted by ED_{50} or EDR_{50}).

The model GENTOX, which can be used in some cases where a linear induction is observed, followed by a toxic inhibition. The induction phase is fitted to a straight line and the toxic phase uses a Hill equation. The induction potential found is the slope at the origin of the induction phase. The corresponding equation is: $f(x) = \frac{(p_1 + p_4 x)}{1+(\frac{x^{nH}}{EC_{50}^{nH}})}$ with EC_{50} that can be substituted by ED_{50} or EDR_{50}).

Within REGTOX, the fitting is based on the Marquardt algorithm. The confidence intervals on the parameters are estimated by a bootstrap simulation which is entirely non parametric.

4.3 Other models required

In addition to dose-effect modelling, biokinetics models may be useful to report and model change in the internal dose, and consequently to estimate the internal doses. Such models have been largely published, from simple equilibrium ones based on CRs to complex dynamic models. Physiological models may be helpful to model change in a physiological endpoint such as growth or assimilation rate for instance. When models from these categories will be used, assumptions, concepts, equations, limits and associated numerical recipes will be described.

4.4 Some recommendations for reporting results

The final step is reporting the results with three types of information: the raw data, the justification of the methods and assumptions used, the results of the data analysis and conclusions.

Concerning the presentation of the results, the mean and the variation amongst individuals are to be reported in a clearly stated way to avoid confusion between Standard Deviations (*SD*) or Standard Errors and to show the precision of the mean (*SE, n*). The magnitude of any significant effects should be quoted with a confidence interval, a standard deviation or Standard error to indicate its precision, the exact *p*-values (rather than $p<0.05$). Lack of statistical significance should not be used to claim that an effect does not exist. The size of biological effect that an experiment was probably capable of detecting calculated by a power analysis should always be reported. For each experiment, results should be summarised as described in the checklist (see Section 5).

5 Examples of statistical power analysis of data from *Daphnia magna* reproduction test

5.1 Brief description of the test OECD TG 211 (1998)

The primary objective of the test OECD TG 211 (1998) is to assess the effect of a compound on the reproductive output of *D. magna*. Young females aged less than 24 hours at the start of the test, are exposed to the compound of interest added to the water at a range of concentrations. After 21 days, the total number of living offspring produced per parent alive at this time, is assessed. The reproductive output of the animals exposed to the test substance, is 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*). In addition, the data are analysed using a regression model in order to estimate the concentration that would cause an *x %* reduction in reproductive output (*i.e.* EC_{50}, EC_{10}). Survival of parent animals and time to production of first brood must be reported. Effects on other parameters may be examined, such as individual growth, intrinsic rate of increase of the population, number and size of broods per animal, number of aborted broods, number of males or ephippies. Obviously, the test can be followed up during the whole life of the parent and can be repeated for the nth generation.

The basic data processing for statistical analysis is presented in figure 3 illustrated for reproduction data (number of offspring per parent). This processing aims on one hand, at performing hypothesis testing to evaluate the LOEC/NOEC and on the other hand, at fitting a regression model to calculate effect concentrations.

5.2 Statistical basis for power calculation

There are different means to strengthen the conclusions concerning an experiment. The statistical tests and the experimental designs should favour the highest statistical power. The statistical power is the probability of correctly rejecting a false null hypothesis, which is equal to 1-β in classical hypothesis testing as explained in Table 1. The most powerful test is that with the lowest β given an α. Figure 3 shows the recommended statistical procedure to analyse reproduction data; the methods toward the left hand side tend to have the highest power.

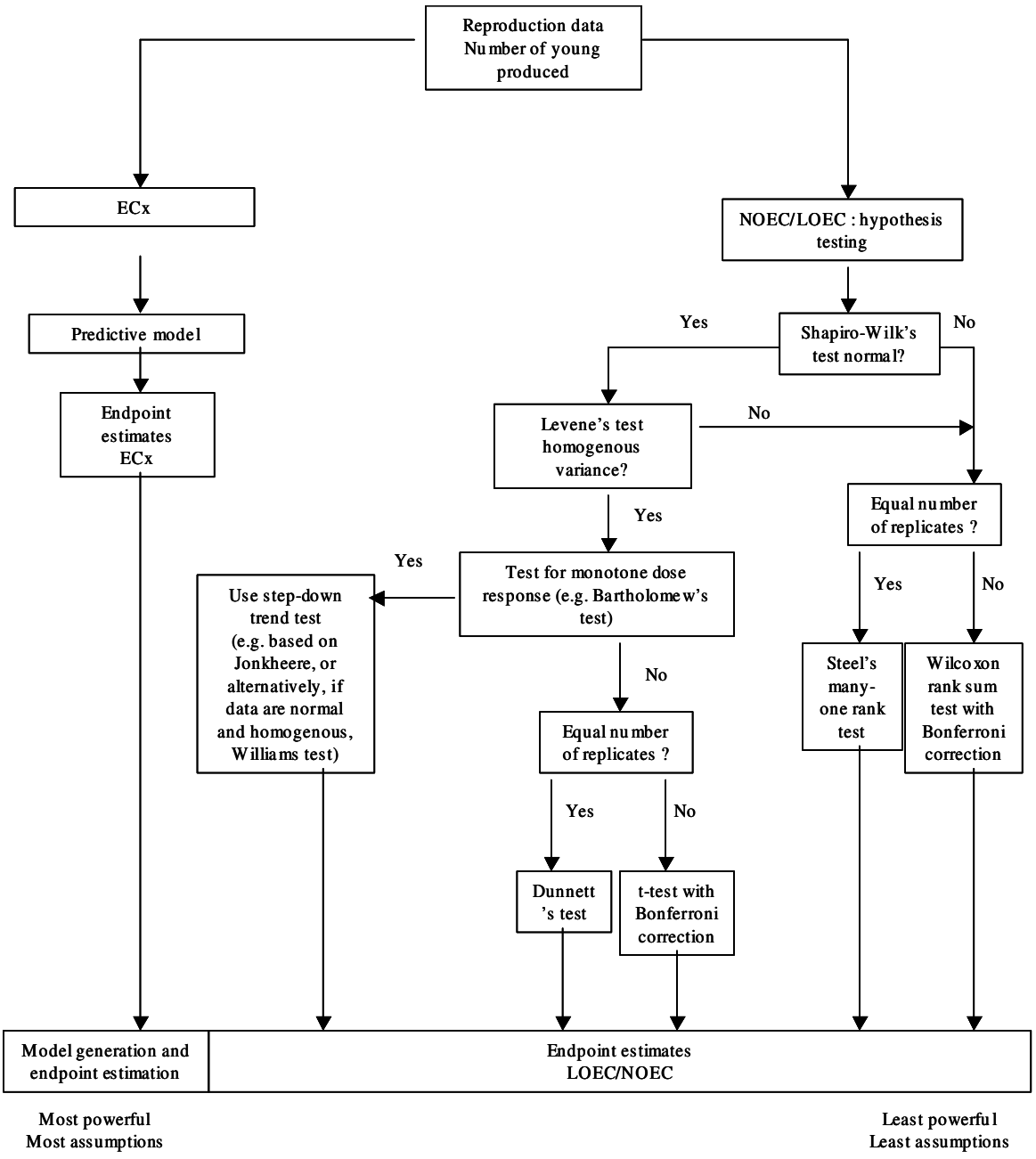


Figure 3. Flowchart for the recommended statistical analysis of D. magna reproduction test. Adapted from Newman (1994).

To illustrate the concept of power, Figure 4 presents a typical graph of the sampling distributions of the sample mean for two samples. The null hypothesis is $H_0 : \mu \leq m_1$ *vs* $H_1 \mu > m_1$. The significance level α, the probability of a type II error β and the power of the test $(1-\beta)$ are indicated.

Table 1. Statistical decisions and different types of errors.

		True State of the null hypothesis	
		H_0 Null is true	H_1 Null is false
Decision	H_0 Retain null	Correct acceptance $(1-\alpha)$	Type II error (false negative) β
	H_1 Reject null	Type I error (false positive) α	Correct rejection $(1-\beta)$=power

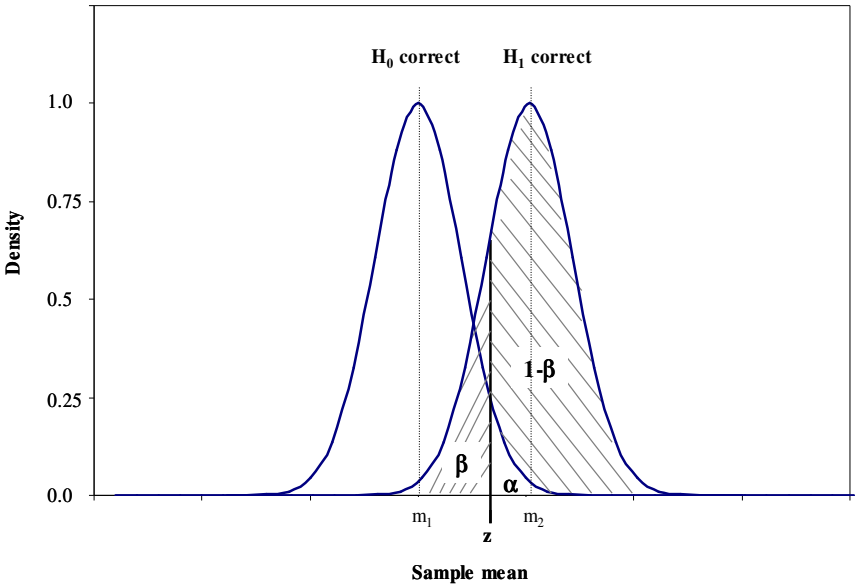


Figure 5. Sampling distributions and error probabilities.

Important factors need to be considered to calculate the statistical power: the kind of statistical test performed, the sample size and the size of experimental effect. To detect a reasonable departure from the null hypothesis, acceptable levels of power are 0.8-0.9. Graphical approach to power analysis requires the construction of graph relating to: power, sample size, the amount by which the null hypothesis is wrong (experimental effect) and type I error rate. This analysis involves different steps:

- specify the type of analysis and H_0;
- power and required sample size are investigated for a reasonable range of effect;
- the sample size required to detect a reasonable experimental effect (*i.e.* departure from the H_0) with a reasonable level of power, is calculated.

Some general-purpose statistics packages have built-in power capabilities, but with limited options. Thomas & Krebs (Thomas and Krebs, 1997) recommend general purpose power packages such as nQueryAdvisor, PASS or Stat Power. It must however be underlined that even these specialised packages offer limited possibilities for calculating statistical power for non-parametric step-down trend test. For example, nQuery Advisor offers only the Mantel-Haenszel (Cochran) test required for quantal data, while for continuous data there is no test available that could be equivalent to the Jonckheere-Terpstra test. Instead, this software uses the Mantel-Haenszel test with a continuity correction. In PASS software, to our knowledge, there is not any non-parametric step-down trend test available.

5.3 Some numerical applications

For the following tests, the trial version of PASS software ((PASS Trial version, Hintze (2001), NCSS and PASS, Number Cruncher Statistical systems, Kaysville, Utah) has been used, since the other recommended tool nQueryAdvisor does not allow to run completely an analysis with its trial version. A statistical power analysis tool such as PASS can be used either as a prospective tool (i) to ensure a powerful experimental design or (ii) for calculating the power of a test on data that have already been collected and analysed.

5.3.1 Prospective analysis

Example 1: reproduction data are analysed using the Dunnett's test with a significance level of 0.05. Previous studies have indicated that the standard deviation is 10 (equal variances for all the groups). The typical mean response level (number of offspring per parent) is 65. The researcher is interested in finding a 50 % decrease in the mean number of offspring. Since 0.5(65) = 32.5, this is the number that will be used as the minimum detectable difference.

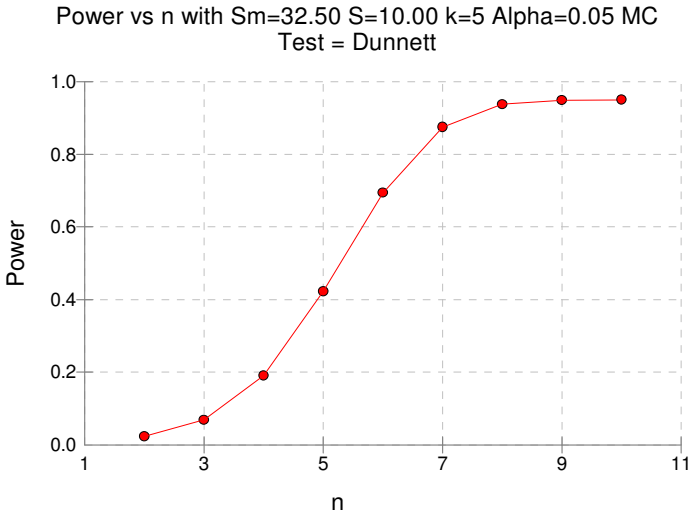
To better understand the relationship between power and sample size (corresponding to the number of replicates or EU in that case), the power is computed for several group sample sizes between 2 and 10. The sample sizes will be equal across all groups (the number of groups is noted *k*). All these main options are set in PASS before running the analysis. The principal part of the statistical report and the plot of results relative to this example are displayed:

Multiple Comparisons Power Analysis

Numeric Results for Multiple Comparison Test: Dunnett (With Control)

Power	Average Standard Size Deviation (n) (S)	k Diff / S	Total N	Alpha	Beta	Minimum Detectable Difference
0.02333	2.00 10.00	5 3.2500	10	0.05000	0.97667	32.50
0.06884	3.00 10.00	5 3.2500	15	0.05000	0.93116	32.50
0.19102	4.00 10.00	5 3.2500	20	0.05000	0.80898	32.50

0.42263	5.00 10.00	5 3.2500	25	0.05000	0.57737	32.50
0.69436	6.00 10.00	5 3.2500	30	0.05000	0.30564	32.50
0.87466	7.00 10.00	5 3.2500	35	0.05000	0.12534	32.50
0.93792	8.00 10.00	5 3.2500	40	0.05000	0.06208	32.50
0.94905	9.00 10.00	5 3.2500	45	0.05000	0.05095	32.50
0.94998	10.00 10.00	5 3.2500	50	0.05000	0.05002	32.50



This plot gives a visual presentation to the results in the Numeric Report. We can see the impact on the power of increasing the sample size. If we need a power higher than 0.8 then the number of replicates need to be at least 7.

Example 2: the same procedure can also be performed for different minimum differences (10, 20, 30, 40, 50).

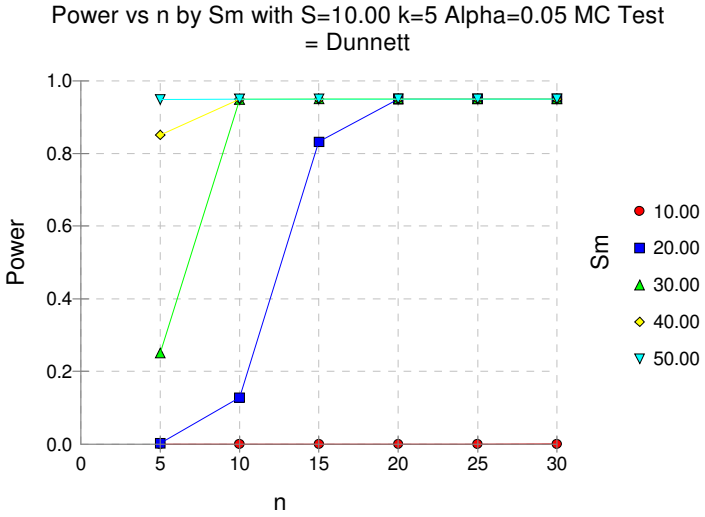
Multiple Comparisons Power Analysis

Numeric Results for Multiple Comparison Test: Dunnett (With Control)

Power	Average Standard Size Deviation (n) (S)	k Diff / S	Total N	Alpha	Beta	Minimum Detectable Difference
0.00000	5.00 10.00	5 1.0000	25	0.05000	1.00000	10.00
0.00000	10.00 10.00	5 1.0000	50	0.05000	1.00000	10.00

0.00000	15.00 10.00	5 1.0000	75	0.05000	1.00000	10.00
0.00000	20.00 10.00	5 1.0000	100	0.05000	1.00000	10.00
0.00000	25.00 10.00	5 1.0000	125	0.05000	1.00000	10.00
0.00006	30.00 10.00	5 1.0000	150	0.05000	0.99994	10.00
0.00228	5.00 10.00	5 2.0000	25	0.05000	0.99772	20.00
0.12810	10.00 10.00	5 2.0000	50	0.05000	0.87190	20.00
0.83173	15.00 10.00	5 2.0000	75	0.05000	0.16827	20.00
0.94990	20.00 10.00	5 2.0000	100	0.05000	0.05010	20.00
0.95000	25.00 10.00	5 2.0000	125	0.05000	0.05000	20.00
0.95000	30.00 10.00	5 2.0000	150	0.05000	0.05000	20.00
0.25065	5.00 10.00	5 3.0000	25	0.05000	0.74935	30.00
0.94872	10.00 10.00	5 3.0000	50	0.05000	0.05128	30.00
0.95000	15.00 10.00	5 3.0000	75	0.05000	0.05000	30.00
0.95000	20.00 10.00	5 3.0000	100	0.05000	0.05000	30.00
0.95000	25.00 10.00	5 3.0000	125	0.05000	0.05000	30.00
Power	Average Standard Size Deviation (n) (S)	k Diff / S	Total N	Alpha	Beta	Minimum Detectable Difference
0.95000	30.00 10.00	5 3.0000	150	0.05000	0.05000	30.00
0.85124	5.00 10.00	5 4.0000	25	0.05000	0.14876	40.00
0.95000	10.00 10.00	5 4.0000	50	0.05000	0.05000	40.00
0.95000	15.00 10.00	5 4.0000	75	0.05000	0.05000	40.00
0.95000	20.00 10.00	5 4.0000	100	0.05000	0.05000	40.00
0.95000	25.00 10.00	5 4.0000	125	0.05000	0.05000	40.00
0.95000	30.00 10.00	5 4.0000	150	0.05000	0.05000	40.00

0.94872	5.00 10.00	5 5.0000	25	0.05000	0.05128	50.00
0.95000	10.00 10.00	5 5.0000	50	0.05000	0.05000	50.00
0.95000	15.00 10.00	5 5.0000	75	0.05000	0.05000	50.00
0.95000	20.00 10.00	5 5.0000	100	0.05000	0.05000	50.00
0.95000	25.00 10.00	5 5.0000	125	0.05000	0.05000	50.00
0.95000	30.00 10.00	5 5.0000	150	0.05000	0.05000	50.00



This plot shows that for a given sample size, there is less power when detecting a small minimum difference. The smallest difference detectable (20/65= 0.3 i.e. 30%) with a power more than 0.8 requires a sample size of 15.

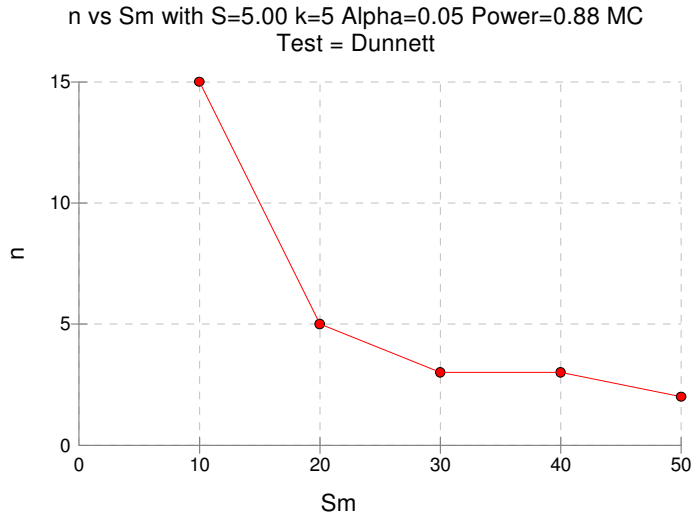
Example 3: the sample size can also be plotted as a function of the minimum difference that the experimenter wants to detect (Sm) for a given power of 0.8, and a standard deviation of 5.

Multiple Comparisons Power Analysis

Numeric Results for Multiple Comparison Test: Dunnett (With Control)

Power	Average Standard Size Deviation (n) (S)	k Diff / S	Total N	Alpha	Beta	Minimum Detectable Difference
0.83173	15.00 5.00	5 2.0000	75	0.05000	0.16827	10.00

0.85124	5.00 5.00	5 4.0000	25	0.05000	0.14876	20.00
0.85483	3.00 5.00	5 6.0000	15	0.05000	0.14517	30.00
0.94862	3.00 5.00	5 8.0000	15	0.05000	0.05138	40.00
0.88397	2.00 5.00	5 10.0000	10	0.05000	0.11603	50.00



This plot shows that the lower the difference is expected, the higher the sample size must be.

5.3.2 Post analysis: power after Dunnett's test

Example 4: an experiment has been performed, with 5 groups, 10 replicates for each group, giving a standard deviation of 5. The experimenter had hoped to show a significant difference of 10. The result of the test is that there is no significant difference. He wants to calculate the power of the experiment:

Numeric Results for Multiple Comparison Test: Dunnett (With Control)

Power	Average Standard Size Deviation (n) (S)	k Diff / S	Total N	Alpha	Beta	Minimum Detectable Difference
0.12810	10.00 5.00	5 2.0000	50	0.05000	0.87190	10.00

The power is only 0.13. Hence, there was little chance of detecting a difference of 10 between a treatment and a control group.

Example 5: it can be of interest to the experimenter to determine how large a sample was needed if the power was to be 0.80. Setting Beta to 0.8 and 'Find/Solve For' "n" resulted in the following report:

Numeric Results for Multiple Comparison Test: Dunnett (With Control)

Power	Average Standard Size Deviation (n) (S)	k Diff / S	Total N	Alpha	Beta	Minimum Detectable Difference
0.83173	15.00 5.00	5 2.0000	75	0.05000	0.16827	10.00

It can be seen that instead of 10 replicates per group, 15 per group were needed.

Example 6: it is also of interest to determine how large a difference between the means could have been detected.

Numeric Results for Multiple Comparison Test: Dunnett (With Control)

Power	Average Standard Size Deviation (n) (S)	k Diff / S	Total N	Alpha	Beta	Minimum Detectable Difference
0.80000	10.00 5.00	5 2.4968	50	0.05000	0.20000	12.48

We see that a study of this size with these parameters could only detect a difference of 12.3. This explains why the results were not significant.

5.3.3 Future application to data obtained within ERICA project

The first experiments performed on the effect of radionuclides on *D. magna* and on *E. fetida* will provide an estimation of the variance of a number of effects endpoints. This will allow a determination of, for instance (as shown in example 1), the sample sizes required to observe significant changes of reproduction rate with a sufficient power. Post-analysis of the dataset will show if there is appropriateness between predicted and observed power. If there is a discrepancy, experimental designs could be optimised to lower the standard deviation, sample size could be increased or additional dose rates could be studied to increase the effect size.

6 Pro forma to describe experiments

This pro forma is to be filled in to define any ERICA experiment. It is adapted from Wood *et al.* (2003) on the basis of the checklist given in the general guidelines. Each key instruction is just presented here as a summary. All details are written in the specific protocols to which section and page refer. To get information on expected details for each key instruction, see the checklist previously cited.

1 – Identification of the experiment			page
1.1	Title		
1.2	Purpose-question to answer		
1.3	Umbrella endpoint of interest		
1.4	Wildlife Group and tested Species		
1.5	Main conclusions from updated literature review including pilot experiment if it exists		
1.6	Null hypothesis to be tested		
2 – General materials and Methods			page
2.1	Reference test if it exist		
2.2	Tested species	Strain: Life stage Feeding rate and diet	
2.3	Tested medium		
2.4	Maintenance conditions (constant factors such as temperature, light regime, diet)		
2.5	Acclimation period		
2.6	Exposure pathway		
2.7	Irradiation Type (internal/external/mixed)		
2.8	Facilities/radioactive sources required (*e.g.* Cs-137 source)		
2.9	Dosimetric (*biokinetics)* models/measurements used		
2.10	Specific endpoint(s) to study (*e.g.* No. of eggs produced)		
2.11	All analytical techniques used		
2.12	Planned statistical methods		
2.13	Other needed models	Physiological model Population Dynamic model	

3 – Experimental design			page
3.1	Natural range of variation of the effects		
3.2	First power analysis		
3.3	Number of test groups		
3.4	Identification of the Experiment Unit		
3.5	Exposure duration		
3.6	Experiment duration		
3.7	Treatment groups definition	Range spacing	
3.8	Dose(rate)s to use (nominal) (e.g. Background, 10, 20, 40, 80, 160, 320 etc Gy or µGy/h). Where known also indicate, in the brackets, the total dose received at each dose rate	Background = Dose rate 1 = ( ) Dose rate 2 = ( ) Dose rate 3 = ( ) Dose rate 4 = ( ) Dose rate 5 = ( )	
3.9	Dose(rate)s used (measured)		
3.10	Randomisation (method)		
3.11	Number of replicates		
3.12	Number of individuals in each replicate		
3.13	Tested endpoint (unit, time of sampling)		
4 – Results and analysis (to be duplicated for each tested endpoint)			
4.1	Raw data and graphics		
4.2	Outliers		
4.3	Descriptive statistics		
4.4	p-value for normality p-value for homoscedacity		
4.5	Statistical methods		
4.6	Actual power of the test		
4.7	Estimated or derived effect values and uncertainties		

4.8	Any other information		

6.1.1 Pro forma for earthworm experiments

1 – Identification of the experiment		
1.1	Title	Effects of chronic external gamma irradiation on reproduction endpoints in different lifestages of the earthworm *Eisenia fetida*:
1.2	Purpose-question to answer	The study aims to establish dose(rate)-effect relationships for external gamma irradiation and a variety of endpoints in the earthworm *E. fetida*. The endpoints studied relate primarily to reproduction, including number of cocoons produced, survival and growth of offspring, sexual maturation, and DNA damage in sperm. Although the study includes effects on the F0 generation, as is standard in ecotoxicology tests, the main focus is on effects on the F1 generation. The rationale is to extend the ecological relevance of the studies and provide a better basis for population effect modelling. For example, as discussed below, the time to sexual maturation for pre-adults is an important parameter for population dynamics. Also the inclusion of a wide number of reproduction endpoints, including state-of-the-art molecular techniques to study DNA damage in spermatogenic cells should contribute to a better understanding of the mechanisms involved and provide important data on potential biomarker applications.
1.3	Umbrella endpoint of interest	Reproduction
1.4	Wildlife Group and tested Species	Wildlife Group: Soil fauna, invertebrate Tested species: *Eisenia fetida* Common name: Earthworm/ Compost worm

1.5	Main conclusions from updated literature review	The most sensitive endpoints for *Eisenia fetida* are linked to reproduction (nb. of spermatogonia and hatchability of cocoons). In a recent study performed for the UK Environment Agency, *E. fetida* (F0) was exposed to 5 dose rates ranging from 0.2 – 8.5 mGy/h for 16 weeks. No significant effects on mortality, histopathological anomalies, weight and reproductive capacity (number of cocoons and viable offspring) in the exposed groups was found compared to individuals in the background groups. (R&D technical report P3-101/SP7, in press). However, the experiments suffered from a number of confounding factors that reduced the statistical power of the experiments. While the data is relevant for estimating gross expected effects in F0 generations, it is of less use in providing expected sample variation and in the statistical design of these particular experiments. In traditional reproduction tests the rate of cocoon production (number of cocoons per worm per week), hatchability, and number of hatchlings emerging from each cocoon are measured. Other parameters that could have a great influence on the population dynamics are the rate of growth and sexual maturation of the offspring. Modelling results indicate that earthworm populations are more sensitive to toxicants that retard maturation. Klok and de Roos (1998) therefore recommended that laboratory tests on the toxicity of chemicals also should include chronic tests for pre-adult growth. This is also recognised as an important endpoint by Spurgeon et al. (2003) and Reinecke and Reinecke (2004).
1.6	Hypothesis to be tested	H0:Increasing dose rates do not modify the reproductive capacity. H0: No variation in sensitivity between different life stages (adult F0 reproduction; cocoon hatching; juvenile (F1) growth and sexual maturation, Adult F1 reproduction).
2 – General materials and Methods		
2.1	Reference test	Earthworm reproduction test (*Eisenia fetida/andrei*) (OECD, draft 2000), ISO guideline (1998)
2.2	Tested species	*Eisenia fetida* (originally from Centre for Soil Research, Norway) Life stage: Adults (age 2 mon-1yr), cocoons, juveniles (from hatching to adult (sexual mature))
2.3	Tested medium	Artificial soil (as recommended in OECD guideline)
2.4	Maintenance conditions (constant factors such as temperature, light regime, diet)	21 ± 2 °C 16 hours light, 8 hours dark Soil moisture: ~ 57 % of maximum water holding capacity of the soil. Soil pH ~ 6 Diet and feeding rate: Air dried and rewetted horse manure: 0.5 g dry w/ worm/ week
2.5	Acclimation period	2 weeks in artificial soil, supplied with the same feed as in the rest of the experiment
2.6	Exposure pathway	External (Co-60 source in front of the experimental units)
2.7	Irradiation Type(internal/external/mixed)	External gamma irradiation

2.8	Facilities/radioactive sources required	Co-60 source (5 Ci) in a controlled, irradiation facility. The source and experimental room are housed in a temperature-regulated building.
2.9	Dosimetric (biokinetics) models/measurements used	Pre-calibrated external dose rates (including with soil-boxes) Dosimetry calibration using ionisation chamber supplied with measurement with portable dosimetry. Irradiation field characterisation under experiments, dose measured with Mg,Ti thermoluminescence detectors (TLD).. Dose calculation models (NRPA, FASSET)
2.10	Specific endpoint(s) to study (e.g. No. of eggs produced)	**Reproduction and maturation study:** **Adult F0 reproduction:** Viability, weight, morbidities, number of cocoons produced per worm per week, hatchability, number of hatchlings (F1) per cocoon, DNA damage in somatic and spermatogenic cells **Hatchlings/juveniles (F1)**: Viability, morbidities, growth and sexual maturation rate **Adult F1 reproduction**: Viability, weight, morbidities, number of cocoons produced per worm per week, hatchability, number of hatchlings (F2) per cocoon, DNA damage in somatic and spermatogenic cells **Irradiation of cocoons** (from unirradiated adults): Incubation time for hatching, hatchability (%), number of hatchlings per cocoon
2.11	All analytical techniques used	pH, water content (OECD guideline) Ge(Li)-detector for determination of natural radioactivity in the soil and worms Determination of heavy metals in the soil and worms using ICP-AES, ICP-MS, AAS Cell staining and microscopy for determination of viability Comet assay SCSA (sperm chromatin structure assay)
2.12	Planned statistical methods	A multiple concentration- or dose(rate) design – This design can be used to establish a *NOEDR*, to build regression for dose rate - effect and to calibrate population dynamic models
2.13	Other needed models	Earthworm life cycle and population dynamic models do exist, but to date have been used mostly for other species than *E. fetida* within ecological studies. Applications of generic population models (eg. Leslie matrix) to the results of experiments has been carried out separately (see D5)
3 – Experimental design		
3.1	Natural range of variation of the effects/endpoints	The control values of the endpoints will depend on factors like temperature, soil moisture, density, substrate and food supply. Number of cocoons per worm per week: 2 - 5 Incubation time for hatching of cocoons: 3 - 4 weeks Number of hatchlings per cocoon: 1-6 Time from hatching to sexual mature adult: 8 – 12 weeks (20 ± 2 °C)(OECD, 2000)

		Comet assay: few data in the literature. To be completed during the test Validity of test: OECD recommendations on acceptable control statistics will be applied.
3.2	First power analysis	To be completed
3.3	Number of test groups	1 control + 5 dose rates (cocoons: 1 control + 6 dose rates)
3.4	Identification of the experiment unit	**Reproduction and maturation study:** Test container with 500 g dry-weight equivalent soil and 10 adult or juvenile earthworms **Irradiation of cocoons:** Small vial with 2 g dry-weight equivalent soil and 1 cocoon
3.5	Exposure duration	**Reproduction and maturation study:** 13 weeks for adult (F0) reproduction 8 -12 weeks? for juvenile (F1) growth and sexual maturation 13 weeks for Adult (F1) reproduction **Irradiation of cocoons:** 3 weeks
3.6	Treatment groups definition	Range ~0.15 - 40 mGy/h (cocoons: ~ 0.15- 100 mGy/h) Spacing factor ranging from 2 to 10
3.7	Dose(rate)s to use (nominal)	**Reproduction and maturation study:** Background < 5 µGy/h Dose rate: — Total dose (4 - 13 weeks): Dose rate 1 = 0.15 mGy/h — (0.10 - 0.33 Gy) Dose rate 2 = 1.5 mGy/h — (1.0 - 3.3 Gy) Dose rate 3 = 3.5 mGy/h — (2.4 – 7.6 Gy) Dose rate 4 = 10 mGy/h — (6.7 - 21.8 Gy) Dose rate 5 = 40 mGy/h — (26.9 – 87.4 Gy) **Irradiation of cocoons:** Background < 5 µGy/h Dose rate: — Total dose (3 weeks): Dose rate 1 = 0.15 mGy/h — (0.08 Gy) Dose rate 2 = 1.5 mGy/h — (0.76 Gy) Dose rate 3 = 10 mGy/h — (5.0 Gy) Dose rate 4 = 20 mGy/h — (10.1 Gy) Dose rate 5 = 40 mGy/h — (20.2 Gy) Dose rate 6 = 100 mGy/h — (50.4 Gy)
3.8	Dose(rate)s used (measured)	**Reproduction and maturation study:** Dose rates, adult (F0): — Total dose (4 - 13 weeks): Dose rate 1 ~ 0.19 mGy/h — (0.11 - 0.37 Gy) Dose rate 2 ~1.8 mGy/h — (1.1 - 3.6 Gy) Dose rate 3 ~ 4.2 mGy/h — (2.7 – 8.6 Gy) Dose rate 4 ~ 11 mGy/h — (7.1 - 23 Gy) Dose rate 5 ~ 43 mGy/h — (26 – 85 Gy) Dose rates, juveniles (F1): — Total dose (at week 11):

		Dose rate 1 ~ 0.18 mGy/h (0.30 Gy) Dose rate 2 ~1.7 mGy/h (2.9 Gy) Dose rate 3 ~ 4.0 mGy/h (7.0 Gy) Dose rate 4 ~ 11 mGy/h (19 Gy) Dose rates, adult (F1): Total dose (week 12-24): Dose rate 1 ~ 0.18 mGy/h (0.44 – 0.64 Gy) Dose rate 2 ~1.7 mGy/h (4.2 – 6.1 Gy) Dose rate 3 ~ 4.0 mGy/h (10 - 15 Gy) Dose rate 4 ~ 11 mGy/h (27 - 40 Gy)
3.9	Randomisation (method)	Allocation of worms to boxes (OECD) Rotating boxes (replicates) Choice of individual/group to be analysed Return of juveniles to experimental study (random selection)
3.10	Number of replicates	**Reproduction and maturation study:** Control: 12, exposed (dose rate 1-4): 4, highest dose rate: 1 **Irradiation of cocoons:** Control: 25, exposed: 10
3.11	Number of individuals in each replicate	**Reproduction and maturation study:** 10 adult (F0) or 10 hatchlings (F1) **Irradiation of cocoons:** 1 cocoon per replicate
3.11	Tested endpoint (units, sampling time)	**Reproduction and maturation study:** All sampling times from start of irradiation. Total experiment duration shall be approximately 34 - 38 weeks. **Adult F0 reproduction:** Day 28 and 56: Adult (F0) worms are taken out, weighed, transferred to new soil and irradiation is continued. Number of cocoons and hatchlings (F1) are registered in the old soil. Cocoons are transferred to small boxes with new soil and irradiation is continued. Hatchability of cocoons is registered after an additional 4,7 and 9 weeks. Day 91: Adult (F0) worms are taken out, weighed, and a sub-sample taken for analysis of DNA damage (comet assay, SCSA). Number of cocoons and hatchlings (F1) are registered. Cocoons are transferred to a small amount of new soil and irradiation is continued. **F1 growth and maturation:** Day 98: 10 hatchlings (F1) (from cocoons produced between Day 56 and 91) from each replicate are weighed, transferred to new soil, and irradiation is continued. Day 112, 126, 140, 154,168 and 182: determination of weight and sexual maturity (development of clitellum), registration of viability and morbidities. **Adult F1 reproduction:** Day 154, 168 or 182: If/when all F1 worms are sexually mature, they will be transferred to new soil, and irradiation will be continued for 13 weeks, with registration of reproduction capacity

		as described for F0. At the end of exposure, adult (F1) are taken out weighed and a sub-sample taken for analysis of DNA damage (comet assay, SCSA). **Irradiation of cocoons** (from unirradiated adults): Day 21: the cocoons are removed from the irradiation source. Every second day from Day 21 to 55: Registration of hatchability (hatching), and number of hatchlings emerging from each cocoon

6.1.2 Pro forma for daphnia experiments

1 – Identification of the experiment		
1.1	Title	Biological effects of **external chronic gamma irradiation** in the freshwater microcrustacean *Daphnia magna*.
1.2	Purpose-question to answer	This set of experiments aims at examining the effects of external irradiation at continuous low level dose rates on parthenogenetic populations of *Daphnia* magna. A robust set of data is needed: (1) to suggest **methods for extrapolating effects** of external irradiation **from individual to population** using a science-based reasoning (effects on vital rates and population dynamics modelling). (2) to assess the relevancy of weighting factors for internal *vs.*external irradiation at the population level by comparing results with those obtained during chronic internal irradiation experiments (see set 2).
1.3	Umbrella endpoint of interest	Reproduction **Life table** approach will include measurement of individual variables such as **survival probability** (Si) and **reproductive output** (Fi) in controlled conditions. Measurements will be carried out at least during 21 days, as recommended in the standard reproduction test.
1.4	Wildlife Group and tested Species	Wildlife Group : Crustacean Tested species: *Daphnia magna* Common name: Water flea
1.5	Main conclusions from updated literature review	An SSD from **FRED data** on *Daphnia pulex* (no data for *D. magna*) shows that effects occur from 20 to 600 mGy/h according to the sensitivity of the endpoint. The most sensitive endpoints are morbidity parameters linked to biomass production. An **endpoint linked to the biomass production** (size, dry weight of adults, eggs and neonates) is included.
1.6	Hypothesis to be tested	H0: Increasing dose rates for external irradiation do not modify the reproductive capacity at the population level.
2 – General materials and Methods		
2.1	Reference tes	ASTM Standard guide E 1193 – 97. OECD Guidelines for testing of chemicals 211 ISO 10706 USEPA Test Method 1002.0 (EPA-821-R-02-013)
2.2	Tested species and Feeding	Strain: *Daphnia magna* clone A (from INERIS, France) Life stage: juvenile (<24h) at the beginning of the test Feeding rate and diet: *Chlamydomonas reinhardtii* (1x/d)
2.3	Tested medium	M4 (recommended by OECD procedure)
2.4	Maintenance conditions (constant factors such as temperature, light regime, diet)	20°C(+/- 1) in thermostated incubators light illumination with fluorescent tubes (300lx) 6hrs/8hrs day/night food: 3.10^6 cells/daphnid/day (morning), i.e. 100 µg carbon per

		daphnid
2.5	Acclimation period	3 generations in a media-adapted husbandry
2.6	Exposure pathway	External (^{137}Cs source in front of the experimental units) medium renewed daily
2.7	Irradiation Type(internal/external/mixed)	External gamma irradiation (^{137}Cs)
2.8	Facilities/radioactive sources required	1 Line Source ^{137}Cs: (1.85 GBq, active length 50mm) 1 ^{137}Cs olution (200MBq; 5ml) diluted according to the different dose-rates required from CERCA (Framatome, Pierrelatte, France)
2.9	Dosimetric models/measurements used	**MCNP** Monte Carlo code (version 4C2, F. BRIESMEISTER, LA-13709-M (04/2000), "MCNP - A General Monte Carlo N-Particle Transport Codes, Version 4C") – Design for real-time measurement under examination
2.10	Specific endpoint(s) to study (e.g. No. of eggs produced)	Critical individual variables for extrapolating effects at the population level (measured during at least 21 days of exposure: number of living / dead females average time to first brood / time between broods number of living / dead neonates per female female, eggs and neonate dry weights
2.11	All analytical techniques used	pH, conductivity, O_2 by traditional techniques major ion concentrations measured by ion chromatography Dionex DX 120 DOC by Organic Carbon analyser
2.12	Planned statistical methods	A multiple concentration- or dose(rate) design – This design can be used to establish a *NOEDR*, to build regression for dose rate - effect and to calibrate population dynamic models
2.13	Other needed models	Models (eg. Leslie matrix) will be used in order to validate the suitability of various parameters as significant indexes of radiological effects at the population level.
3 – Experimental design		
3.1	Natural range of variation of the effects	Vital rates for daphnids in the husbandry
3.2	First power analysis	To be completed
3.3	Number of test groups	1 control + 3 dose-rates
3.4	Exposure duration	At least 21-day reproduction test defined in standard protocols
3.5	Treatment groups definition Rangespacing	Dose rates equally spaced on $\log_{10}$ scale
3.6	Dose(rate)s to use (nominal)	Background = <1 µGy/h Dose rate 3 = 400 µGy/h Dose rate 4 = 4000 µGy/h Dose rate 5 = 40 000 µGy/h
3.7	Dose(rate)s used (measured)	With Thermo Luminescent Dosimeters: Al2O3 ø = 3.97 mm, of dimensions 4x4x1 mm LiF ø = 2.64, of dimensions 4x4x0.05 mm placed on the front and on the back sides of the e.u..: Background = <1 µGy/h Dose rate 3 = 400 µGy/h

		Dose rate 4 = 4200 µGy/h Dose rate 5 = 31 000 µGy/h
3.8	Randomisation (method)	completely randomised design
3.9	Number of replicates	10 replicates per test group (2x10 replicates when possible, i.e for control group and 31 000 µGy/h group)
3.10	Number of individuals in each replicate	1 single individual per replicate (in 50ml)
3.11	Tested Effect endpoint (units, sampling time)	Daily measurements: number of living / dead females number of living / dead neonates per female Time to first brood / time between broods (days) Femaleand egg dry weight (µg): at time for brood 1, 3 and 5 Neonate dry weight (µg): on day of production

1 – Identification of the experiment		
1.1	Title	Biological effects of **internal chronic alpha irradiation** to the freshwater microcrustacean *Daphnia magna*.
1.2	Purpose-question to answer	This set of experiments aims at examining the **effects of internal irradiation** at **low level dose rates** on parthenogenetic **populations of *Daphnia* magna**. A robust set of data is needed: (1) to suggest **methods for extrapolating effects** of internal irradiation **from individual to population** using a science-based reasoning (effects on vital rates and population dynamics modelling). (2) to assess the relevancy of weighting factors for internal *vs.*external irradiation at the population level by comparing results with those obtained during chronic internal irradiation experiments (see set 1).
1.3	Umbrella endpoint of interest	Reproduction **Life table** approach will include measurement of individual variables such as **survival probability** (Si) and **reproductive output** (Fi) in controlled conditions. Measurements will be carried out at least during 21 days, as recommended in the standard reproduction test.
1.4	Wildlife Group and tested Species	Wildlife Group : Crustacean Tested species: *Daphnia magna* Common name: Water flea

1.5	Main conclusions from updated literature review	No data on the toxicity of americium is available. The concentration factor from water is evaluated at **2200 Bq.kg^{-1}.Bq^{-1}.L^{-1}** (Veran, 1998) for adult stage at equilibrium. In marine organisms, the target organs and tissues of bioaccumulation of stable and radioactive elements (^{238}U, ^{239}Pu and ^{241}Am), were shown to be mainly digestive gland, gill and exoskeleton (Chassard-Bouchaud, 1996). Americium in water is not very soluble. The use of an organic ligand to maintain Am in solution (eg. EDTA) will be used to avoid adsorption on experimental unit walls and exoskeleton of daphnids. Note that ^{241}Am experiments are limited to 6000Bq at eah time, for radioprotection reasons.
1.6	Hypothesis to be tested	H0: Increasing dose rates for internal irradiation do not modify the reproductive capacity at the population level
2 – General materials and Methods		
2.1	Reference test	ASTM Standard guide E 1193 – 97. OECD Guidelines for testing of chemicals 211 ISO 10706 USEPA Test Method 1002.0 (EPA-821-R-02-013)
2.2	Feeding	Strain: *Daphnia magna* clone A (from INERIS, France) Life stage: juvenile (<24h) at the beginning of the test Feeding rate and diet: *Chlamydomonas reinhardtii* (1x/d)
2.3	Tested medium	M4 (recommended by OECD procedure / contains EDTA)
2.4	Maintenance conditions (constant factors such as temperature, light regime, diet)	20°C(+/- 1) in thermostated incubators light illumination with fluorescent tubes (300lx) 6hrs/8hrs day/night food: 3.10^{6} cells/daphnid/day (morning)
2.5	Acclimation period	3 generations in a media-adapted husbandry
2.6	Exposure pathway	medium renewed each day
2.7	Irradiation Type(internal/external/mixed)	Mean concentrations of ^{241}Am in the medium over the course of the experiments are fixed to 0.4, 4.0 and 40 Bq/ml, respectively corresponding to 0.4-0.5 Bq, 2.5- 3.5 Bq and 27.5- 28.3 Bq per daphnid after a 23-day exposure, distributed in the cuticle and soft tissues. Contribution of cuticle is assessed by the measurement of 241Am content of molt (up to 45-50% of total 241Am on Day 7. and 16% after Day 10). Dose rates are mainly the result of internal alpha radiation from bioaccumulated ^{241}Am in daphnid tissues and cuticles. Daphnid were exposed to total dose rates of 0.01, 0.07 and 0.80 mGy/h respectively at the ^{241}Am concentrations of 0.4, 4.0 and 40 Bq per ml.
2.8	Facilities/radioactive sources required	^{241}Am (>99.9%, 2.0x10^{5} Bq in 5ml 2M HNO_3) from ISOTRAK AEA Technology QSA.(n° AMP10030).
2.9	Dosimetric (biokinetics) models/measurements	Calculations of dose conversion coefficients (DCC) for gamma and beta radiations are performed with the Monte Carlo code MCNP 4C (Briesmeister, 2000). DCC calculations for alpha

		radiations are performed with the Monte Carlo code McDPUC (Jasserand, 2003). The number of particles and their energies emitted by disintegration of an atom of ^{241}Am are determined by EMIXG code (Jasserand, 2002) using the international database JEF 2.2. The duration of all calculations is adjusted so that the statistical uncertainty associated with the results was lower than 1%.
2.10	Specific endpoint(s) to study (e.g. No. of eggs produced)	Critical individual variables for extrapolating effects at the population level (measured during at least 21 days of exposure: number of living / dead females average time to first brood / time between broods number of living / dead neonates per female female, eggs and neonate dry weights
2.11	All analytical techniques used	pH, conductivity, O_2 by traditional techniques major ion concentrations measured by ion chromatography Dionex DX 120 DOC by Oragnic Carbon analyzer ^{241}Am measured by alpha liquid scintillation (>0.03Bq)
2.12	Planned statitiscal methods	A multiple concentration- or dose(rate) design – This design can be used to establish a *NOEDR*, to build regression for dose rate - effect and to calibrate population dynamic models
2.13	Other needed models Physiological modelPopulation Dynamic model	Models (eg. Leslie matrix) will be used in order to validate the suitability of various parameters as significant indexes of radiological effects at the population level.
3 – Experimental design		
3.1	Natural range of variation of the effects	Vital rates for daphnids in the husbandry
3.2	First power analysis	To be completed
3.3	Number of test groups	1 control + 3 dose-rates
3.4	Exposure duration	At least 21-day reproduction test defined in standard protocols
3.5	Treatment groups definition Rangespacing	Dose rates equally spaced on $\log_{10}$ scale
3.6	Dose(rate)s to use (nominal)	Background = <1 µGy/h Dose rate 1 = 10 µGy/h Dose rate 2 = 100 µGy/h Dose rate 3 = 1 000 µGy/h
3.7	Dose(rate)s used (measured)	Background = <1 µGy/h Dose rate 1 = 10 µGy/h Dose rate 2 = 70 µGy/h Dose rate 3 = 800 µGy/h
3.8	Randomisation (method)	completely randomised design
3.9	Number of replicates	24 replicates per test group
3.10	Number of individuals in each replicate	1 single individual per replicate (in 50ml)
3.11	Tested Effect endpoint (units, sampling time)	Daily measurements: number of living / dead females number of living / dead neonates per female

		Time to first brood / time between broods (days) Femaleand egg dry weight (µg): at time for brood 1, 3 and 5 Neonate dry weight (µg): on day of production

7 Specific Glossary

Definitions are all adapted from (OECD, 2003). Terms are listed by alphabetic order.

▪Confidence interval

A *x %* confidence interval for a parameter is an interval of values that theoretically covers the true value of the estimated parameter with x % of confidence. Note that the confidence level reflects the proportion of cases that the confidence interval would contain the true parameter value in a long series of repeated random samples under identical conditions.

▪Continuous data

Data are continuous when they can theoretically take any value in an open interval

▪EC_x , ED_x, EDR_x

In ecotoxicology, the term EC_x is defined as the concentration associated with an effect x where x is defined as the percent change in the (average) level of the endpoint considered $x\% = 100\left(\frac{y(EC_x)}{y(0)} - 1\right)\%$. The same definition can apply for the Dose (ED_x) or the dose rate (EDR_x). These parameters are estimated by modelling (concentration-effects, dose-effects or dose rate-effect modelling).

▪Effect

An effect is the change in an endpoint under consideration when it is compared to a control.

▪Endpoint

In toxicity testing and evaluation it is the biological response that is measured. Endpoints vary with the level of biological organization being examined and include response at the subcellular level to the community level such as biomarkers (subcellular level), survival, growth, reproduction (individual level), primary production, and changes in structure (and abundance) and function in a community (population or community level). Endpoints are used in toxicity tests as criteria for effects.

▪Experimental Unit/replicate

The experimental unit is the smallest unit of experimental material to which a treatment can be allocated independently of all other units. By definition, experimental units (*e.g.* aquariums, beakers, or plant pots) must be able to receive different treatments. Each experimental unit may contain multiple sampling units (*e.g.* fish, daphnia or plants) on which measurements are taken. Within each experimental unit, sampling units may not be independent. However, in some special case situations, individual organisms (housed in common units) can be treated as the experimental units: these special cases require some proof or strong argument of independence of organisms

▪Exposure concentration, dose or dose rate

The exposure concentration, dose or dose rate is the "amount" that an organism is exposed to. For a chemical element radioactive or stable, it can be expressed as a concentration (quantity of the substance per volume or mass of the exposure source, *in mol or g or Bq per L or per g)*. For a radionuclide, the dose is the total quantity of ionising radiation absorbed by the organism (*in Gy)*; the

absorbed dose rate refers to the quantity of ionising radiation released over a specified unit of time (*e.g. µGy/h*).

▪LOEC (or LOED or LOEDR) and NOEC (or NOED or NOEDR)

The Lowest Observed Effect-Concentration is the lowest Concentration out of the tested Concentration at which a statistically significant difference from the control group is observed.

The No Observed Effect -Concentration is the tested concentration just below the LOEC. They are obtained by hypothesis testing.

The same definitions apply for Dose and Dose Rate (in place of Concentration).

▪Monotonous /non-monotonous

A monotonic concentration- or dose(rate)-effect relationship exhibits an increase or a decrease over the range of concentrations or dose(rate)s in the study. In a non-monotonic relationship, the variation in effects are not consistent across the concentrations or dose(rate)s.

▪Power

Power is the probability of rejecting the null hypothesis in favour of the alternative hypothesis, given that the alternative hypothesis is the true. Power of a test varies with sample size, variance of the measured response, the size of an effect that it is of interest to detect, and the choice of statistical test. Power to detect differences can be increased by increasing the sample size and/or reducing variation in the measured responses.

▪Quantal data

These data can exhibit two states: *eg* an individual shows an effect or not. Typically, these data are presented as the number of individuals showing the property out of a total number in the experimental unit.

▪Response

A response corresponds to an observed value of any endpoint. This term has been avoided as far as possible to avoid confusion.

▪Statistical significance

In hypothesis testing, a result is statistically significant at the chosen level α if the test statistic falls in the rejection region. The finding of statistical significance implies that the observed deviation from what was expected under the null hypothesis is unlikely to be attributable to chance variation. In general, the α-level will be 0.05 unless otherwise stated.

▪Type I and Type II errors

Type I errors (false positives) occur when the null hypothesis is the true but the hypothesis test results in a rejection of the null hypothesis in favour of the alternative hypothesis. The probability of making a Type I error is often referred to as α and is usually specified by the data analyst – often at 0.05, or 5 %. Type II errors (false negatives) occur when the alternative hypothesis is true but the test fails to reject the null hypothesis (*i.e.* there is insufficient evidence to support the alternative hypothesis). The probability of making a Type II error is often referred to as β (1 – power).

References

ASTM. 2003. Standard Guide for Conducting Laboratory Soil Toxicity or Bioaccumulation Tests with the Lumbricid Earthworm Eisenia fetida and the Enchytraeid Potworm Enchytraeus albidus. ASTM International. Report nr E1676-04.

ASTM. 2004. Standard Guide for Conducting Daphnia magna Life-Cycle Toxicity Tests. ASTM International. Report nr E1193-97 (Reapproved 2004).

Box, GEP, Hunter, WG, Hunter, JS. 1978. Statistics for experimenters: an introduction to design, data, analysis and model building. West Sussex, England: John Wiley and Sons Ltd. 672 p

Briesmeister, J. (2000). MCNP – A General Monte Carlo N-Particle Transport Code, Version 4C. *LA-13709-M (04/2000)*

Daniel, D, Garnier-Laplace, J, Gilek, M, Kautsky, U, Larsson, C, Pentreath, J, Real, A, Skarphedinsdottir, H, Sundbell-Bergman, S, Thorring, H, Woodhead, D, Zinger, I, editors. 2003. Radiation effects on plants and animals Deliverable 4: FASSET Project Contract FIGE-CT-2000-00102, Woodhead and Zinger (Eds). 196 p.

Festing, MFW. (2003). Principles: The need for better experimental design. *Trends in Pharmacological Sciences* **24(7)**:341-345

Festing, MFW, Altman, DG. (2002). Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR Journal* **43(4)**:244-258

ISO. 1998. Soil quality -- Effects of pollutants on earthworms (Eisenia fetida) -- Part 2: Determination of effects on reproduction. Geneva: International Organization for Standardization

ISO. 2000. Water quality -- Determination of long term toxicity of substances to Daphnia magna Straus (Cladocera, Crustacea). Geneva: International Organization for Standardization. Report nr ISO 10706.

Jasserand, F. (2002). EMIXG : Outil de manipulation de spectres neutron, photon et alpha, basé sur les données internationales JEF 2.2 *T MILLENNIUM 02/DSP/10.03 Ind.02.*

Jasserand, F. (2003). Dosimétrie pour l'environnement - EDEN : Formalismes de calcul des DPUC. *NT MILLENNIUM 03/026-2/10.01 ind. 02*

Kammenga, J, Laskowski, R. 2000. Demography in Ecotoxicology: Wiley. 318 p

OECD. 1998a. Daphnia reproduction test. Section 211 (Adopted 1998). OECD guidelines for the testing of chemicals. Paris: Organisation for Economic Cooperation and Development.

OECD. 1999. OECD Series on principles of good laboratory practice and compliance monitoring Number 1. Paris: Organisation for Economic Cooperation and Development. Report nr ENV/MC/CHEM(98)17, Number 1.

OECD. 2003. Draft guidance Document on the statistics analysis of ecotoxicity data. Paris: Organisation for Economic Cooperation and Development. 214 p

OECD. 2004. Earthworm reproduction test. Section 222 (adopted february 2004). OECD guidelines for testing of chemicals. Paris: Organisation for Economic Cooperation and Development.

Sparks, T. 2000. Statitistics in Ecotoxicology. West Sussex, England: John Wiley and Sons Ltd. 334 p

Thomas, L, Krebs, C. (1997). A review of statistical power analysis software. *Bull. Ecolog. Soc. America* **78(2)**:126-139

Vindimian, E. (2003). REGTOX :macro Excel™ for dose-response modelling. *http://eric.vindimian.9online.fr/en_index.html*

Wood, M, Knowles, J, Wittaker, J, Copplestone, D, Malcolm, H, Beilby, S. 2003. Developing experimental protocols for chronic irradiation studies on wildlife. R&D Tech. Report P3-101/SP2. Environment Agency, Bristol.