Environment and Climate Change Canada
Environnement et Changement climatique Canada

Biological Test Method:

Tests for Measuring Avoidance Behaviour or Reproduction of Earthworms (*Eisenia andrei* or *Dendrodrilus rubidus*) Exposed to Contaminants in Soil

Method Development and Applications Unit
Science and Technology Branch
Environment and Climate Change Canada
Ottawa, Ontario

STB 1/RM/43
Second Edition
August 2022

Canada

Cat. No.: En14-497/2022E-PDF
ISBN: 978-0-660-44285-3
EC22095

Environment and Climate Change Canada
Public Inquiries Centre
12th Floor, Fontaine Building
200 Sacré-Coeur Boulevard
Gatineau QC K1A 0H3
Telephone: 819-938-3860
Toll Free: 1-800-668-6767 (in Canada only)
Email: enviroinfo@ec.gc.ca

Aussi disponible en français

Biological Test Method:
Tests for Measuring Avoidance Behaviour or Reproduction of Earthworms (*Eisenia andrei* or *Dendrodrilus rubidus*) Exposed to Contaminants in Soil

Method Development and Applications Unit
Science and Technology Branch
Environment and Climate Change Canada
Ottawa, Ontario

STB 1/RM/43
Second Edition
August 2022

Readers' Comments

Comments regarding the content of this report should be addressed to:

Richard Scroggins, Chief
Biological Assessment and Standardization Section
Science and Technology Branch
Environment and Climate Change Canada
335 River Road
Ottawa, Ontario
K1A 0H3

Leana Van der Vliet, Manager
Method Development and Applications Unit
Biological Assessment and Standardization Section
Science and Technology Branch
Environment and Climate Change Canada
335 River Road
Ottawa, Ontario
K1A 0H3

General inquiries regarding this method can be addressed to:
methods@ec.gc.ca

Review Notice

This report has been reviewed by the staff at the Science and Technology Branch, Environment and Climate Change Canada, and approved for publication. Mention of trade names or commercial products does not constitute endorsement by Environment and Climate Change Canada for use. Other products of similar value are available.

Abstract

Revised methods now recommended by Environment and Climate Change Canada for conducting biological test methods for measuring soil toxicity using earthworms (*Eisenia andrei* and *Dendrodrilus rubidus*) are described in this report. This revised version of Report EPS 1/RM/43 includes numerous updates such as: guidance for the collection, handling and testing of soils; the removal of two test species (*Eisenia fetida* and *Lumbricus terrestris*) and the addition of another test species (*Dendrodrilus rubidus*) specifically for testing soils from the boreal or taiga ecozones; revised test designs; and updated guidance for culturing and testing, testing with a reference toxicant, and for the statistical analysis of data. In addition, due to lack of demand and lack of sensitivity relative to the reproduction test, the 14-day acute lethality test has been removed as a test option herein. This revised report supersedes the first edition of this document published as Report EPS 1/RM/43 in June 2004 and revised in June 2007. This test method document provides detailed procedures, conditions, and guidance for preparing for and conducting each of two discrete biological test methods for measuring soil toxicity using earthworms (*Eisenia andrei* or *Dendrodrilus rubidus*). The test methods described herein are as follows:

i) a 56-day test for effects on reproduction of adult earthworms exposed to one or more samples or concentrations of contaminated or potentially contaminated soil; and

ii) a 48-hour test for avoidance of contaminated soil by adult earthworms.

Each test method is conducted as a static (i.e., no renewal) test, using one or more samples of contaminated or potentially contaminated soil or one or more concentrations of chemical(s) or chemical product(s) spiked in negative control (or other) soil. Worms are fed (Magic® Worm Food or organic mixed grains) only during the reproduction test.

The 56-day reproduction test uses laboratory-cultured *E. andrei* or *D. rubidus*. This test is initiated by placing four adult worms in each of a series of glass jars (500 mL for *E. andrei* and 250 mL for *D. rubidus*) containing a measured wet weight equivalent to ~350 mL (for *E. andrei*) or 200 mL (for *D. rubidus*) of test or clean (negative control or reference) soil. For multi-concentration tests that will use regression for data analysis, a minimum of five replicate vessels are prepared for each treatment. For single-concentration tests that will use hypothesis testing for data analysis, the minimum number of replicate vessels is based on the earthworm species used and the target effect size. Following a 28-day exposure, survival for the replicate groups of adult worms in each treatment is determined. The test is continued for an additional 28 days with their progeny only. At the end of the 56-day test period, the number of live juvenile worms produced in each replicate and treatment is determined and the treatment means compared.

The avoidance test is performed as a 48-hour sublethal test using a series of circular test units constructed of stainless steel or Plexiglas™. Each test unit has a circular central chamber devoid of substrate with holes leading to each of six pie-shaped, interconnected test compartments. Three of the test compartments contain aliquots of the same sample (or concentration) of test material, and three (in alternating positions) contain aliquots of clean (i.e., negative control or reference) soil. Ten adult worms of the same species (laboratory-cultured *E. andrei* or *D. rubidus*) are placed in the central chamber upon test initiation. The number of worms in each of the test compartments is determined at the end of a 48-hour test period during which the worms in each test unit are able to distribute themselves in clean or test soil. This test uses ≥ 5 replicated test units per test soil or concentration if a single-concentration test, and ≥ 2 test units per test soil or concentration if a multi-concentration test; however, fewer test units may be used if the avoidance test is used for screening or range-finding purposes only.

General or universal conditions and procedures are outlined for test preparation and performance. Additional conditions and procedures specific to the intended use of each test are stipulated. Both of these biological test methods are suitable for measuring and assessing the toxicity of samples of field-collected soil, biosolids, sludge,

or similar particulate material; or of natural or artificial soil spiked (mixed) in the laboratory with test chemical(s) or chemical product(s). Instructions and requirements are included for test facilities, sample collection, handling and storing samples, culturing and/or acclimating test organisms, preparing soil or spiked-soil mixtures and initiating tests, specific test conditions, appropriate observations and measurements, endpoints and methods of calculation, and the use of a reference toxicant. Specific guidance for the collection, handling, and preparation of boreal forest and taiga soils and testing of these soils is also provided.

Foreword

This is one of a series of **recommended methods** for measuring and assessing the toxic effect(s) on single species of terrestrial or aquatic organisms, caused by their exposure to samples of toxic or potentially toxic substances or materials under controlled and defined laboratory conditions. Recommended methods are those that have been evaluated by Environment and Climate Change Canada (previously Environment Canada), and are favoured:

- for use in Environment and Climate Change Canada environmental toxicity laboratories;
- for testing that is contracted out by Environment and Climate Change Canada or requested from outside agencies or industry;
- in the absence of more specific instructions, such as are contained in regulations; and
- as a foundation for the provision of very explicit instructions as might be required in a regulatory *protocol* or standard *reference method*.

The different types of tests included in this series were selected because of their acceptability for the needs of environmental protection and management programs carried out by Environment and Climate Change Canada. These reports are intended to provide guidance and to facilitate the use of consistent, appropriate, and comprehensive procedures for obtaining data on the toxicity to terrestrial or aquatic life of samples of specific test substances or materials destined for or within the environment. Depending on the biological test method(s) chosen and the environmental compartment of concern, substances or materials to be tested for toxicity could include samples of chemical or chemical product, soil or similar particulate material, sediment or similar particulate material, effluent, elutriate, leachate, or receiving water. Appendix A lists the biological test methods and supporting guidance documents published to date by Environment and Climate Change Canada as part of this series.

Words defined in the Terminology section of this document are italicized when first used in the body of the report according to the definition.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Chemical Formulae

AES	atomic emission spectrophotometry
Al	aluminum
ANOVA	analysis of variance
B	boron
C	carbon
°C	degree(s) Celsius
Ca	calcium
$CaCl_2$	calcium chloride
$CaCO_3$	calcium carbonate
$Ca(OH)_2$	calcium hydroxide
CCME	Canadian Council of Ministers of the Environment
CEC	cation exchange capacity
Cl	chlorine
cm	centimetre(s)
Cu	copper
CV	coefficient of variation
d	day(s)
DQO	data quality objective
EC50	median effective concentration
ECp	effective concentration for a (specified) percent effect (e.g., EC25)
Fe	iron
g	gram(s)
H	hydrogen
h	hour(s)
H_3BO_3	boric acid
HCl	hydrochloric acid
HNO_3	nitric acid
H_2O	water
HPLC	high performance liquid chromatography
ICP	inductively coupled plasma
ICp	inhibiting concentration for a (specified) percent effect (e.g., IC25)
K	potassium
KCl	potassium chloride
kg	kilogram(s)
L	litre(s)
LC50	median lethal concentration
LCp	lethal concentration for a (specified) percent effect (e.g., LC25)
LED	light-emitting diode
LOEC	lowest-observed-effect concentration
m	metre(s)
M	mole(s) (concentration)
Mg	magnesium
mg	milligram(s)
mL	millilitre(s)
mm	millimetre(s)
Mn	manganese
Mo	molybdenum
mS	millisiemens
MW	molecular weight
MWF	Magic® Worm Food
n	sample size
N	nitrogen
Na	sodium
NH_4	ammonium
nm	nanometre(s)
NO_3	nitrate
NO_2	nitrite
NOEC	no-observed-effect concentration
O	oxygen
OM	organic matter
P	phosphorus
p	probability
PAHs	polycyclic aromatic hydrocarbons
QA/QC	quality assurance/quality control
®	Registered Trade Mark
SD	standard deviation
S	sulphur
s	second
sp.	species (singular)
spp.	species (plural)
t	time
TOC	total organic carbon
TM (™)	Trade Mark
v:m	volume-to-mass
v:v	volume-to-volume
WHC	water-holding capacity
wt	weight
Zn	zinc
α	alpha, denotes Type I error
ß	beta, denotes Type II error
µg	microgram(s)
µm	micrometre(s)
µmhos	micromhos
µmol	micromole(s)
>	greater than
<	less than
≥	greater than or equal to
≤	less than or equal to

%	percentage or percent	÷	divided by
=	equals	/	per; alternatively, “or” (e.g., holding/acclimation)
+	plus		
−	minus	≈	approximately equal to
±	plus or minus	~	approximately
×	times		

Terminology

Note: All definitions are given in the context of the procedures in this report, and might not be appropriate in another context.

Grammatical Terms

Must is used to express an absolute requirement.

Should is used to state that the specified condition or procedure is recommended and ought to be met if possible.

May is used to mean "is (are) allowed to."

Can is used to mean "is (are) able to."

Might is used to express the possibility that something could exist or happen.

Technical Terms

Acclimation is physiological adjustment to a particular level of one or more environmental factors such as temperature. The term usually refers to the adjustment to controlled laboratory conditions.

Adult (worm) is an earthworm that is sexually mature and bears an apparent *clitellum*. (See also *clitellum*, *juvenile*, and *sub-adult*.)

Amphimictic refers to reproduction involving the fertilization of an ovum by a sperm.

Biomass is the total weight (mass) of a group of animals or plants.

Canopy for the purpose of this method is the more or less continuous cover produced by the foliage of plants.

Clitellum is the fleshy "ring" or "saddle" of glandular tissue found on certain mid-body segments of *lumbricid* earthworms. It is the most visible feature of an *adult* earthworm, and is prominent only in sexually mature (i.e., *adult*) individuals. Adolescent or younger worms, which are sexually immature, are distinguished from *adults* by the absence of a *clitellum*. The *clitellum* secretes the *cocoon* into which eggs and sperm are deposited. During copulation, it also exudes mucous that envelops the anterior ends of the two individuals.

Cocoon is the protective egg case formed by the *clitellum* of earthworms, from which neonates emerge. (See also *clitellum*.)

Compliance means in accordance with governmental regulations or requirements for issuing a permit.

Culture, as a noun, means the stock of organisms raised in the laboratory under defined and controlled conditions through one or more generations, to produce healthy test organisms. As a verb, it means to carry out the procedure of raising healthy test organisms from one or more generations, under defined and controlled conditions.

Ecological risk assessment (ERA) is the process of *risk* analyses and evaluation of the adverse *effects* of *contaminated* environmental media (e.g., air, *soil*, water) on non-human organisms with respect to the nature, extent, and probability of the occurrence of these *effects* (ISO, 2005).

Electrical conductivity is a numerical expression of the ability of a solution to carry an electric current. This ability depends on the *concentrations* of ions in solution, their valence and mobility, and on the solution's temperature. For these methods, *electrical conductivity* is measured at 25 °C, and is reported as micromhos per centimetre (μmhos/cm) or as millisiemens per metre (mS/m); 1 mS/m = 10 μmhos/cm.

Epigeic refers to litter-dwelling species of earthworms that are active primarily in the detritus and feed primarily on fresh organic *material*.

Epilobic (prostomium) refers to the type of *prostomium* in which the tongue of the earthworm partly divides the *peristomium*. (See also *prostomium* and *peristomium*.)

Genital tumescences refer to areas of modified epidermis (i.e., glandular swellings) on the body of an earthworm without distinct boundaries and through which follicles of genital *setae* open (Reynolds, 1977).

Growth is the increase in size or weight as the result of proliferation of new tissues. In this test, it refers to an increase in dry weight.

Hatchling is an earthworm that has recently emerged from a *cocoon*, and has begun to actively feed and grow. (See also *juvenile*.)

Hormesis is an observed stimulation of performance (e.g., reproduction) among test organisms, compared with the *control* organisms, at low *concentrations* in a *toxicity test*.

Juvenile (worm) is an earthworm that is sexually immature and lacks an apparent *clitellum*. This classification can include *hatchlings* (e.g., worms within 48 hours of emergence from *cocoons*), as well as all other sexually immature life stages for hatched earthworms up to and including *sub-adults*. (See also *adult*, *clitellum*, *hatchling*, and *sub-adult*.)

L, F, and H layers refer to the combined LFH layer of a *soil*. This is an organic layer that occurs on the surface of the mineral *soil*, and is usually composed of the accumulation of leaves, twigs, and woody *materials*. The components of the L (leaf) layer, which is at the top, are usually identifiable. The next layer down (F) is distinguished by the original *materials* being difficult to identify as a result of the initiation of decomposition, while the H layer is composed of decomposed organic *materials* that are indiscernible. The H layer might be intermixed with mineral particles from the mineral *soil* below.

Light-emitting diode (LED) is a type of light source. It is a semi-conductor diode that glows when a voltage is applied. *LED* differs from fluorescent and incandescent light sources in the mechanism used to generate light.

Lumbricid refers to earthworms that are members of the family Lumbricidae, superfamily Lumbricoidea, order Haplotaxida, subclass Oligochaeta, class Clitellata, and phylum Annelida.

Lux is a unit of illumination based on units per square metre. One *lux* = 0.0929 foot-candles and one foot-candle = 10.76 *lux*. For conversion of *lux* to quantal flux [μmol/(m^2 · s)], the spectral quality of the light source must be known. Light conditions or irradiance are properly described in terms of quantal flux (photon fluence rate) in the photosynthetically effective wavelength range of approximately 400–700 nm. The relationship between quantal flux and *lux* or foot-candles is highly variable and depends on the light source, the light meter used, the geometrical arrangement, and the possibilities of reflections (see ASTM, 2014). Approximate conversions between quantal flux and *lux*, however, are:

- for cool-white fluorescent light: 1 *lux* ≈ 0.014 μmol/(m^2 · s);
- for full-spectrum fluorescent light (e.g., Vita-Lite® by Duro-Test®): 1 *lux* ≈ 0.016 μmol/(m^2 · s); and
- for incandescent light: 1 *lux* ≈ 0.019 μmol/(m^2 · s) (Deitzer, 1994; Sager and McFarlane, 1997).

Monitoring is the routine (e.g., daily, weekly, monthly, quarterly) checking of quality or collection and reporting of information. In the context of this report, it means either the periodic (routine) checking and measurement of certain biological or *soil* quality variables, or the collection and testing of *soil* samples for *toxicity*.

Papillae refer to protruding dermal structures on the external body of an earthworm (Reynolds, 1977).

Peristomium refers to the first body segment of an earthworm, which is without *setae* and contains the mouth. (See also *seta*.)

pH is the negative logarithm of the activity of hydrogen ions in gram equivalents per litre. The *pH* value expresses the degree or intensity of both acidic and alkaline reactions on a scale from 0 to 14, with 7 representing neutrality, numbers < 7 indicating increasingly greater acidic reactions, and numbers > 7 indicating increasingly basic or alkaline reactions.

Photoperiod is the duration of illumination and darkness within a 24-h period.

Pollution is the addition of a *substance* or *material*, or a form of energy such as heat, to some component of the environment, in such an amount as to cause a discernible change that is deleterious to some organism(s) or to some human use of the environment. Some national and international agencies have formal definitions of *pollution*, which should be honoured in the appropriate contexts.

Pretreatment means treatment of a sample of *soil*, or portion thereof, before exposure of the test organisms.

Progeny means the young or offspring (i.e., immediate descendants) of a sexually mature (*adult*) earthworm.

Prostomium refers to the anterior lobe projecting in front of the *peristomium* and above the mouth. (See also *peristomium*.)

Protocol is an explicit set of procedures for a test, formally agreed upon by the parties involved, and described precisely in a written document.

Quality assurance (QA) is a program within a laboratory, intended to provide precise and accurate results in scientific and technical work. It includes selection of proper procedures, sample collection, selection of limits, evaluation of data, *quality control*, and qualifications and training of personnel.

Quality control (QC) consists of specific actions within the program of *quality assurance*. It includes standardization, calibration, replication, *control* samples, and statistical estimates of limits for the data.

Redox potential (also known as the oxidation-reduction potential) is a measure (in volts) of the affinity of a *substance* for electrons relative to hydrogen.

Reference method refers to a specific *protocol* for performing a *toxicity test*, i.e., a biological test method with an explicit set of test procedures and conditions, formally agreed upon by the parties involved, and described precisely in a written document. Unlike other multipurpose (generic) biological test methods published by Environment and Climate Change Canada, the use of a *reference method* is frequently restricted to testing requirements associated with specific regulations.

Remediation is the management of a *contaminated site* to prevent, minimize, or mitigate damage to human health or the environment. *Remediation* can include both direct physical actions (e.g., removal, destruction, and containment of *toxic substances*) and institutional controls (e.g., zoning designations or orders).

Risk is the probability or likelihood that an adverse *effect* will occur.

Risk assessment – see *ecological risk assessment*.

Seminal vesicles refer to the storage sacs for an earthworm's own sperm, until copulation.

Septa refer here to internal partitions at intersegmental furrows along the outer body of an earthworm. *Septa* also act as supporting membranes for internal organs (Reynolds, 1977).

Seta refers to a solid rod or bristle secreted by cells at the internal end of a tubular epidermal ingrowth referred to as the setal follicle. *Setae* are of various types (e.g., general, genital, or penial; see Reynolds [1977] for further description).

Spermathecae refer to the pouches developed in the *septa* of an earthworm, which receive sperm from another individual during copulation; the sperm are stored here until the period of *cocoon* laying (Reynolds, 1977). (See also *septa*.)

Sub-adult (worm) is a *juvenile* "adolescent" earthworm that is sexually immature and lacks an apparent *clitellum*. (See also *adult, clitellum,* and *juvenile*.)

Tubercula pubertatis refers to a glandular swelling on an earthworm, which appears near the ventrolateral margins of the *clitellum*. It is not always present, and it might be continuous or discontinuous and of varied size and shape (Reynolds, 1977).

Terms for Test Materials or Substances

Artificial soil is a laboratory-formulated *soil*, prepared to simulate a natural *soil* using a specific ratio of natural constituents of sand, clay, and peat. *Artificial soil* may be used as a *negative control soil*, and as a diluent to prepare multiple *concentrations* of *site soil(s)* or *chemical-spiked soil(s)*.

Batch means the total amount of a particular *test soil* (or specific *concentration* thereof) prepared for each *treatment* (*concentration*) in a test. A *batch* is any hydrated *test soil* ready for separation into *replicates*. A *batch* might also refer to a single group of worms received from a source outside the laboratory at a discrete time.

Bulk soil samples are *unconsolidated*, typically large (>1 L) *point samples* that consist of more than one individual block of *soil* removed from one sample location by a sampling device, and therefore are *point samples*, not *composite samples*. *Bulk soil samples* are often collected to satisfy the large volume requirements for biological testing. (See also *point sample[s]* and *composite sample[s]*.)

Cation exchange capacity (CEC) is the sum total of exchangeable cations that a *soil* can adsorb. It is sometimes called total-exchange capacity, base-exchange capacity, or cation-adsorption capacity. It is expressed in milliequivalents per 100 grams of *soil* (or other adsorbing *material* such as clay) (AAFC, 1998).

Chemical is, in this report, any element, compound, formulation, or mixture of a *substance* that might be mixed with, deposited in, or found in association with *soil* or water, or that might enter the environment through spillage, application, or discharge.

Chemical-spiked soil is natural or *artificial soil* (usually *negative control soil*, *reference soil*, or other *clean soil*) to which one or more *chemicals* or *chemical products* have been added, and mixed thoroughly to evenly distribute the *substance(s)* throughout the *soil* at a specific *concentration* to form a *batch* for use in a *soil toxicity test*. (See also *spiked soil*.)

Clean soil is *soil* that does not contain *concentrations* of any *substance(s)* or *material(s)* causing discernible *toxic effects* to the test organisms.

Composite sample(s) are *soil* samples consisting of *point* or *bulk soil samples* combined from two or more sample locations at a *site* (Crépin and Johnson, 1993).

Concentration means the ratio of the weight of a test *substance* or *material* to the weight of *soil*, and is frequently expressed as the weight of test *substance* or *material* per kg of dry *soil* (mg/kg). *Concentration* might also be expressed as a percentage of the test *substance* or *material* (e.g., *contaminated site soil*) per dry weight of *soil*.

Consolidated sample is synonymous with undisturbed sample and is a sample obtained from *soil* using a method designed to preserve the *soil* structure (ISO, 2005). (See also *unconsolidated sample.*)

Contaminant is a *substance* or *material* that is present in a natural system, or present at an increased *concentration*, often because of some direct or indirect human activity. The term is frequently applied to *substances* or *materials* present at *concentrations* that have the potential to cause adverse biological *effects*.

Contaminated (soil) means (*soil*) containing *chemical substances* or *materials* at *concentrations* that pose a known or potential threat to environmental or human health.

Control is a *treatment* in an investigation or study that duplicates all the conditions and factors that might affect results, except the specific condition being studied. In *toxicity tests*, the *control* must duplicate all the conditions of the exposure *treatment(s)*, but must contain no *contaminated* test *material* or *substance*. The *control* is used as a check for the absence of measureable *toxicity* due to basic test conditions such as temperature, health of test organisms, or *effects* due to their handling. *Control* is synonymous with *negative control*, unless indicated otherwise.

Control soil – see *negative control soil*.

Core is a sample of *soil* that has been collected using a corer.

Data quality objectives (DQOs) are predefined criteria for the quality of data generated or used in a particular study so as to ensure that the data are of acceptable quality to meet the needs for which they were collected.

Definitive (soil toxicity test) means decisive (as opposed to a preliminary, *range-finding test*). (See also *range-finding [test]*.)

Deionized water is water that has been purified by passing it through resin columns or a reverse osmosis system, for the purpose of removing ions such as Ca^{++} and Mg^{++}.

Distilled water is water that has been passed through a distillation apparatus of borosilicate glass or other *material*, to remove impurities.

Fertility (of soil) refers to the potential of a *soil* to supply nutrient elements in the amounts, forms, and proportions required for optimal plant *growth*. *Soil fertility* is measured directly in terms of the ions and compounds important for plant nutrition. The fundamental components of *fertility* are the essential nutrients

(macronutrients including C, H, O, N, P, K, Ca, Mg, S and micronutrients including Fe, Mn, Mo, B, Cu, Zn, and Cl). Indirectly, *soil fertility* is measured by demonstrating its productivity (i.e., the capacity of the *soil* to produce plants that supply essential food and fibre; Hausenbuiller, 1985).

Horizon – see *soil horizon*.

Hydration water means water used to hydrate *test soils*, to create a specific *moisture content* suitable for the test organisms. The water used for hydration is normally *test water*, and is frequently *deionized* or *distilled water*, reverse osmosis water, or dechlorinated tap water. Depending on study design and intent, a surface water or groundwater from the *site* might be used instead of *deionized* or *distilled water* for the hydration of each *test soil* (including *negative control soil*). (See also *test water*, *deionized water*, and *distilled water*.)

Material is the *substance* or *substances* from which something is made. A *material* would have more or less uniform characteristics. *Soil*, sediment, or surface water are *materials*. Usually, the *material* would contain several or many *substances.*

Moisture content is the percentage of water in a sample of *test soil*, based on its wet or dry mass. It is determined by measuring both the wet and dry weights of a subsample of the *soil*. The *soil's moisture content* is then calculated and expressed on a dry-weight basis, by dividing the mass of water in the subsample (wet mass – dry mass) by the mass of dry *soil*, and then multiplying by 100. Units for mass (i.e., g or mg) must be the same in each instance.

Negative control – see *control*.

Negative control soil is *clean soil* that does not contain *concentrations* of one or more *contaminants* that could affect the survival, reproduction, or behaviour (i.e., avoidance) of the test organisms. *Negative control soil* might be natural *soil* from an uncontaminated *site*, or *artificial* (formulated) *soil*. This *soil* must contain no added test *material* or *substance*, and must enable acceptable survival and performance of the test organisms during the test. The use of *negative control soil* provides a basis for interpreting data derived from *toxicity tests* using *test soil(s)* and gives information about the state of health (i.e., quality) of the test organisms coming from a *culture*.

Organic matter (OM) in *soil* consists primarily of plant and animal residues, at different stages of decomposition, including *soil* humus. The accumulation of OM within *soil* is a balance between the return or addition of plant and animal residues and their subsequent loss due to the decay of these residues by *soil* microorganisms. For many types of *soil*, the following equation (AESA, 2001) is suitable for estimating the total OM content of *soil* from *total organic carbon (TOC)* measurements: % OM = % TOC × 1.78; however, the relationship between *TOC* and *OM* is slightly different among *soils*, and therefore the *total organic carbon* content should also be determined by laboratory analysis. (See also *total organic carbon [TOC]*.)

Point sample(s) are individual blocks of *soil* removed from one sample location by a sampling device (e.g., *soil corer*).

Positive control soil is *contaminated soil* that contains *concentrations* of one or more *contaminants* that adversely affect the reproduction or behaviour (i.e., avoidance) of the test organisms using the biological test methods defined herein. *Positive control soil* might be used as a *reference toxicant* to assess the sensitivity of the test organisms at the time the test *material* or *substance* is evaluated, and to determine the *precision* of results obtained by the laboratory for that *reference toxicant*.

Product is a commercial formulation of one or more *chemicals*. (See also *chemical*.)

Range-finding (test) means a preliminary *soil toxicity test*, performed to provide an initial indication of the *toxicity* of the test *material* under defined conditions and to assist in choosing the range of *concentrations* to be used in a *definitive* multi-concentration test. (See also *definitive [soil toxicity test]*.)

Reference site is a *site* uninfluenced by source(s) of contamination but within the general vicinity of the *sites* where samples of *test soil* are collected. (See also *site*.)

Reference soil is typically *clean* field-collected *soil* or formulated (*artificial*) *soil* that is selected for use in a particular *toxicity test* together with a *negative control soil* and one or more samples of *test soil*. *Reference soil* used in a test frequently exhibits physicochemical properties (e.g., *texture*, *organic matter* content, *total organic carbon* content, *pH*, and *electrical conductivity*) closely matching those of the *test soil* sample(s), except that it is free from the source of contamination being assessed. In tests involving samples of *site soil*, one or more samples of *reference soil* are often selected from the general location of *test soil* sampling, and thus might be subject to other sources of contamination aside from the one(s) being studied. *Reference soil* is used to describe matrix *effects* in the test, and may also be used as a diluent to prepare *concentrations* of the *test soil*. In tests involving *chemical-spiked soil*, one or more samples of *artificial* (formulated) *soil* with differing physicochemical characteristics might be chosen to investigate the influence of certain *soil* properties (e.g., *soil texture*, or percent *organic matter*) on the *toxicity* of a *chemical* mixed in each of these *soil* types. (See also *negative control soil*, *site soil*, *test soil*, *clean*, *artificial soil*, and *chemical-spiked soil*.)

Reference toxicant is a standard *chemical* used to measure the sensitivity of the test organisms to establish confidence in the *toxicity* data obtained for a test *material* or *substance*. In most instances, a multi-concentration *toxicity test* with a *reference toxicant* or a positive control *concentration* prepared using a *reference toxicant* is used to assess the sensitivity of the organisms at the time the test *material* or *substance* is evaluated, and the *precision* and reliability of results obtained by the laboratory for that *chemical*.

Reference toxicity test is a multi-concentration test conducted using a *reference toxicant* in conjunction with a *soil toxicity test*, to appraise the sensitivity of the organisms and the *precision* and reliability of results obtained by the laboratory for that *chemical* at the time the test *material* or *substance* is evaluated. Deviations outside an established normal range indicate that the sensitivity of the test organisms, and the performance and *precision* of the test, are suspect and should be investigated as to the cause. A *reference toxicity test* with earthworms is performed as a *spiked-soil* test, using a standard *chemical*.

Sampling location means a specific location, within a *site*, where the sample(s) of field-collected *soil* are obtained for *toxicity tests* and associated physicochemical analyses (and is considered the same as a sampling station).

Site means a delineated tract of land that is being used or considered as a study area, usually from the perspective of it being *contaminated* or potentially *contaminated* by human activity. (See also *reference site*.)

Site soil is a field-collected sample of *soil*, taken from a location (i.e., a *site*) thought to be *contaminated* with one or more *chemicals*, and intended for use in the *toxicity test* with earthworms. In some instances, the term includes *reference soil* or *negative control soil* from a *reference site*.

Soil is whole, intact *material* representative of the terrestrial environment, that has had minimal manipulation following collection or formulation. In the natural environment, it is formed by the physical, chemical, and biological weathering of rocks and the decomposition and recycling of nutrients from *organic matter* originating from plant and animal life. Its physicochemical characteristics are influenced by biological activities (e.g., microbial, invertebrate [including earthworm], and plant) and abiotic factors therein, and by anthropogenic activities.

Soil horizon is a layer of mineral or organic *soil material* approximately parallel to the land surface, which has characteristics altered by processes of *soil* formation. It differs from adjacent *horizons* in properties such as colour, structure, *texture*, and consistency and in *chemical*, biological, or mineralogical composition.

Solvent control soil is a sample of (usually *artificial*) *soil* included in a test involving *chemical-spiked soil*, in which an organic solvent is required to solubilize the test *chemical* before mixing it in a measured quantity of *negative control soil*. The amount of solvent used when preparing the *solvent control soil* must contain the same *concentration* of solubilizing agent as that present in the highest *concentration* of the test *chemical(s)* in the sample of *chemical-spiked soil* to be tested. This *concentration* of solvent should not adversely affect the performance of earthworms during the test. Any test that uses an organic solvent when preparing one or more *concentrations* of *chemical-spiked soil* must include a *solvent control soil* in the test. (See also *artificial soil*, *negative control soil*, and *chemical-spiked soil*.)

Spiked soil is natural or *artificial soil* (usually *negative control soil*, *reference soil*, or other *clean soil*) to which one or more *chemicals*, *chemical products*, or other test *substances* or *materials* (e.g., a sample of sludge or drilling mud) have been added in the laboratory, and mixed thoroughly to evenly distribute the *substance(s)* or *material(s)* throughout the *soil* at a specific *concentration* to form a *batch* for use in a *soil toxicity test*. (See also *chemical-spiked soil* and *spiking*.)

Spiking refers to the addition of a known amount of *chemical(s)*, *chemical product(s)*, or other test *substance(s)* or *material(s)* (e.g., a sample of sludge or drilling mud) to a natural or *artificial soil*. The *substance(s)* or *material(s)* is (are) usually added to *negative control soil*, *reference soil*, or another *clean soil*, but sometimes to a *contaminated* or potentially *contaminated soil*. After the addition ("*spiking*"), the *soil* is mixed thoroughly. If the added test *material* is a *site soil*, Environment and Climate Change Canada documents typically do not call this *spiking*, but instead refer to the manipulation as "dilution," "amendment," or simply "addition." (See also *chemical-spiked soil* and *spiked soil*.)

Stock solution means a concentrated solution of the *substance(s)* to be tested, followed by the addition of a measured quantity of this solution to a sample of natural or *artificial soil* and thorough mixing to prepare a *batch* of *chemical-spiked soil*. To prepare the required strength of the *stock solution*, measured weights or volumes of test *chemical(s)* or *chemical product(s)* are added to *test water* (*deionized* or *distilled water*, or equivalent), with or without the inclusion of an organic solvent.

Substance is a particular kind of *material* having more or less uniform properties. The word *substance* has a narrower scope than *material*, and might refer to a particular *chemical* (e.g., an element) or *chemical product*.

Test soil is a sample of field-collected *soil* (e.g., *site soil*) that is *contaminated* or potentially so, or *chemical-spiked soil* that is to be evaluated for *toxicity* to earthworms. Boreal and taiga *test soils* are collected as separate *soil horizons*. In some instances, the term also applies to any solid-phase sample or mixture thereof (e.g., *negative control soil*, *positive control soil*, *reference soil*, sludge, drilling mud) used in a *soil toxicity test*.

Test water is water used to prepare *stock solutions*, rinse test organisms, or rinse glassware and other apparatus used for culturing earthworms and for other purposes associated with the biological test method (e.g., to hydrate samples of *test soil*). *Test water* must be *deionized* or *distilled water* or better (e.g., reagent-grade water produced by a system of reverse osmosis, carbon, and ion exchange cartridges). (See also *hydration water*.)

Texture is defined based on a measurement of the percentage by weight of sand, silt, and clay in the mineral fraction of *soils*. Classification as to *texture* confers information on the general character and behaviour of *substances* in *soils*, especially when coupled with information on the structural state and *organic matter*

content of the *soil*. *Texture* in the context of this guidance document is described according to the Canadian System of Soil Classification (AAFC, 1998), not the Unified Soil Classification, the United States Soil Conservation Service Classification, or any other *soil* classification system used for *soil* science, engineering, or geology. *Soil texture* is determined in the laboratory by measuring the particle size distribution using a two-step procedure whereby the sand particles (coarse fragments) are initially separated by sieving from the silt and clay particles; followed by separation of the silt and clay particles by their sedimentation in water. Textural classification systems typically refer to groupings of *soil* based on specific ranges in relative quantities of sand, silt, and clay. There are three main textural classes:

i) coarse *texture* (sands, loamy sands, and sandy loams);
ii) medium *texture* (loams, silt loams, silts, and very fine sandy loams); and
iii) fine *texture* (clays, silty clay loams, sandy clay loams, silty clays, and sandy clays).

Further distinction as to *texture* (e.g., "sandy clay," "silt loam," "loam") can be made based on the Canadian classification schemes using the relative amounts of percent sand, percent silt, and percent clay in the *soil* (AAFC, 1998).

Total organic carbon (TOC) refers to the organic carbon content of *soil* exclusive of carbon from undecayed plant and animal residues, as determined by dry combustion analysis (ISO, 1995). (See also *organic matter*.)

Unconsolidated sample is synonymous with disturbed sample and is a sample obtained from *soil* without any attempt to preserve the *soil* structure (ISO, 2005). (See also *consolidated sample*.)

Water-holding capacity (WHC) refers to the maximum quantity of water that a *soil* can retain, following complete saturation. It is usually determined gravimetrically, and is generally expressed as the percentage of water (by mass; water weight:dry *soil* weight) retained in a sample of *soil* that has been saturated with water.

Statistical and Toxicological Terms

A priori literally refers to something that is independent of experience. In the context of test design and statistics, *a priori* tests are ones that have been planned before the data were collected. Test objectives and test design would influence the decisions on which *a priori* tests to select.

Acute means within a short period of exposure (seconds, minutes, hours, or a few days) in relation to the lifespan of the test organism and is generally used to describe the length of a test or exposure duration.

Acute toxicity is a discernible adverse *effect* (*lethal* or *sublethal*) induced in the test organisms within a short period (usually a few days) of exposure to *test soil(s).*

Battery of toxicity tests is a combination of several *toxicity tests*, normally using different species of test organisms (e.g., a series of *soil toxicity tests* using earthworms, plants, or springtails), different biological *endpoints* (e.g., *lethal* and various *sublethal*), and different durations of exposure (e.g., *acute* and *chronic*).

Bioassay is a test (= assay) in which the strength or potency of a *substance* is measured by the response of living organisms. In standard pharmacological usage, a *bioassay* assesses the unknown potency of a given preparation of a drug, compared to the known potency of a standard preparation. *Toxicity test* is a more specific and preferred term for environmental studies.

Chronic means occurring within a relatively long period of exposure (weeks, months, or years), usually a significant portion of the lifespan of the organism, and is generally used to describe the length of a test or exposure duration.

Chronic toxicity refers to discernable adverse *effects* observed during or after relatively long-term exposures to one or more *contaminants*, which are related to changes in reproduction, *growth*, metabolism, ability to survive, or other biological variables (e.g., behaviour) being observed.

Coefficient of variation (CV) is the standard deviation (SD) of a set of data divided by the mean of the data set, expressed as a percentage. It is calculated according to the following formula: CV (%) = 100 × (SD ÷ mean).

EC50 is the *median effective concentration*, i.e., the *concentration* (e.g., % or mg/kg) of *substance(s)* or *material(s)* in *soil* that is estimated to cause some defined *toxic effect* on 50% of the test organisms. The *EC50* and its 95% confidence limits are usually derived by statistical analysis of the percentages of organisms affected (e.g., showing an avoidance response) in five or more test *concentrations*, after a fixed period of exposure. The duration of exposure must be specified (e.g., 48 hours). The *EC50* describes *sublethal quantal effects* (e.g., *effects* with binomial responses such as avoidance or no avoidance) and is not applicable to continuous *quantitative effects* (e.g., *effects* that can be measured along a numerical continuum such as number of *juveniles* or weight) (see *ICp*). Depending on the study objectives, an ECp other than *EC50* (e.g., an EC25) might be calculated instead of or in addition to the *EC50*.

Effect, in *toxicology*, means a measurable biological change. The change could be structural, physiological, behavioural, etc. In a *toxicity test*, the biological change should be assessed against a background of measurements on organisms in *control* conditions. The statistical analysis generally considers the degrees of *effect* that are beyond the *control* measurements, and are therefore presumed to result from exposure to *toxic* components of the *material* being tested.

Endpoint means the response of the test organisms that is measured (e.g., *adult* death, number of *progeny* produced, or avoidance response), or the value(s) that characterize the results of a test (e.g., *EC50*, *LC50*, IC25).

Environmental toxicology (or *ecotoxicology*) is a branch of *toxicology* with the same general definition. However, the focus is on ecosystems, natural communities, and wild living species, without excluding humans as part of the ecosystems.

Geometric mean is the mean of repeated measurements, calculated logarithmically. It has the advantage that extreme values do not have as great an influence on the mean as is the case for an arithmetic mean. The *geometric mean* can be calculated as the n^{th} root of the product of the "n" values, and it can also be calculated as the antilogarithm of the mean of the logarithms of the "n" values.

Heteroscedasticity refers herein to data showing heterogeneity of the *residuals* within a scatter plot (see EC, 2005a). This term applies when the variability of the *residuals* changes significantly with that of the independent variables (i.e., the test *concentrations* or *treatment* levels). When performing statistical analyses and assessing *residuals* (e.g., using Levene's test), for test data demonstrating *heteroscedasticity* (i.e., non-homogeneity of *residuals*), there is a significant difference in the variance of *residuals* across *concentrations* or *treatment* levels. (See also *homoscedasticity* and *residual*.)

Homoscedasticity refers herein to data showing homogeneity of the *residuals* within a scatter plot (see EC, 2005a). This term applies when the variability of the *residuals* does not change significantly with that of the independent variables (i.e., the test *concentrations* or *treatment* levels). When performing statistical analyses and assessing *residuals* (e.g., using Levene's test), for test data demonstrating *homoscedasticity* (i.e., homogeneity of *residuals*), there is no significant difference in the variance of *residuals* across *concentrations* or *treatment* levels. (See also *heteroscedasticity* and *residual*.)

ICp is the inhibiting *concentration* for a (specified) percent *effect*. It represents a point estimate of the *concentration* of test *substance* or *material* that causes a designated percent inhibition ("p") compared with the *control*, in a *quantitative* (continuous) biological measurement such as number of *progeny* produced by individuals at the end of the test (e.g., IC25 or IC50).

LC50 is the *median lethal concentration*, i.e., the *concentration* (e.g., % or mg/kg) of *substance(s)* or *material(s)* in *soil* that is estimated to be *lethal* to 50% of the test organisms. The *LC50* and its 95% confidence limits are usually derived by statistical analysis of percent mortalities in five or more test *concentrations* after a fixed period of exposure. The duration of exposure must be specified (e.g., 28-day *LC50*). Depending on the study objectives, an LCp other than *LC50* (e.g., an LC25) might be calculated instead of or in addition to the *LC50*.

Lethal means causing death by direct action. Death of test organisms is defined as the cessation of all visible signs of movement or other activity indicating life.

LOEC is the *lowest-observed-effect concentration*. This is the lowest *concentration* of a test *substance* or *material* for which a statistically significant adverse *effect* on the test organisms was observed, relative to the *control*.

NOEC is the *no-observed-effect concentration*. This is the highest *concentration* of a test *substance* or *material* at which no statistically significant adverse *effect* on the test organisms was observed, relative to the *control*.

Normality (or *normal distribution*) refers to a symmetric, bell-shaped array of observations. The array relates frequency of occurrence to the magnitude of the item being measured. In a *normal distribution*, most observations will cluster near the mean value, with progressively fewer observations toward the extremes of the range of values. The normal distribution plays a central role in statistical theory because of its mathematical properties. It is also central in biological sciences because many biological phenomena follow the same pattern. Many statistical tests assume that data are normally distributed, and therefore it can be necessary to test whether that is true for a given set of data.

Power is the probability of correctly concluding that there is a statistically significant difference between the variables being tested. By definition, it is "the probability of rejecting the null hypothesis when it is in fact false and should be rejected." In effect, it is the inverse of making a *Type II error*, in which an investigator accepts the null hypothesis when there is actually a difference. The probability of making that *Type II erro*r is called ß, and *power* is represented by (1 – ß). *Power* cannot be directly and precisely set by the investigator before doing a *toxicity test*. *Power* can be increased, however, by optimizing the *toxicity test* design (more organisms, more *replicates*, reducing variability, etc.).

Precision refers to the closeness of repeated measurements of the same quantity to each other, i.e., the degree to which data generated from replicate measurements are the same. It describes the degree of certainty around a result, or the tightness of a statistically derived *endpoint* such as an *ICp*.

Quantal effects in a *toxicity test* are those in which each test organism responds or does not respond. For example, an animal might respond by dying in or avoiding a *contaminated test soil*. Generally, *quantal effects* are expressed as numerical counts or percentages thereof. (See also *quantitative*.)

Quantitative effects in a *toxicity test* are those in which the measured *effect* is continuously variable on a numerical scale. Examples would be number of *progeny* produced at test end. Generally, *quantitative effects* are determined and expressed as measurements. (See also *quantal*.)

Replicate (treatment, test vessel, or *test unit)* refers to a single test vessel containing a prescribed number of organisms in either one *concentration* of the test *material* or *substance*, or in the *control* or reference *treatment(s)*. A *replicate* of a *treatment* must be an independent test vessel; therefore, any transfer of

organisms or test *material* from one test vessel to another would invalidate a statistical analysis based on the replication (see Sections 5.1 and 5.6.1 herein, and Section 2.5 of EC, 2005a). For the avoidance test described herein, the avoidance test unit is considered a single replicate.

Replicate samples are field-replicated samples of *soil* collected independently from the same *sampling location*, to provide an estimate of the sampling error or to improve the *precision* of estimation. A single *soil* sample from a *sampling location* is treated as one *replicate*. Additional samples are considered to be additional *replicate samples* when they are treated identically (regardless of whether they are *point* or *composite samples* from the same location), but stored in separate sample containers (i.e., not composited or, if already *composite samples*, not composited further).

Residual, in the context of Section 6.4.2.1, refers to the difference between the predicted estimate (based on the model) and the actual value observed, as determined by subtracting the former from the latter. (See also *heteroscedasticity* and *homoscedasticity*.)

Static describes a *toxicity test* in which the *test soil* (or any *chemical* or *chemical product* therein) is not renewed or replaced during the test.

Sublethal (toxicity) means detrimental to the organism, but below the *concentration* or level of contamination that directly causes death within the test period.

Sublethal effect is an adverse *effect* on an organism resulting from exposure to the *concentration* or level of contamination below that which directly causes death within the test period.

Target effect size is the magnitude of adverse *effect* in a particular study that is deemed to be important, expressed as the percent reduction from the *control*. In this test method, the *effect* refers particularly to a reduction in number of *progeny*. The *target effect size* can be linked to a policy statement, decided based on expert judgement, chosen to align with other *effect* sizes in a *battery of toxicity tests*, or derived through other means. The *target effect size* is selected before testing begins. Note that selecting a *target effect size* does not imply that adverse *effects* will be observed in a particular test; the selection of *target effect size* only links the number of *replicates* with the ability of the test to "detect" (in terms of statistical significance) an *effect*, if it does exist.

Toxic means poisonous. A *toxic chemical* or *material* can cause adverse *effects* on living organisms if present in sufficient amount at the right location (i.e., receptor/organ). *Toxic* is an adjective and, in some situations, a noun (usually found in the plural). In this context, *toxicant* is the better choice for the noun.

Toxicant is a *toxic substance* or *material*.

Toxicity is the inherent potential or capacity of a *substance* or *material* to cause adverse *effect(s)* on living organisms. These *effect(s)* could result from exposure to either *lethal* or *sublethal concentrations* of *contaminants* in *soil*.

Toxicity test is a determination of the adverse *effect(s)* of a *substance* or *material* that results from exposure of a group of selected organisms of a particular species (e.g., *Eisenia andrei* or *Dendrodrilus rubidus*), under defined conditions. A *toxicity test* involving samples of *test soil* usually measures (a) the proportions of organisms affected (*quantal*), and/or (b) the degree of *effect* observed (*quantitative* or graded), after exposure of the test organisms to the whole sample (e.g., undiluted *site soil*) or specific *concentrations* thereof.

Toxicology is a branch of science that studies the *toxicity* of *substances*, *materials*, or conditions. There is no limitation on the use of various scientific disciplines, field or laboratory tools, or studies at various levels of

organization, whether molecular, single species, populations, or communities. Applied *toxicology* would normally have a goal of defining the limits of safety of *chemical* or other agents. (See also *environmental toxicology*.)

Treatment refers to a specific *test soil* (e.g., a *site soil*, *reference soil*, or *negative control soil*) from a particular *sampling location*, or a *concentration* of *chemical-spiked soil* (or a mixture of *test soil* diluted with *clean soil*) prepared in the laboratory. *Test soils* representing a particular *treatment* are typically replicated in a *toxicity test*. (See also *replicate* and *replicate samples*.)

Type I error, commonly designated as α (alpha), occurs when an investigator rejects a null hypothesis that is true. In other words, the investigator concludes that there is a significant difference, when there is in fact none.

Type II error, commonly designated as ß (beta), occurs when an investigator fails to reject the null hypothesis when it is false (i.e., concludes that there is no significant difference, when there is in fact one).

Warning chart is a graph used to follow changes over time in the *endpoints* for a *reference toxicant*. The date of the test or test number is on the horizontal axis. For multi-concentration tests, the effect concentration is plotted on the vertical logarithmic scale, whereas for positive controls, the percent *effect* relative to the *control* is plotted on the vertical arithmetic scale.

Warning limit is plus or minus two standard deviations of the mean from tests with a *reference toxicant*. For multi-concentration tests, a *warning limit* is calculated logarithmically from a historical *geometric mean* of the *endpoints* (i.e., IC50), whereas for positive controls, a *warning limit* is calculated arithmetically from a historical mean of *endpoints* (i.e., % *effect* relative to *control*).

Acknowledgements

The first edition of this biological test method document, published in June 2004, was co-authored by D.J. McLeay (McLeay Environmental Ltd., Victoria, BC) and G.L. Stephenson (Aquaterra Environmental, Orton, ON). J.A. Miller (Miller Environmental Sciences Inc., Innisfil, ON) and J.I. Princz (Biological Methods Division, Environmental Technology Centre, EC, Ottawa, ON) are sincerely thanked for their contributions to certain sections of the first edition of this document. Input from J.B. Sprague (Sprague Associates Ltd., Salt Spring Island, BC) with respect to certain definitions in the Terminology section and the statistical guidance herein is gratefully acknowledged. B.A. Zajdlik (Zajdlik & Associates, Rockwood, ON) is thanked for his input to the statistical guidance on regression analyses. N.C. Feisthauer (Stantec Consulting Ltd., Guelph, ON) and J. McCann (University of Waterloo, Waterloo, ON) contributed much useful technical guidance and specific advice included herein.

R.P. Scroggins (Biological Methods Division, Environmental Technology Centre, EC, Ottawa, ON) acted as Scientific Authority for the first edition of this method document and provided technical input and direction throughout the work. The studies resulting in the biological test methods defined in the first edition of this method document were developed by G.L. Stephenson, as part of her doctoral thesis. Additional studies were directed by Ms. Stephenson, and managed and performed by N.C. Feisthauer and co-workers with Stantec Consulting Ltd. (formerly ESG International Inc.) (Guelph, ON). Laboratory assistance was also provided by many undergraduate and graduate students (University of Guelph). R. Pandey of Guelph Chemical Laboratories Ltd. (Guelph, ON) and V. Marsielle-Kerslake of Analytical Services (University of Guelph, Guelph, ON) are thanked for providing assistance on methods for physicochemical analyses.

The interlaboratory studies undertaken to validate the test methods described in the first edition of this method document were coordinated by J. Princz (Biological Methods Division, Environment Canada, Ottawa, ON) and conducted by the following participating laboratories: Environment Canada's Pacific Environmental Science Centre (North Vancouver, BC), Environment Canada's Atlantic Environmental Science Centre (Moncton, NB), Environment Canada's Soil Toxicology Laboratory at the Environmental Technology Centre (Ottawa, ON), NRC Biotechnology Research Institute (Montreal, PQ), Centre d'expertise en analyse environnementale du Québec (Sainte-Foy, PQ), Ontario Ministry of the Environment (Etobicoke, ON), Saskatchewan Research Council (Saskatoon, SK), BC Research Inc. (Vancouver, BC), EVS Consultants Ltd. (North Vancouver, BC), HydroQual Laboratories Ltd. (Calgary, AB), Stantec Consulting Ltd. (formerly ESG International Inc.) (Guelph, ON), Pollutech EnviroQuatics Ltd. (Point Edward, ON), and Bodycôte Essais Matériaux Canada Inc. (Sainte-Foy, PQ). The contribution of Buchanan Environmental Ltd. (Fredericton, NB) to discussions leading to these method-validation studies is also acknowledged with thanks. Names of laboratory personnel participating in this series of tests are provided in the technical report on these studies (EC, 2004a).

We gratefully acknowledge the many useful comments provided by each member of Environment Canada's Scientific Advisory Group (SAG) responsible for initial and final reviews of the first edition of this report, as follows: C. Bastien (Centre d'expertise en analyse environnementale du Québec, Ministère de l'Environnement, Saint-Foy, PQ), C. Edwards (Department of Entomology, Ohio State University, Columbus, OH), R. Kuperman (Geo-Centers, Inc., Aberdeen Proving Ground, MD), R. Lanno (Department of Entomology, Ohio State University, Columbus, OH), F. Riepert (Biologische Bundesantalt fur Land- und Forstwirtschaft, Berlin, Germany), J. Römbke (ECT Oekotoxikologie GmbH, Flörsheim am Main, Germany), G. Sunahara (Biotechnology Research Institute, National Research Council, Montreal, PQ), G. van Aggelen (Pacific Environmental Science Centre, EC, North Vancouver, BC), C.A.M. van Gestel (Institute of Ecological Science, Vrije Universiteit, Amsterdam), and S. Visser (Department of Biological Sciences, University of Calgary, Calgary, AB). Contact information for each SAG member for the first edition test method document is provided in Appendix D.

In addition to the SAG members who reviewed the first edition of this document, the following people also reviewed the initial and/or final drafts of the first edition and provided useful comments: K. Becker-van Slooten (École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland); S. Campiche (École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland); W. Diehl (Department of Biological Sciences, Mississippi State University, MS); J. Filser (Zentrum für Umweltforschung und Umwelttechnologie, University of Bremen, Bremen, Germany); P. Hankard (CEH Monks Wood, Abbots Ripton, Huntington, UK); J. Hatcher (HydroQual Laboratories Ltd., Calgary, Alberta); M. Hughes (University of Northern BC, Prince George, BC); K. Hund-Rinke (Fraunhofer Institut fuer Molekularbiologie und Angewandte Oekologie, Fraunhofer-IME, Schmallenberg, Germany); L. Kapustka (Ecological Planning and Toxicology Inc., Corvallis, OR); G. Linder (United States Geological Survey Field Station, Brooks, OR), J. McCann (Department of Biology, University of Waterloo, Waterloo, ON); J. Miller (Miller Environmental Sciences Inc., Innisfil, ON), J. Princz (Biological Methods Division, Environment Canada, Ottawa, ON); J.M.L. Rodrigues (Departmento de Biologia, Universidade de Aveiro, Portugal); M. Rutherford (University of Northern BC, Prince George, BC); M. Schaefer (Zentrum für Umweltforschung und Umwelttechnologie, University of Bremen, Bremen, Germany); J. Scott-Fordsmand (Department of Terrestrial Ecology, National Environmental Research Institute, Silkeborg, Denmark; J.P. Sousa (Instituto do Ambiente e Vida, Universidade de Coimbra, Coimbra, Portugal); J.B. Sprague (Sprague Associates Ltd., Salt Spring Island, BC); B.-J. Unis (HydroQual Laboratories Ltd., Calgary, AB); and M. Warne (Environment Protection Authority, Sydney, NSW, Australia).

This (second) edition was prepared by Jennifer Miller (Miller Environmental Sciences Inc., Uxbridge, ON), with assistance and guidance from Juliska Princz, Patrick Boyd, and Jessica Velicogna (Soil Toxicology Laboratory, Environment and Climate Change Canada [ECCC], Ottawa, ON) and Rick Scroggins (ECCC, Ottawa, ON). Studies related to the revision of these methods for *Eisenia andrei* and the development of the biological test methods defined herein for *Dendrodrilus rubidus*, carried out at ECCC's Soil Toxicology Laboratory, were directed by Juliska Princz and Rick Scroggins and grateful acknowledgement is made of the following laboratory personnel who participated in these studies over time: Patrick Boyd, Heather Lemieux, Jessica Velicogna, Emily Welsh-Crack, Christopher Fraser, Ellyn Ritchie, Leana Van der Vliet, and the various co-op students who provided technical assistance. Leana Van der Vliet and Carolyn Martinko (ECCC, Ottawa, ON) are sincerely thanked for their guidance and input on power analysis. Thanks also to Rick Scroggins who acted as Scientific Authority and provided guidance throughout the preparation of this report. We gratefully acknowledge the many useful review comments provided by Rick Scroggins, Juliska Princz, Patrick Boyd, Emily Welsh-Crack, Heather Lemieux, and Jessica Velicogna who provided feedback on initial drafts of this document and by Dr. Gladys Stephenson who provided feedback on the final draft of this document. Thanks also to Sylvain Trottier and Charles Faille at the Quebec Laboratory for Environmental Testing for their review of the French translation.

The Biological Assessment and Standardization Section would like to thank Leana Van der Vliet, Dr. Jill Clapperton, and Dr. John Reynolds for assisting ECCC's Soil Toxicology Laboratory in collecting and identifying earthworm specimens from the field, through which *D. rubidus* was identified and reared for these test methods. Furthermore, we thank the University of Guelph's Barcode of Life initiative with the genomic identification of the test species, including efforts by Stacey Saucier within the Soil Toxicology Laboratory for genomic confirmation. We also gratefully acknowledge Jason Nelson of Ecodynamics Consulting Inc. and Mary Moody for their efforts in identifying and collecting the field reference and contaminated boreal forest soils. We are very appreciative of the funding support provided by the Program for Energy Research and Development (PERD), which allowed ECCC to conduct research to improve these test methods.

The ongoing support of members of ECCC's Regional Environmental Testing Laboratories (Appendix C) and the Inter-Governmental Exotoxicological Testing Group (Appendix B) is also acknowledged.

Section 1

Introduction

1.1 Background

The Method Development and Applications Unit (MDAU) of Environment and Climate Change Canada (ECCC; previously Environment Canada) is responsible for the development, standardization, and publication (see Appendix A) of a series of biological test methods for measuring and assessing the *toxic effect(s)* on single species of terrestrial or aquatic organisms caused by their exposure to samples of test *materials* or *substances* under controlled and defined laboratory conditions. In 1994, MDAU, the Canadian Association of Petroleum Producers (CAPP), and the federal Program for Energy Research and Development (PERD) initiated a multi-year program to research, develop, validate, and publish a number of standardized biological test methods for measuring the *toxicity* of samples of *contaminated* or potentially contaminated *soil*, using appropriate species of terrestrial test organisms. The goal was to develop biological test methods applicable to diverse types of Canadian soils using terrestrial species that were representative of Canadian soil ecosystems. In a 2003 workshop convened by Environment Canada's MDAU, it was recommended that priority should be given to dedicating resources for the development of test methods using species that are more reflective on non-agricultural soils and/or habitats. With over 50% of Canada's total land mass being comprised of the boreal and taiga ecozones, and the contribution of resources within these ecozones to Canada's economy via oil and gas, mining, and forestry industries, priority was given to the development of standardized tests applicable to the assessment of *contaminants* present in boreal soils. Since then, several years of research have been completed on the selection of suitable and sensitive test organisms for measuring soil toxicity, including those from boreal forest soils (EC, 2010, 2013a), to meet the needs of industry, Canadian regulatory and monitoring requirements, and on the development of appropriate biological test methods. There have been three comprehensive reviews of existing biological test methods used internationally to evaluate the toxicity of contaminants to soil invertebrates (Bonnell Environmental Consulting, 1994; Römbke *et al.*, 2006; van Gestel, 2012).

ECCC's initiative resulted in the publication of five standardized soil *toxicity test* methods: i) Tests for Toxicity of Contaminated Soil to Earthworms (*Eisenia andrei*, *Eisenia fetida*, or *Lumbricus terrestris*), EPS 1/RM/43 (EC, 2004b amended 2007); ii) Test for Measuring Emergence and Growth of Terrestrial Plants Exposed to Contaminants in Soil, EPS/1/RM/45 (EC, 2005b, amended 2007); iii) Test for Measuring Survival and Reproduction of Springtails Exposed to Contaminants in Soil, EPS 1/RM/47 – 2nd edition (EC, 2014a); iv) Test for Growth in Contaminated Soil using Terrestrial Plants Native to the Boreal Region, EPS 1/RM/56 (EC, 2013b); and v) Test for Measuring Reproduction of Oribatid Mites Exposed to Contaminants in Soil, STB 1/RM/61 (ECCC, 2020a).

Since its publication in 2004, the first edition of this test method document (Test for Toxicity of Contaminated Soil to Earthworms (*Eisenia andrei*, *Eisenia fetida*, or *Lumbricus terrestris*), has been used by numerous government and private sector testing laboratories for soil toxicity testing (EC, 2004b). After 16 years of application, however, ECCC's MDAU recognized the need for the methods to be updated and a second edition test method document to be prepared. In 2014, ECCC commissioned a review of historical data in order to address technical issues and investigate test performance of the earthworm survival, reproduction, and *growth* method using the primary test species, *Eisenia andrei* (MESI, 2014).[1] The

[1] A total of 68 data sets for the earthworm survival, reproduction, and growth test using *E. andrei* were collected from two laboratories. The failure rate for data from both laboratories combined was 20%. For the valid tests, variability associated with the 28- or 35-day survival endpoint was much lower (mean CVs of 7.8%, 7.5%, and 3.4% for artificial, natural, and reference soils,

recommendations of this review helped direct several years of earthworm method refinement research conducted by staff of the ECCC's Soil Toxicology Laboratory (ECCC, 2020b). The results of this research, in addition to the development of *culture* conditions and test procedures required for the addition of a new test species representative of boreal regions of Canada (*Dendrodrilus rubidus*) are included in this second edition test method document (EC, 2010, 2013a; ECCC 2020b).

Detailed procedures and conditions for preparing and performing two biological test methods are defined herein. These include:

i) a 56-day test for effects on earthworm reproduction, and

ii) a 48-hour *sublethal* test of avoidance responses.

Universal procedures for preparing and conducting soil toxicity tests using a selected species of earthworm (i.e., *E. andrei* or *D. rubidus*) are described. Guidance is also provided for specific sets of conditions and procedures that are required or recommended when using either of these biological test methods for evaluating different types of substances or materials (e.g., samples of field-collected soil or similar particulate waste, or samples of one or more *chemicals* or chemical *products* experimentally mixed into or placed in contact with natural or formulated soil).

The flowchart in Figure 1 illustrates the universal topics covered herein, and lists topics specific to testing samples of field-collected soil, similar particulate waste (e.g., sludge, drilling mud, or dredged material), or soil spiked experimentally with chemical(s) or chemical product(s).

These biological test methods are intended for use in evaluating the *lethal* and sublethal toxicity of samples of material such as:

- field-collected soil that is contaminated or potentially contaminated;
- soils under consideration for removal and disposal or *remediation* treatment;
- dredged material destined or under consideration for land disposal after dewatering;
- industrial or municipal sludge and similar particulate wastes that might be deposited on land; and
- *clean* or contaminated *soil* (natural or *artificial*), spiked with one or more chemicals or chemical products (e.g., for *risk assessment* of new or current-use chemicals).

In formulating these biological test methods, an attempt has been made to balance scientific, practical, and cost considerations, and to ensure that the results will be sufficiently precise for most situations in which they will be applied. It is assumed that the user has a certain degree of familiarity with soil toxicity tests. Explicit instructions that might be required in a regulatory *protocol* are not provided in this report, although it is intended as a guidance document useful for that and other applications.

For guidance on the implementation of these and other biological test methods, and on the interpretation and application of *endpoint* data for soil toxicity, the reader should consult Sections 4.1.2, 5.5, and 5.6.4 of Environment Canada's *Guidance Document on Application and Interpretation of Single-Species Tests in Environmental Toxicology* (EC, 1999).

---

respectively), than those for the reproduction endpoint (mean CVs of 62%, 67%, and 60%, for artificial, natural, and reference soils, respectively). One main source of the variability associated with the control soils that was identified was the occurrence of "duds," where in a given replicate, despite full survival of the adults on Day 28 or 35, zero progeny were produced (MESI, 2014). These data led to numerous investigations for method improvement, including the development of a new test design with more adult worms being used in fewer replicates (i.e., 4 adult worms in each of 5 replicates vs. the first edition method, which required 2 adult worms in each of 10 replicates), and the shift from the traditional oatmeal as a source of food to Magic® Worm Food. Numerous other method improvements were investigated, and the results are incorporated in this second edition test method document (ECCC, 2020b).

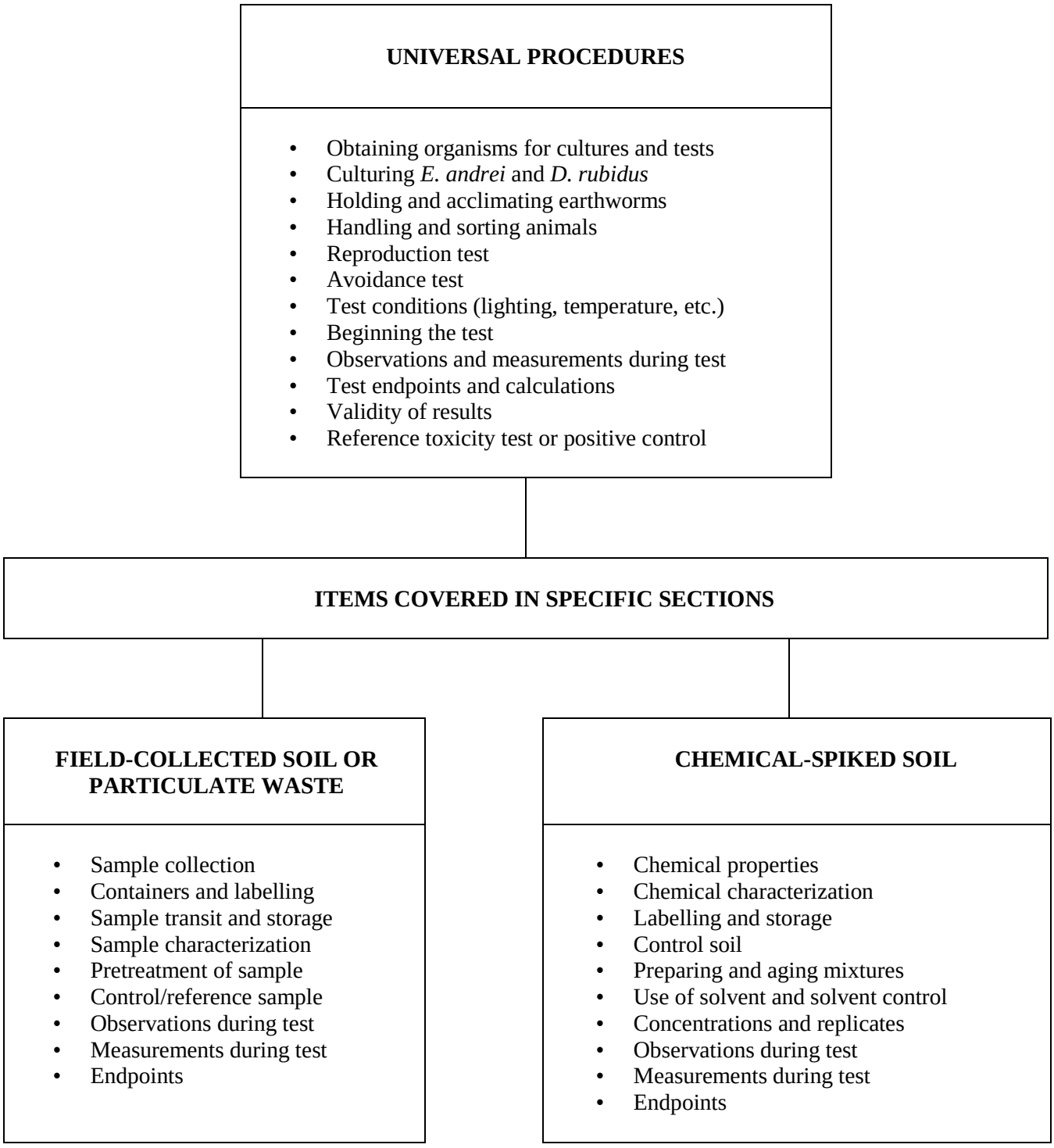

Figure 1 Considerations for preparing and performing soil toxicity tests using earthworms (*E. andrei* or *D. rubidus*) and various types of test materials or substances

In-depth direction on the use of statistics in determining effect endpoints in *ecotoxicology* testing is available in Environment Canada's *Guidance Document on Statistical Methods for Environmental Toxicity Tests* (EC, 2005a).

1.2 Identification, Distribution, and Life History of E. andrei *and* D. rubidus

The species of earthworms to be used for either of the biological test methods described herein (i.e., *Eisenia andrei* or *Dendrodrilus rubidus*) belong to the Lumbricidae family (phylum, Annelida; class, Clitellata; subclass, Oligochaeta; order, Haplotaxida; superfamily, Lumbricoidea; family, Lumbricidae). The *lumbricids* are not native to Canadian soils, and were most likely introduced from Europe by early settlers (Bonnell Environmental Consulting, 1994). Definitive information regarding the identification, distribution, biology, and life history of lumbricid earthworms including *E. andrei* and *D. rubidus* is found in a number of publications, including: Edwards and Lofty, 1977; Reynolds, 1977, 1994, 2015; Bengtsson *et al*., 1986; Fender, 1985; Sims and Gerard, 1985, 1999; Curry, 1988; Bouché, 1992; Frenot, 1992; Christensen and Mather, 1994; Edwards and Bohlen, 1996; Blakemore, 2008 cited in Csuzdi *et al*., 2017; Addison, 2009; Berman *et al*., 2010; Domínguez and Edwards, 2010; and Coulson *et al*., 2013. Lumbricid earthworms are important members of the soil fauna, and are appropriate organisms for use in the assessment of potentially toxic soils. Together with other earthworms, they constitute up to 92% of the invertebrate *biomass* of soil and are important in the maintenance of soil structure and nutrient cycling (Edwards and Lofty, 1977; Lee, 1985). Additionally, lumbricid and other earthworms represent a significant component of the diet of many species of birds, small mammals, reptiles, amphibians, and invertebrates (Macdonald, 1983; Cooke *et al.*, 1992). Earthworms can accumulate a variety of organic and inorganic compounds that might (or might not) affect them adversely (Edwards and Bohlen, 1992). A major change in the abundance of lumbricid earthworms could have serious adverse ecological effects on the entire terrestrial system (ASTM, 2012).

1.2.1 Eisenia andrei

Eisenia andrei is commonly referred to as the *red wiggler*, *compost worm*, or *manure worm* (Aquaterra Environmental, 1998). Taxonomists have difficulty distinguishing *E. andrei* from *E. fetida*, and morphological features alone are insufficient to enable this (R. Blakemore, VermEcology, personal communication, 2000; W.J. Diehl, Mississippi State University, personal communication, 2000). Historically, in much of the literature, *E. fetida* has been misidentified as *E. andrei* (or as *E. fetida andrei*), with *E. andrei* being the sibling species found most commonly in North American composts or cultures from commercial suppliers of earthworms of *Eisenia* spp. (W.J. Diehl, Mississippi State University, personal communication, 2000; McCann, 2004; Römbke *et al*., 2016). Early methods for the detection of genetic variation, however, indicated that a definitive identification could be made based on differing electrophoretic patterns of certain enzymes for these two species (Jaenike, 1982; Øien and Stenersen, 1984; McElroy and Diehl, 2001; McCann, 2004). Today, DNA barcoding is recommended for proper identification of *E. andrei* (see Section 2.1).

Historically, *E. fetida* has been referred to by some as a "species complex" (Bouché, 1992; Christensen and Mather, 1994). *E. fetida*/*andrei* has also been described by certain taxonomists as having two morphologically similar subspecies or races (i.e., *E. fetida*, which typically has transverse striping or banding on its segments, and *E. fetida andrei*, which lacks this and has a variegated reddish colour). This (now outdated) means of classification was adopted in certain biological test methods for measuring soil toxicity using earthworms (OECD, 1984; ISO, 1993; ASTM, 1999, 2012). However, earthworm taxonomists now classify *E. andrei* and *E. fetida* as distinct species, while recognizing that morphological characteristics including colouration and segmental banding or striping patterns are insufficient to distinguish them with complete confidence (R. Blakemore, VermEcology, personal communication, 2000; W.J. Diehl, Mississippi State University, personal communication, 2000). More recent methodology documents also refer to *E. fetida* (or *E. fetida fetida*) and *E. andrei* (or *E. fetida/andrei*) as distinct species (Sheppard, 1988; ISO, 2008, 2012; OECD, 2016). This approach is supported by the results of breeding experiments involving these two species, which found that *cocoons* were produced when *E. fetida* and *E. andrei*

were cross-bred, although none of these were viable (Ferreiro *et al.*, 2002; Domínguez *et al*., 2005). To date, the species of *Eisenia* typically used in Canadian laboratories for soil toxicity tests has now been confirmed to be *E. andrei* (W.J. Diehl, Mississippi State University, personal communication, 2000; McCann, 2004; Römbke *et al*., 2016), although it was formerly identified as *E. fetida* (e.g., Aquaterra Environmental, 1998; Aquaterra Environmental and ESG, 2000). There are few studies on the relative sensitivity of these two related species to samples of contaminated soil. Side-by-side laboratory tests by Ingraldi *et al.* (2004), in 14-day *acute* lethality tests, showed that *E. andrei* was somewhat more sensitive than *E. fetida* to boric acid in *artificial soil*, with seven-day *LC50s* of 3236 mg/kg and 4365 mg/kg, respectively, and 95% confidence limits that did not overlap. Similarly, results for 14-day LC50s performed concurrently by these investigators using each of these two species exposed to a sample of condensate-contaminated soil mixed in an uncontaminated clay loam soil, indicated a somewhat greater sensitivity of *E. andrei*. Comparative 48-hour avoidance tests with multiple *concentrations* of this same condensate-contaminated soil in clean clay loam soil, performed according to Section 4.3 herein, also showed a greater sensitivity of response by *E. andrei* to the contaminated soil (Ingraldi *et al.*, 2004). Although *E. fetida* was included as a species option in the first edition of this test method document, it has been dropped as a test species option in this second edition version (see Section 2.1).

Eisenia andrei is commonly found in North American composters and is sold commercially for fish bait (as "red wigglers") and composting (as "compost worms"). *Adult* worms of this species have an average body length of 35–130 mm and an average diameter of 3–5 mm, with between 80 and 110 body segments. Diagnostic characteristics include an *epilobic prostomium*, first dorsal pore on 4/5 or sometimes 5/6 segments, and a *clitellum* on segments 24–32 (Reynolds, 1977). The *tubercula pubertatis* is found on segments 28–30. The *setae* are closely paired with a characteristic variation in patterning that differs from the anterior to the posterior end of the worm. *Genital tumescences* might be present around any of the setae on segments 9–12 of the cylindrical body that can vary in colour from red to dark red, brownish red, or purple, with alternating bands of red-brown pigment and pigmentless yellow intersegmental areas (Reynolds, 1977). The male pores usually have large glandular *papillae* on segment 15. The *spermathecae* are two pairs with ducts, which open on segments 9/10 and 10/11. Four pairs of *seminal vesicles* are found on segments 9–12.

E. andrei is native to the Palearctic, and is also found in Europe, North and South America, Asia, Africa, Iceland, and Australasia (Reynolds, 1977). This gregarious species is generally associated with anthropogenic activities, and is commonly found across North America in gardens, compost, and manure piles (Edwards and Lofty, 1977). Within Canada, *E. andrei* has been found in the provinces of British Columbia, Alberta, Ontario, Quebec, New Brunswick, Nova Scotia, and Prince Edward Island (Reynolds, 1977; M.J. Clapperton, Agriculture and Agri-Food Canada, personal communication, 2000). This species prefers moist soils with high levels of *organic matter*. It is generally restricted to the upper layers of soil and is considered to be *epigeic*, selectively feeding (with little ingestion of soil) on organic material dispersed throughout the soil (Wallwork, 1983). *E. andrei* is tolerant of a wide range (i.e., 4 to 8) of soil *pH* values (Stephenson, 2002), although it prefers soils with a pH between 7 and 8 (Edwards and Lofty, 1977).

E. andrei is thought to have a lifespan of four to five years, although between one and two years is more common (Reynolds, 1977). *E. andrei is* obligatorily *amphimictic*, although uniparental reproduction has been reported (Reynolds, 1977, 1995). *E. andrei* copulates and casts below ground. This species reproduces rapidly at temperatures ranging within 20–25 °C and can reach sexual maturity within 52 days. Time for completion of a life cycle is appreciably slower at cooler temperatures (e.g., >166 days at 13 °C) (ASTM, 2012). Cocoons are produced at a frequency of one or two, every three or four days; each cocoon can produce as many as six or more *hatchlings*, although one to four offspring per cocoon is more commonly observed (Reinecke and Viljoen, 1991; Reinecke *et al.*, 1992; Edwards and Bohlen, 1996). These characteristics (i.e., rapid rate of cocoon production, large number of offspring, short generation time, rapid maturation time) and the fact that *E. andrei* can be easily

cultured in the laboratory (ASTM, 2012; ISO, 2008, 2012; OECD, 2016) make this earthworm one of the most commonly used test species for earthworm reproduction tests (Aquaterra Environmental, 1998).

1.2.2 Dendrodrilus rubidus

Dendrodrilus rubidus is commonly referred to as the jumping red wiggler, red trout worm, red wiggler, jumbo red worm, pink worm, bank worm, tree worm, cockspur, gilt tail, gold-tailed brandling, and European barkworm (Sims and Gerard, 1985; GISD, 2020). It is generally acknowledged by taxonomists that *D. rubidus* is polymorphic and comprised of four recognized morphs or subspecies: *rubidus* (Savigny 1826), *tenuis* (Eisen 1874), *norvegicus* (Eisen 1874), and *subrubicundus* (Eisen 1874) (Sims and Gerard, 1985, 1999; Frenot, 1992; GISD, 2020). Some morphs reproduce sexually, while others reproduce parthenogenetically. The *tenuis* morph completely lacks tubercles as well as spermathecae. In the *rubidus* morph, the spermathecae are sometimes present but usually empty, and indistinct tubercles can be seen in 29–30. Spermathecae are present in *norvegicus*, and in the *subrubicundus* morph, even filled spermathecae can be seen and tubercles are easily recognized on 28–30 (Blakemore, 2008 cited in Csuzdi *et al*., 2017). Adult worms of this species are small with an average body length of 20–100 mm and an average diameter of 2–5 mm, with between 50 and 120 body segments. Diagnostic characteristics include an epilobic prostomium, inconspicuous dorsal pores starting on the 5/6 segment, and the clitellum (saddle-shaped) on segments 25, 26–31, 32. The tubercles, when present, are found on segments 29–30 or 28–30 either as a broad rectangular band or reduced to a slender strip (when over two segments) and interrupted by furrow 29/30 (Sims and Gerard, 1985). Setae are moderately paired, closer ventrally, and wider laterally. *D. rubidus* is heavily pigmented red-violet, darker on the dorsal surface, pale ventrally, and yellowish-orange in colour on the tail end. The female genital pore is located on segment 14, dorsad of setae b. Male pores are located on segment 15, equatorial just above setae b, on a small porophore confined to its own segment. *Septa* 5/6–10/11 are slightly thickened. Calciferous glands are in segments 10–20 with large diverticula present in segment 10. The excretory system is holoic. Nephridial bladders are U-shaped throughout, with forward-bent ental limbs. Typhlosoles are well developed and lamelliform (Reynolds, 1977; Sims and Gerard, 1985; Blakemore, 2008 cited in Csuzdi *et al*., 2017).

D. rubidus is a Holarctic species with a cosmopolitan distribution, having been found on every continent in the world except Antarctica (Berman *et al*., 2010). Considered as resident in large parts of the Holarctic, and a natural component of the sub-Arctic and Arctic terrestrial fauna, its wide distribution in northern climates is well known (Bengtsson *et al*., 1986; Frenot, 1992; Reynolds, 1977, 1994; Berman *et al*., 2010; Coulson *et al*., 2013). Within Canada, *D. rubidus* is found in all of the provinces including the Yukon Territory, but has yet to be identified in the Northwest Territories or Nunavut (Addison, 2009; Reynolds, 1977, 2015). Although the worms are sensitive to the cold, failing to survive even after a brief exposure to temperatures below 0 °C, the cocoons (i.e., embryos) can withstand temperatures as low as -196 °C (Berman *et al.*, 2010). Thus, only cocoons overwinter in colder climates. *D. rubidus* is an epigeic species, most often found in the uppermost (i.e., 10 cm) soil layer, inhabiting a wide range of habitats and various soil types. Although it has been shown to tolerate soils with low pH (Edwards and Bohlen, 1996), a decrease in cocoon production, survival, and growth has been shown in soils with pH ≤ 4.5 (Bengtsson *et al*., 1986; Rundgren and Nilsson, 1997). It has also been shown to avoid soils with lower pH (EC, 2010). It prefers substrates rich in organic material (EC, 2010), typically inhabiting coniferous forests and cultivated soils, but it has also been associated with branches on cave floors or mine support beams. It is frequently found in leaf litter, detritus, and under the bark of decaying logs (McAlpine and Reynolds, 1977). It is an active species, and on damp nights has been noted crawling on the surface of the ground and climbing trees (Reynolds, 1977). *D. rubidus* is considered a primary decomposer, feeding predominantly on litter rather than detritus (Scheu and Falca, 2000).

D. rubidus has a high reproduction rate, making it favourable for use as fish bait and composting worms. At 20–25 °C, it has a mean maturation time of 51–54 days and a mean cocoon production rate of 0.2–0.4 cocoons per earthworm per day. Hatching success is 85% with a mean incubation time of 22 days and one to three hatchlings emerging from

each cocoon (Elvira *et al*., 1997). *D. rubidus* can complete its life cycle within 75 days, however maturity is typically reached between 18 and 20 weeks, with the clitellum developing at 100 days and disappearing at 320 days. Its lifespan is typically one year, having only one period of sexual activity during which it produces 40–95 cocoons (Bengtsson *et al*., 1986). These characteristics (i.e., rapid rate of cocoon production, large number of offspring, short generation time, rapid maturation time) make *D. rubidus* ideal for culturing in the laboratory.

1.3 Historical Use of Earthworms in Toxicity Tests

Earthworms are frequently exposed to toxic chemicals in soil. Besides the myriad of fertilizers, insecticides, herbicides, and fungicides from agricultural and domestic applications, earthworms are sometimes exposed to heavy metals, petroleum hydrocarbons, or other chemicals such as wood preservatives (e.g., pentachlorophenol) or nitroaromatic explosive compounds in contaminated soils.

Earthworms are widely used as test organisms in single-species laboratory toxicity tests intended to measure the toxicity of pure chemicals, chemical products, or samples of soil contaminated or potentially contaminated with chemicals in the field or (for experimental purposes) in the laboratory. The toxicity of various chemicals or chemical products to earthworms, as determined in the laboratory under standardized conditions using lethal and/or sublethal endpoints and acute (hours or a few days) or prolonged (several weeks) exposures has been well studied (Natal-da-Luz *et al*., 2008; Chelinho *et al*., 2011; Hirano and Tamae, 2011; Sivakumar, 2015; Princz *et al*., 2017; Uwizeyimana *et al*., 2017).

The use of earthworm toxicity tests as "ecotoxicological assessment tools" for appraising the toxicity of contaminated or potentially contaminated *site soil* is increasing in Canada and elsewhere (Callahan, 1988; Menzie *et al.*, 1992; Römbke *et al*., 1994; Kula and Larink, 1997; Spurgeon *et al*., 1994; Spurgeon and Hopkin, 1995, 1996a; Yeardley *et al*., 1996; Chang *et al.*, 1997; Meier *et al*., 1997; Stephenson *et al.*, 1997; Aquaterra Environmental, 1998; Saterbak *et al.*, 1999; Stephenson *et al.*, 2002; Stephenson, 2003a; Princz *et al.*, 2012; Renoux *et al*., 2013). Studies comparing the results of single-species toxicity tests performed in the laboratory with related field surveys for effects on terrestrial biota have generally found a strong correlation between the laboratory findings and the field results (Edwards and Bohlen, 1992; Kula and Kokta, 1992; Menzie *et al.*, 1992; van Gestel, 1992, 1997; Heimbach, 1993, 1997; Christensen and Mather, 1994; Kula, 1995). Scientists, however, have frequently commented that it is difficult to extrapolate results for single-species laboratory tests with earthworms to the field situation. A number of researchers have discussed how to improve the predictive worth of the laboratory toxicity tests (i.e., their ability to discern adverse environmental conditions or effects). Promising improvements include reliable procedures for estimating the bioavailability of inorganic and organic contaminants in soil, tiered testing approaches, and risk assessment schemes for soil toxicity that include earthworm toxicity tests (Bouché, 1988; Callahan, 1988; Lofs-Holmin and Bostrom, 1988; NERI, 1993; Keddy *et al.*, 1995; Leon and van Gestel, 1994; Christensen and Mather, 1994; Sauvé *et al.*, 1996, 1998, 2000; Barber *et al.*, 1997; Meier *et al.*, 1997; Saterbak *et al.*, 1999; Conder and Lanno, 2000; Wells and Lanno, 2001; Chelinho *et al*., 2011; Princz *et al*., 2012; Velicogna *et al*., 2012, 2016; Cermak *et al*., 2013; Renoux *et al*., 2013; Brami *et al*., 2017; Ritchie *et al*., 2017; Niemeyer *et al*., 2018; Renaud *et al*., 2018; de Santo *et al*., 2019; Prodana *et al*., 2019; Kilpi-Koski *et al*., 2020).

A number of investigators have studied the effects of variations in natural characteristics of *chemical-spiked soil* or site soil, on the soil's toxicity to earthworms. Variables investigated include soil pH, organic carbon content, particle size, and *moisture content* (Heimbach and Edwards, 1983; van Gestel and van Dis, 1988; van Gestel, 1991; Christensen and Mather, 1994; Spurgeon and Hopkin, 1996b; Yeardley *et al*., 1996; Bauer and Römbke, 1997; Puurtinen and Martikainen, 1997; Meharg *et al*., 1998; Aquaterra Environmental and ESG, 2000; Robidoux *et al*., 2004; Bradham *et al*., 2006; Natal-da-Luz *et al*., 2008; Chelinho *et al*., 2011; Scheffczyk *et al*., 2014; Alves *et al*., 2018; Lanno *et al*., 2019; Velicogna *et al*., 2020). The influence of these soil variables on chemical toxicity depends on

interactions between the physicochemical characteristics of the soil and the type(s) and speciation of chemical contaminant(s) therein.

Laboratory tests that measure the effects of contaminated soil on the avoidance behaviour of earthworms are increasingly used (see Section 1.3.2), as are those that measure the effects of prolonged exposures on earthworm reproduction (see Section 1.3.1). Certain researchers have also studied or reviewed other *sublethal* (e.g., gametogenic, teratogenic, neurotoxic, immunotoxic, cytotoxic, or genotoxic) *effects* of chemical-contaminated soil on earthworms (Drewes *et al.*, 1984; Zoran *et al.*, 1986; Edwards and Bohlen, 1992; Fitzpatrick *et al.*, 1992; Cikutovic *et al.*, 1993; Goven *et al.*, 1993, 1994; Christensen and Mather, 1994; Suzuki *et al.*, 1995; Brousseau *et al.*, 1997; Giggleman *et al.*, 1998; Scott-Fordsmand *et al.*, 2000; Plytycz *et al.*, 2009; Button *et al.*, 2010, 2012; Vasseur and Bonnard, 2014; Demuynck *et al.*, 2016; Cao *et al.*, 2017; Lackmann *et al.*, 2018; Tatsi *et al.*, 2018; Bouguerra *et al.*, 2019; Saggioro *et al.*, 2019; Chen *et al.*, 2020; Pereira *et al.*, 2020; Ramires *et al.*, 2020).

1.3.1 Reproduction Tests

The effects of exposure to toxic substances or materials on the survival, reproduction, and growth of a single species of test organism, under controlled laboratory conditions, are recognized and accepted by environmental toxicologists as ecologically relevant responses. From an ecological viewpoint, these biological effects represent "ideal endpoints" for laboratory toxicity tests with earthworms (Christensen and Mather, 1994). Christensen and Mather (1994) recommended their inclusion in an assessment protocol, following their review of the use of earthworms as test organisms for evaluating the ecological *risk* of toxic chemicals in soil.
In 1988, international efforts were initiated to develop and standardize tests for measuring the effects of long-term exposure to contaminants in the soil on survival, reproduction, and growth of earthworms (van Gestel *et al.*, 1988). A number of standard methods or guidelines were developed using *E. andrei* and *E. fetida*; these are now commonly applied and their use is expanding.
E. andrei and *E. fetida* are the preferred test organisms for studying the effects of prolonged exposure to contaminants on the survival, reproduction, and growth of earthworms because of the widespread knowledge and experience in culturing these species, their rapid life cycle, their international distribution, and their frequent use in toxicity tests (OECD, 1984, 2016; USEPA, 1989, 2012; ISO, 2008, 2012; ASTM, 2012). The development, growth, and reproductive biology of *E. andrei* and *E. fetida* under laboratory conditions has been extensively studied and is well documented (e.g., Edwards and Lofty, 1977; Tsukamoto and Watanabe, 1977; Sheppard, 1988; van Gestel *et al.*, 1992a). The toxic effects of prolonged exposure to contaminated soil on the survival, reproduction, and/or growth of *E. andrei* and *E. fetida* have also been well documented in laboratory studies involving samples of soil spiked or contaminated with:

- pesticides (Lofs-Holmin, 1980; Venter and Reinecke, 1988; Neuhauser and Callahan, 1990; van Gestel *et al.*, 1992b; Riepert and Kula, 1996; Bauer and Römbke, 1997; Heimbach, 1997; Kula and Larink, 1997; ESG and Aquaterra Environmental, 2002; Rico *et al.*, 2016; Lackmann *et al.*, 2018; Alves *et al.*, 2019; de Santo *et al.*, 2019; Saggioro *et al.*, 2019; de Lima e Silva *et al.*, 2020; Pereira *et al.*, 2020);

- heavy metals (Neuhauser *et al.*, 1984; van Gestel *et al.*, 1989, 1992b; Spurgeon *et al.*, 1994; Reinecke and Reinecke, 1996; Spurgeon and Hopkin, 1996a; Fischer and Molnár, 1997; Kula and Larink, 1997; Aquaterra Environmental and ESG, 2000; Scott-Fordsmand *et al.*, 2000; ESG, 2002; Bradham *et al.*, 2006; Renoux *et al.*, 2013; Velicogna *et al.*, 2016; Jesmer *et al.*, 2017; Ritchie *et al.*, 2017, 2019; Alves *et al.*, 2018; Kilpi-Koski *et al.*, 2020; McGuirk *et al.*, 2020);

- petroleum hydrocarbons (Aquaterra Environmental, 1998; Stephenson *et al.*, 1998, 1999a, 1999b, 2000a; ESG, 2001; Princz *et al.*, 2012; Cermak *et al.*, 2013); and

- other chemicals including *reference toxicants* (Hartenstein, 1982; van Gestel *et al.*, 1989, 1992b; Neuhauser and Callahan, 1990; Gibbs *et al.*, 1996; Aquaterra Environmental, 1998; Robidoux *et al.*, 2000, 2001; Becker *et al.*, 2011; Velicogna *et al.*, 2012, 2016; Dodard *et al.*, 2013; Ritchie *et al.*, 2013, 2017; Scheffczyk *et*

al., 2014; Bouguerra *et al*., 2016; Jesmer *et al*., 2017; Princz *et al.*, 2017).

In their initial efforts to develop a standardized test method for determining the effect of chemical substances on the reproduction of *E. andrei/fetida*, van Gestel *et al.* (1988) performed five-week incubation studies that measured cocoon viability and numbers of hatchlings per cocoon, following cocoon recovery from earlier chemical-exposure studies with adult earthworms and their incubation in water or artificial soil. Subsequently, van Gestel *et al.* (1989) described a Dutch test method whereby adult *E. andrei/fetida* were preconditioned for one week in artificial soil and exposed thereafter to a range of concentrations of chemical-spiked artificial soil, after which the cocoons produced were incubated for a further five weeks in untreated artificial soil to assess hatchability.

In 1990, a German working group established by the Biologische Bundesanstalt für Land- und Forstwirtschaft (BBA) and joined by experts from the Netherlands and Switzerland developed a slightly different test method, whereby adult *E. andrei/fetida* were exposed to chemical-spiked artificial soil for four weeks. After this time the worms were removed, and the exposure continued with their *progeny* for a further four weeks. This draft method was introduced by Germany to the ISO Working Group "Soil Fauna" (WG 2) in 1990, and was later published by the BBA (1994) as a guideline for testing the toxicity of pesticides. Following further evaluation and consideration by other scientists, ISO (1998) published a modified version of the BBA (1994) method. The standard method published by ISO (1998) consists of a four-week exposure of adult *E. andrei/fetida* to a range of concentrations of chemical-spiked soil or contaminated site soil with observations thereafter (adult survival and their increase or decrease in wet weight), and a subsequent four-week exposure to the same chemical-spiked soils with an endpoint measurement of number of offspring (*juveniles*) produced per *treatment* (see Appendix E). This ISO standard was further updated following a periodic review (ISO, 2012), and details are outlined in Appendix E. Currently, this ISO standard (11268-2) is being amended again to add new annexes with instructions on culturing and testing using the boreal species, *D. rubidus* and other alternative earthworm species. Preparation of the *D. rubidus* annex to ISO 11268-2 is being led by Canadian experts. The Organisation for Economic Co-operation and Development (OECD) has also published a similar test method (OECD, 2016; see Appendix E for specifics). A shorter (28-day) test, restricted to determinations of the survival and weight change of adult *E. andrei/fetida* during exposure, has been published by the USEPA (2012) as a method for screening samples of contaminated soil. This shorter test method, however, is not widely used because it does not measure the effects on the reproduction of earthworms and the survival of their progeny.

In 2006, Römbke *et al*. identified *D. rubidus* as a good candidate species for inclusion in ECCC's biological test method series, as part of a battery of test options relevant to Canadian boreal forest, taiga, and tundra ecozones. The criteria for candidate species selection included habitat, frequency and abundance, origin, taxonomy, practicability, and stress tolerance. In 2012, Princz *et al.* further evaluated the practicability and sensitivity of *D. rubidus* in 56-day reproduction toxicity tests using site soils impacted by petroleum hydrocarbon (PHC) and salt contamination. Results of this study demonstrated the applicability of this boreal species with comparable performance and sensitivity to standard (agronomic) test species (i.e., *E. andrei*). Several years of research undertaken by ECCC's Biological Assessment and Standardization Section resulted in defined culturing procedures for *D. rubidus* and a test design to assess for effects on survival and reproduction (EC, 2010, 2013a; ECCC, 2020b).

Standardized procedures and conditions for performing a biological test method that measures the toxic effects of exposure to chemical-spiked soil or site soil on the reproduction of *E. andrei* and *D. rubidus* are defined herein (see Section 4.2).[2] This

[2] The growth of juvenile earthworms (measured as individual dry weight of juveniles at the end of the test) was a required endpoint in the first edition of this test method document. Following a review of test performance data for the earthworm survival, reproduction, and growth test using *Eisenia andrei*, it was concluded that individual dry weight of worms can be

biological test method is largely in keeping with ISO (2012) and OECD (2016).

1.3.2 Avoidance Tests

Lumbricid earthworms including *E. andrei* and *D. rubidus* are known to be highly mobile (Karnak and Hamelink, 1982; Mather and Christensen, 1992; McAlpine and Reynolds, 1977). A number of researchers have concluded that a behavioural avoidance response by earthworms to sublethal concentrations of chemicals in soil can have ecological relevance at the population level (Christensen and Mather, 1994; Tomlin, 1995; Yeardley *et al.*, 1996). Some evidence suggests that these and certain other terrestrial invertebrates are able to minimize exposure to harmful chemicals through such behaviour (Yeardley *et al.*, 1996; Haimi and Paavola, 1998). Christensen and Mather (1994) reviewed the use of earthworms as test organisms for evaluating chemical hazards and as part of *ecological risk assessments* for the Danish Environmental Protection Agency. They concluded that, from an ecological viewpoint at the population level, toxicity tests that measured effects on migratory (avoidance) behaviour were amongst those considered as "ideal endpoints," and recommended their application. Advantages of tests for an avoidance response include their short duration (relative to more prolonged tests for effects on reproduction) and their sensitivity (i.e., their ability to detect a behavioural response at sublethal concentrations). Earthworms exposed to contaminated soil typically show an avoidance response to sublethal concentrations within 24–72 h of exposure (Wentsel and Guelta, 1988; Yeardley *et al.*, 1996; Slimak, 1997; Hund, 1998; Stephenson *et al.*, 1998; Hund-Rinke and Wiechering, 2001; ESG and Aquaterra Environmental, 2002; Hund-Rinke *et al*., 2003, 2005; Schaefer, 2003; Stephenson, 2003a; ISO, 2003, 2008). Avoidance tests with *E. andrei* given a choice between *negative control soil* (natural or artificial) and various concentrations of a condensate-contaminated site soil diluted with the respective negative control soil showed a concentration-dependent avoidance response at sublethal concentrations. Associated prolonged-exposure tests with *E. andrei* and the same sample of condensate-contaminated site soil indicated that the threshold concentration avoided by this species of earthworm was similar to the threshold-effect concentration that reduced reproductive success and subsequent growth of offspring. Similar results were obtained for a site soil contaminated with amines and glycol products in that the earthworms avoided sublethal concentrations in soil that resulted in adverse effects on reproduction (Aquaterra Environmental, 1998; Stephenson, 2003a).

Laboratory tests that measure avoidance are particularly useful, from an ecological perspective, when performed in conjunction with standard toxicity tests such as those that measure sublethal effects on reproduction. The earthworm avoidance test has been identified as a useful screening tool for assessing the habitat function of soils. Avoidance behaviour towards organic chemicals and heavy metals proved to be a suitable rapid screening method for identifying *sites* or soils with impaired habitat function, and for selecting soil samples for which more *definitive* assays (i.e., 56-day reproduction test) might be necessary (Hund-Rinke and Wiechering, 2001; Hund-Rinke *et al*., 2003, 2005; ISO, 2008; EC, 2012). The ecological relevance of findings for an avoidance test in the absence of comparable data for standard toxicity tests (i.e., 56-day reproduction test) might produce confusing or questionable results, since the earthworms might avoid concentrations of contaminants that are not damaging to their tissues or might fail to avoid concentrations that are. However, the earthworm avoidance test has proven extremely effective and relevant as a screening tool for potential sublethally-toxic areas before or after chemical characterization in risk assessments (Tiers

---

influenced, and therefore confounded, by the number of juveniles produced in a given replicate (MESI, 2014). In addition, based on the data reviewed, there was a strong tendency for the reproduction endpoint to be more sensitive than the growth endpoint; however, the large variability with the growth and reproduction endpoints results in very wide, and in most cases, overlapping confidence intervals (MESI, 2014). Following consideration by ECCC's Biological Assessment and Standardization Section (BASS), it was decided that the revised method should focus on endpoints that are less variable and more sensitive (i.e., adult 28-day survival and 56-day reproduction), and therefore the growth (i.e., juvenile dry weight) endpoint is no longer required herein. It has been retained, however, as an optional endpoint (see footnote 44 in Section 4.2), as there might be occasions where dry weight might be of interest to researchers in a given study.

1 and 2), or as a *range-finding test* for further toxicological assessment (i.e., 56-day reproduction test) of contaminated or *spiked soils* (Hind-Rinke *et al*., 2005; EC, 2012).

Apparatus and procedures used for measuring avoidance responses of earthworms to contaminated soil have been varied. Using both a two-compartment unit and a six-compartment unit, Hund-Rinke *et al.* (2005) determined that *E. fetida* avoided soil contaminated with organic chemicals and heavy metals in 48-hour exposures at sublethal concentrations at or below those causing reproductive effects in adults in 56-day reproduction tests. In tests using circular units, Yeardley *et al.* (1996) found that *E. andrei/fetida* avoided sublethal concentrations of chemical-spiked soils or toxic site soils when exposures were as brief as one to two days. ECCC (2020b) performed 24- and 48-hour avoidance tests with *D. rubidus* exposed to a range of concentrations of boric acid spiked into clean artificial soil. There were no obvious trends in avoidance at 24-h, but there was a concentration-dependent avoidance response to boric acid observed in the 48-hour exposure (ECCC 2020b).

Confounding effects due to differing physicochemical characteristics of soil (e.g., particle size, organic carbon content, total nitrogen content, pH, *water-holding capacity*) have been found to be minimal (Yeardley *et al*., 1996; Hund, 1998; Hund-Rinke and Wiechering, 2001), however Delgadillo *et al*. (2017) determined that *electrical conductivity* and organic matter content in the control substrate can influence the avoidance response. A method published by ISO in 2008 (ISO 17512-1), provides a standardized approach for conducting 48-hour avoidance tests with *E. andrei* or *E. fetida* in contaminated or spiked soil using either a two-compartment or a six-compartment test unit for both single-and multi-concentration exposures (ISO, 2008).

During the past many years, a number of studies have been performed in Canada to develop and further standardize the avoidance test for soil toxicity using *E. andrei* and *D. rubidus* (Stephenson, 2003a; EC, 2010; ECCC, 2020b). The experimental apparatus used, and recommended herein, is the six-compartment test unit illustrated photographically in Stephenson *et al.* (1998) and schematically in Section 3.2.3 (as Figures 2 and 3).

Standardized procedures and conditions for performing a biological test method that measures the avoidance response of earthworms (*E. andrei* or *D. rubidus*) to chemical-spiked soil or site soil are defined in Section 4.3.

Section 2

Test Organisms

2.1 Species and Life Stage

Both the 56-day reproduction test (Section 4.2) and the 48-hour avoidance test (Section 4.3), described herein, must be performed using *Eisenia andrei* or *Dendrodrilus rubidus* (Sections 1.2.1 and 1.2.2).[3] The identification, distribution, and life history of *E. andrei* and *D. rubidus* are summarized in Section 1.2. Species identification must be confirmed and documented[4] upon establishment of a new culture, and/or with each new *batch* of earthworms obtained for testing or introduced to the laboratory culture (Römbke *et al*., 2016). Cultures of *E. andrei* and *D. rubidus* held for a prolonged period at a testing laboratory should be identified to species at least once every two years. Species identification may be made using the distinguishing taxonomic features described and illustrated in taxonomic keys by qualified personnel experienced with identifying the intended species (see Section 1.2) of earthworm to be used in the toxicity test, or using DNA-based taxonomic identification (i.e., barcoding) (ISO, 2019). It is highly recommended that species identification be conducted using DNA-based taxonomic identification for *E. andrei* and *D. rubidus*.[5] The wet weight of each *adult* worm used to start either the reproduction test or the avoidance test must range within 250–600 mg for *E. andrei* and 50–200 mg for *D. rubidus* (see Sections 4.2.1 and 4.3.1). In addition, worms used to start either test must be clitellated (i.e., adult worms).

2.2 Source

Laboratory-cultured earthworms must be used as the source of test organisms for both the 56-day reproduction test (Section 4.2) and the 48-hour avoidance test (Section 4.3). *E. andrei* and *D. rubidus* may either be cultured in the laboratory conducting the testing (i.e., in-house; see Section 2.3), or obtained from an outside source and acclimated to laboratory conditions (Section 2.4) before the test is initiated.[6] Sources of *E. andrei* and *D. rubidus* for use in toxicity testing must be

[3] *E. fetida* and *L. terrestris* were both species options for testing in the first edition of this test method document. Both have been removed as options herein due to lack of demand for testing using these two species and lack of a national supply for *E. fetida*. *D. rubidus* has been added as a species option herein as it is representative of the boreal regions of Canada, and therefore relevant for testing soils collected from the boreal and taiga ecozones (see Section 1.2.2).

[4] Acceptable forms of documentation include identification of laboratory specimens by a qualified taxonomist, and identification of laboratory specimens by molecular analysis (such as DNA barcoding).

[5] It is commonly accepted that it is very difficult to distinguish *E. andrei* from *E. fetida* based on morphological features alone. In a study carried out by Römbke *et al.* (2016), 28 participating laboratories submitted samples of both species to assess whether DNA barcoding could be used to distinguish the species from one another, and to determine which of the two species were being used in toxicity testing. The results indicated that all laboratory participants who submitted *E. andrei* to the study had correct identification of the species, whereas only 56% of the laboratories' *E. fetida* cultures were actually confirmed to be *E. fetida*. The remaining 44% of the worms believed to be *E. fetida* were genetically identified as *E. andrei* (Römbke *et al*., 2016). It is evident from this study that DNA barcoding can be used to distinguish the two species from one another, and that the possibility exists for *E. fetida* contamination of *E. andrei* cultures.

DNA sequences for the 5′ region of the mitochondrial cytochrome c oxidase subunit I gene are available for *D. rubidus* through the Barcode of Life project at the University of Guelph (EC, 2010).

[6] These worms should be obtained from in-house cultures maintained in the testing laboratory using the guidance provided in Section 2.3. In situations where a testing laboratory is unable to provide test organisms using in-house cultures, the worms may be obtained from another source that uses culturing conditions, procedures, and quality assurance consistent with the guidance in Section 2.3. In this situation, however, the worms must be held and acclimated to laboratory conditions within the testing laboratory, according to the guidance in Section 2.4, before their use in these biological test methods. The use of earthworms collected from the field or purchased from a commercial supplier with unknown or no quality assurance or quality control is unacceptable for the purposes of this test method document.

government or private laboratories that are culturing these species of earthworms for soil toxicity tests. These same sources are recommended for establishing in-house laboratory cultures of *E. andrei* or *D. rubidus*.[7] In-house cultures can be established using juvenile or adult worms. Cocoons (rather than juvenile or adult worms) may be obtained to establish a culture more quickly or to standardize the age and weights of individual worms within the culture.

For current information on suppliers for *E. andrei* and *D. rubidus*, contact:

Method Development and Applications Unit
Science and Technology Branch
Environment and Climate Change Canada
335 River Road
Ottawa, Ontario K1A 0H3
Email: methods@ec.gc.ca

All earthworms used in a soil toxicity test must be derived from the same population. Worms to be used as a source of breeding stock or test organisms should be transported to the laboratory using a portion of the soil or other substrate to which they are adapted. Additional quantities of this substrate can be obtained for culturing and holding purposes, depending on culturing (Section 2.3) and *acclimation* (Section 2.4) conditions and requirements. Shipping and transport containers should be insulated to minimize changes in temperature during transit, and the temperature should be maintained at 17–23 °C. Live organisms should be transported quickly to ensure their prompt (i.e., within 24 hours) delivery. Excessive crowding of animals during shipment or transport should be avoided to minimize stress during transit.

Upon arrival at the laboratory, organisms may be held in the soil (or other substrate) used in transit while temperature adjustments are made, or they may be transferred to other culturing substrate (Section 2.3.5) or that for holding and acclimating test organisms (Section 2.4.5). If the nature (including the *texture* and moisture content) of the substrate in which worms were initially held (e.g., by a supplier) or transported differs markedly from that in which they are to be cultured (Section 2.3.5) or acclimated (Section 2.4.5), it is prudent to adapt the worms to an increasing percentage of the new substrate over several weeks until they are held in 100% of this substrate.[8]

Soil temperature should be adjusted gradually (e.g., ≤ 3 °C per day) to the temperature to be used during culturing (Section 2.3.4), or when acclimating the worms to test conditions (Section 2.4.4). Guidance on handling worms given in Sections 2.3.7 and 2.4.7 should be followed when transferring worms from an outside source to culture vessels (Section 2.3.2) or those for acclimating worms (Section 2.4.2). Other conditions during this interim holding period for acclimation of breeding stock or test organisms to laboratory conditions should be as similar as possible to those used for maintaining cultures (Section 2.3) or for acclimating worms obtained for use in tests (Section 2.4).

2.3 Culturing of E. andrei *and* D. rubidus

2.3.1 General

General guidance and recommendations for culturing *E. andrei* and *D. rubidus* in preparation for soil toxicity tests are provided here. In keeping with the premise, *"What might work well for one laboratory might not work as well for another*

[7] Investigators might be concerned with the effects of excessive inbreeding of laboratory cultures, or might wish to use progeny from organisms that were collected from a particular location. Accordingly, cultures can also be established using wild populations or can be genetically enhanced by introducing breeding stock from different sources. If animals are obtained from a wild population or a commercial supplier, their taxonomy must be confirmed and their progeny should be evaluated for sensitivity to reference toxicant(s) before being used in toxicity tests. Ideally, any site from which field-collected specimens are taken should be known to be free of any applications or sources of pesticides or fertilizers during the past five years or longer.

[8] Experience at Environment Canada's Soil Toxicology Laboratory (River Road S&T Science Laboratory, Ottawa, ON) indicates that survival in cultures can be poor if the nature of the substrate used by a commercial supplier to culture *E. andrei* differs markedly from the laboratory's culturing substrate. Survival is markedly improved, in this instance, if the percentage of the laboratory's culturing substrate is increased gradually over several weeks until the earthworms are held in 100% of the culturing substrate (J. Princz, Environment Canada, personal communication, 2004).

laboratory" (EC, 2005b, 2013b, 2014a; ECCC, 2020a), explicit directions regarding many aspects of culturing, including the choice of culture vessel, number of organisms per vessel, soil-renewal conditions, culturing substrate, and food type and ration, are left to the discretion and experience of laboratory personnel, although guidance and recommendations are provided herein. Performance-based indices[9] are used to evaluate the suitability of the cultured organisms for tests, and the acceptability of the test results. Cultures must have low mortalities to be suitable for use in tests, and the cultured organisms must appear healthy and behave and feed normally.[10] Additionally, those used as *controls* in the test must meet all criteria for a valid toxicity test (see Sections 4.2.3 and 4.3.3). The acceptability of the culture is also demonstrated by *reference toxicity tests* or positive controls using a reference toxicant (see Section 4.4). If a culture of organisms fails to meet these criteria, its cause should be investigated. Care must be taken to ensure that each culture is not contaminated with other similar species (i.e., mixed with different worm species). Periodic (e.g., every two years) taxonomic checks of the laboratory's cultures are recommended (see Section 2.1).

It is the responsibility of the laboratory to demonstrate its ability to obtain consistent, precise results using a reference toxicant when initially setting up to perform soil toxicity tests with cultured *E. andrei* or *D. rubidus*. For this purpose, intralaboratory *precision*, expressed as a *coefficient of variation* for the respective endpoint data, should be determined by performing five or more tests with different lots (groups) of test organisms from the same source, using the same reference toxicant and identical procedures and conditions for each test (see Sections 3.2.1 and 4.4).

When routinely performing soil toxicity tests with *E. andrei* or *D. rubidus*, consistency must be demonstrated either through the inclusion of a positive control concentration with each definitive test (Section 4.4) or through reference toxicity tests conducted a minimum of twice per year with the laboratory's cultures, using the conditions and procedures outlined in Section 4.4. Additionally, the performance of any cultures that have been established recently using new breeding stock (Section 2.2) should be checked with a reference toxicity test or positive control, and the results determined to be acceptable (see Sections 2.3.9 and 4.4) before these cultures are used to provide test organisms.

Cultures of *E. andrei* and *D. rubidus* should be observed at regular intervals (e.g., biweekly). Ideally, records should be maintained documenting:

- the date a culture is started and the estimated number of organisms used to start the culture;
- dates of substrate renewal;
- feeding and watering regime (including type and quantity added on each occasion);
- facility and soil quality (e.g., air temperature, *photoperiod* and light quality, pH of substrate); and
- observations of culture health and density (e.g., behaviour and appearance of earthworms in culture, appearance and odour of substrate, number and location of worms in the container, amount of uneaten food in container).

A checklist of required and recommended conditions and procedures for culturing *E. andrei* and *D. rubidus* to generate organisms for use in soil toxicity tests is given in Table 1. Numerous procedural specifics that have presumably worked well in producing adult *E. andrei* for use in soil toxicity tests are reported elsewhere (ISO, 2008, 2012; ASTM 2012; OECD, 2016), and unless indicated otherwise in this report, might provide useful guidance that may also be applied here.

2.3.2 Facilities and Apparatus

Worms should be cultured in a controlled-temperature laboratory facility. Equipment for

[9] Performance-based indices include those related to the survival and condition of cultured organisms intended for use in the test (Section 2.3.9), as well as the criteria that must be met by control organisms for a test to be valid (Sections 4.2.3 and 4.3.3), and those related to the performance of groups of animals in a positive control concentration run concurrently with each definitive test or in reference toxicity tests (Section 4.4).

[10] Unhealthy worms might have physical anomalies such as discolouration (e.g., yellowing), de-clitellation, pinching, and lesions, and/or might demonstrate behavioural changes such as coiling or lethargy.

Table 1 Checklist of required and recommended conditions and procedures for culturing *E. andrei* and *D. rubidus*, to provide test organisms for use in soil toxicity tests

Source of breeding stock for culture	−	juveniles, adults, and/or cocoons from a government or private toxicity testing laboratory; identification to species (i.e., *E. andrei* or *D. rubidus*) confirmed
Acclimation	−	gradually, for temperature (recommend ≤ 3 °C/day) and substrate differences upon arrival
Culture vessels	−	breeding boxes of 6–50-L capacity are suitable (e.g., plastic trays measuring ~ 30 × 40 × 15 cm or ~ 60 × 40 × 20 cm for *E. andrei*, and smaller plastic trays measuring ~ 32 × 17 × 12 cm for *D. rubidus*; covered with perforated lid to allow air exchange and minimize evaporation); sides and/or lid transparent or translucent to enable light to contact surface of culturing substrate; recommended minimum depth, 10 cm
Air temperature	−	daily average, 20 ± 2 °C; instantaneous, 20 ± 3 °C
Lighting	−	incandescent, fluorescent, or LED; intensity, 400–800 lux at surface of culture vessel; continuous dark or fixed photoperiod (e.g., 16h L:8h D or 12h L:12h D)
Type of substrate	−	optional (e.g., mixture of potting soil, artificial soil, and peat moss)
Hydration of substrate	−	hydrated with test water; moisture content sufficient to keep surface of bedding moist but with no standing water in the bottom of the culture vessel; soil particles should not adhere to earthworms
pH of substrate	−	adjusted to range within 6.0–7.5 using reagent-grade calcium carbonate
Renewal or refreshment of substrate	−	as required; sort and transfer worms and cocoons manually; alternatively, use constant light to move worms out of top layer of old bedding into fresh bedding placed underneath or into bottom half of old bedding; remove and discard top layer of old bedding, and replace with fresh bedding
Monitoring substrate quality	−	temperature, pH, and moisture content measured at regular intervals in each culture vessel
Feeding	−	Magic® Worm Food and ground and sieved organic mixed grains are recommended; oatmeal, alfalfa pellets, and/or other optional food supplements; feed at least once biweekly by placing food in a shallow depression of the substrate, hydrating with deionized water, and then covering it with a thin layer of substrate; amounts adjusted based on observations of food consumption; any excess (uneaten) food and any visible mould, fungi, or mites nearby should be removed before feeding
Maintenance of culture	−	examine substrate in culture vessel at regular intervals (e.g., biweekly); gently turn manually as necessary; remove dead, injured, or atypical (lethargic) worms; record condition of culture; maintain loading density of worms at ≤ 0.03 g/cm^3

Age/size for test	–	clitellated adults for both avoidance and reproduction tests; individual wet wt within the size range of 250–600 mg for *E. andrei* and 50–200 mg for *D. rubidus*
Indices of culture health	–	considered healthy if (1) worms move actively through the substrate, do not try to leave it, and reproduce continuously, and (2) results for reference toxicity tests or positive controls using worms from the culture fall within historic warning limits; reproduction data from negative control soils are monitored; discard culture if >20% of juvenile or adult worms are dead, inactive, or unhealthy at any time

[*] The information in this table is for summary purposes only. Definitive requirements and recommendations of this test method are contained in the main body of this document.

---

temperature control (i.e., an incubator or a room with constant temperature) should be adequate to maintain temperature within the required limits (Section 2.3.4). The culturing area should be isolated from any testing, sample storage, or sample-preparation areas to avoid contamination from these sources. It must be designed and constructed to prevent contamination of cultures (e.g., elimination of copper or galvanized piping or fixtures that could drip metal-contaminated condensates).

All equipment, vessels, and accessories that might contact the organisms or substrate within the culturing facility must be clean, rinsed as appropriate, and made of nontoxic materials (e.g., glass, Teflon™, Type 316 stainless steel, nylon, Nalgene®, porcelain, polyethylene, polypropylene). Toxic materials including copper, zinc, brass, galvanized metal, lead, and natural rubber must not come in contact with this apparatus and equipment, or the culturing substrate or water.

A variety of culture vessels, such as plastic trays or breeding boxes of 6–50-litre capacity (e.g., plastic trays measuring ~ 30 × 40 × 15 cm or ~ 60 × 40 × 20 cm for *E. andrei*, and smaller plastic trays measuring ~ 32 × 17 × 12 cm for *D. rubidus*), are suitable for culturing *E. andrei* and *D. rubidus*. The sides and/or lid should be translucent or transparent to enable light to contact the surface of the culturing substrate (see Section 2.3.3). Each vessel should have a perforated (e.g., holes covered with fibreglass mesh screening) lid to minimize drying of the surface substrate and the risk of contamination, while allowing air exchange and preventing worms from escaping. The use of culture vessels constructed of wood is not recommended, due to the possible presence of toxic contaminants (e.g., plywood glues, antisapstain chemicals, or wood extractives such as resin acids and juvabiones). The choice of size and numbers of culture vessels required might be influenced by the number of adult earthworms required by the testing facility for one or more series of soil toxicity tests. Each culture container should accommodate a minimum depth of 10 cm of soil or other culturing substrate.

2.3.3 Lighting

E. andrei and *D. rubidus* may either be cultured in the dark or with incandescent, fluorescent, or *light-emitting diode* (*LED*) lights and a regulated photoperiod (e.g., 16 h light:8 h dark, or 12 h light:12 h dark). If lights are used, light intensity adjacent to the top of the culture vessels should range within 400–800 *lux*. This range is equivalent to a quantal flux of 5.6–11.2 μmol/(m^2 · s) for cool-white fluorescent, 6.4–12.8 μmol/(m^2 · s) for full-spectrum fluorescent, or 7.6–15.2 μmol/(m^2 · s) for incandescent. The lights should be positioned sufficiently far from the culture vessels to prevent evaporation caused by heat buildup.

2.3.4 Temperature

E. andrei and *D. rubidus* should be cultured in a facility with an air temperature of 20 ± 2 °C as a daily average. Additionally, the instantaneous temperature of the facility should be 20 ± 3 °C.

2.3.5 Culturing Substrate

The choice of substrate for culturing *E. andrei* and *D. rubidus* is left to the discretion and experience of laboratory personnel; however, the following culturing substrate is recommended.

A mixture of potting soil (manure/peat/loam blend),[11] sieved (2-mm mesh) *Sphagnum* sp. peat moss, and artificial soil has proven to be a suitable culturing substrate for both *E. andrei* and *D. rubidus*. A 10-L *batch* of this mixture is prepared as follows:

1. Mix ~ 3 L of potting soil with ~ 4 L of peat moss (both in their "dry form").

2. Then add *deionized* water (~1 L) to the substrate and mix mechanically (handheld mixer) until the moisture content, colour, and texture of this mixture appear to be homogeneous.

3. Thereafter, add ~1.5 L of artificial soil (see Section 3.3.2).

4. Then add deionized water (~1 L) to this mixture while stirring mechanically until a moisture content equivalent to ~70% of the water-holding capacity of the mixture is achieved.

5. Then measure the soil pH, and depending on the value, sprinkle ~30 g of calcium carbonate ($CaCO_3$) on the surface of the culturing substrate using a fine sieve, and mix into the soil using a mechanical mixer until no white powder is visible.

This mixture is stored in a covered container at ambient laboratory temperature for a minimum of three days. The culturing substrate is then stirred and its pH measured to ensure that it is between 6.0 and 7.5. If the pH is below 6.0, additional calcium carbonate is added (G.L. Stephenson, Aquaterra Environmental, personal communication, 2001).

2.3.6 Food and Feeding

Success in culturing *E. andrei* and *D. rubidus* has been achieved using Magic® Worm Food, ground and sieved organic mixed grains (MESI, 2014, 2020; ECCC, 2020b), cooked oatmeal (Aquaterra Environmental, 1998; Stephenson *et al.*, 1999a, 1999b; Aquaterra Environmental and ESG, 2000; EC, 2010, 2013a), or hydrated alfalfa pellets (ASTM, 2012; USEPA, 2012), with optional food supplements (e.g., kitchen compost, fortified baby cereal, organic composted manure). Details for preparing these recommended food types, together with acceptable feeding regimes, are described in this section. Cultures must be fed a minimum of once every two weeks. During culture feeding, any old food accompanied by mould, fungi, or mites appearing on the surface of the culturing substrate should be removed and discarded.

Magic® Worm Food (MWF) is a commercially available worm food used by Canadian toxicity testing laboratories for both culturing and testing worms, and is recommended herein as a primary food source for culturing and testing (56-day reproduction test; see Section 4.2.4).[12, 13] MWF should be processed prior to use to remove any potential contamination of cultures by other organisms. Success in this regard has been achieved by drying it in the oven at 60 °C for 48 hours and then storing it in the freezer (e.g., -20 °C). It can be used directly from the freezer to feed cultures (or test organisms). Using a spoon, ~5–10 mL of MWF should be added to each culture container by making a small depression in the culturing substrate,

---

[11] The potting soil must not contain any added fertilizers, vermiculite, or perlite. In addition, a freeze/thaw cycle is required in order to prevent the introduction of indigenous organisms to the cultures. The freeze-thaw cycle is described in footnote 110 in Section 5.3.

[12] Magic® Worm Food is a commercially available food used for growing worms as fishing bait. It is available from Magic Products Inc. (1-715-824-3100; or https://magicproducts.com/products/). It contains 32 different proteins, fats, minerals, vitamins, and carbohydrates. Listed ingredients include shelled corn, oats, wheat middlings, lime, alfalfa meal, and soybean meal. The exact formulation is proprietary, and it has been found to vary in composition from batch to batch (ECCC, 2020b). Chemical analyses can be performed on each new batch to determine nutritional composition, and to screen for any potential contamination (e.g., pesticides).

[13] In an ECCC laboratory investigations into sources of variability in juvenile production in the earthworm survival and reproduction test, it was determined that using Magic® Worm Food as a food source for culturing and testing both *E. andrei* and *D. rubidus* led to enhanced worm health, higher juvenile production, and fewer test failures than cultures and tests fed with oatmeal (ECCC, 2020b). Earthworm cultures in ECCC's Soil Toxicology Laboratory were changed from being fed solely oatmeal to alternating biweekly from oatmeal to MWF. In addition, testing was performed using MWF as the only food source instead of oatmeal.

inserting the food, hydrating with a small quantity of deionized water, and then covering it with a thin layer of substrate (J. Princz, Environment and Climate Change Canada, personal communication, 2020). The amount of food should be adjusted based on observations of food consumed or not consumed during the preceding feeding event.

Organic mixed grains are also recommended for use as a primary food source for culturing and testing earthworms, in addition to, or as an alternative to, MWF.[14] Like MWF, it is recommended that organic mixed grains be processed prior to use to remove potential contamination of cultures by other organisms. In addition, the organic mixed grains must be ground and sieved (1–2 mm diameter), and then frozen until use. Organic mixed grains can be used to feed cultures following the guidance provided above for MWF.

Cooked oats can be used as a food for feeding both *E. andrei* and *D. rubidus* cultures.[15] Oatmeal prepared from Quaker® Oats is recommended for this purpose since experience with generic or other brands of oatmeal has sometimes indicated problems with respect to excessive mould production in the cultures or during the test. The oatmeal should be hydrated with deionized water (e.g., 1:2 volume of oatmeal to boiled deionized water) and cooled before feeding to cultures. Using a spoon, ~5–10 mL of cooked oatmeal should be added to each culture container by making a small depression in the culturing substrate, inserting the food, and then covering it with a thin layer of substrate to minimize mould growth or the proliferation of mite populations (G.L. Stephenson, Aquaterra Environmental, personal communication, 2001). If removing any old (uneaten) oatmeal, care should be taken to not remove any worms (hatchlings tend to burrow into the oatmeal bolus).

Alfalfa pellets may also be used as a food source for cultured *E. andrei* (ASTM, 2012; USEPA, 2012).[16] Dried pellets can be obtained from agricultural feed and supply stores. Before using, the pellets should be saturated with deionized or *distilled water* (at a ratio of ~1 g of dry pellet per 2 mL water). Although USEPA (2012) recommends that the hydrated alfalfa should be aged for a minimum of two weeks in a covered container, experience at one of Environment Canada's testing laboratories indicates that aging is not necessary and that hydrated alfalfa may be used within hours of hydrating (D. Moul, Environment Canada, personal communication, 2001). At the time of each feeding, any uneaten food observed on the surface of the bedding should be removed with a spoon or forceps, and discarded. Fresh, hydrated alfalfa food is then transferred to the surface of the bedding and covered with a thin layer of bedding substrate to minimize growth of parasites (mites and springtails) (D. Moul, Environment Canada, personal communication, 2001).

The feeding of earthworm cultures with MWF, organic mixed grains, oatmeal, or alfalfa may be supplemented with regular additions of small quantities of composted vegetable matter or composted manure, to improve and sustain the health of the earthworms. The addition of dehydrated compost (e.g., at a rate of 15–30 mL per culture bin containing ~6–8 L of substrate) can be used to supplement the biweekly feeding of MWF,

[14] ECCC's Soil Toxicology Laboratory investigated the use of organic mixed grains as an alternative food source, comparable to MWF, since MWF is commercially produced and only available through one manufacturer. The mixed grains used by ECCC were Dodd's and Erwin organic mixed grains, obtained through Gilmore's Feed Barn in Metcalfe, ON. It was a mixture of full grains consisting of 14% protein, barley, oats, wheat, roasted corn, and roasted soybeans. MWF and the organic mixed grains had similar nutritional value differing only in mineral content. Testing conducted with *E. andrei* to evaluate potential differences between the two food sources indicated that the organic mixed grains appeared to be a valid alternate food source to MWF for worm culturing and testing (EC, 2020b).

[15] Oatmeal is no longer recommended as a primary food source for culturing and testing earthworms (see footnote 13). ECCC's Soil Toxicology Laboratory has recommended that providing a mixed diet for cultures by alternating between MWF and organic mixed grains and/or oatmeal is beneficial for maintaining healthy cultures of both *E. andrei* and *D. rubidus* (J. Princz, Environment and Climate Change Canada, personal communication, 2020; ECCC, 2020b).

[16] This food source has been demonstrated to enable cultures of *E. andrei* to thrive (ASTM, 2012; USEPA, 2012). ECCC's Soil Toxicology Laboratory has no experience with using alfalfa pellets as a food source for *E. andrei* or *D. rubidus* (ECCC, 2020b).

organic mixed grains, cooked and hydrated oatmeal, or alfalfa.[17]

The quantity of food added depends on worm density and developmental stage. The amount of food added to each culture vessel should be based on observations and records of food consumed or not consumed, during preceding feedings.

2.3.7 Handling Organisms and Maintaining Cultures

The embryonic (in cocoons), juvenile, and adult life stages of *E. andrei* and *D. rubidus* should be handled as little as possible, to avoid damage and undue stress. When handling is necessary, it should be done gently, carefully, and quickly to minimize stress to the animals. The use of a gloved hand and/or the arm(s) of rounded forceps are suitable for moving worms to and from culture or test vessels. When handled, any animals that are dropped, injured, or appear stressed should be discarded, and must not be used for testing.

It is recommended that the contents of each culture vessel be inspected at regular intervals (e.g., just before feeding) to determine the apparent condition of the worms and the bedding substrate. If, during this inspection, any excess water is observed to have accumulated at the bottom of the substrate, the bedding within the culture vessel should be turned carefully at this time to redistribute the excess water throughout the culturing substrate.[18] Care must be taken while turning the vessel's contents, to prevent injuries to the earthworms. Any dead worms observed at these times must be removed and discarded. Any injured or apparently atypical (e.g., lethargic) worms observed should also be removed and discarded. Records should be kept of the apparent condition of the culture (worms and substrate) noted during each observation period (Section 2.3.1).

The loading density of worms in each culture vessel should be restricted to prevent overcrowding and the resulting adverse effects on worm growth, reproduction, and culture health. A maximum loading density of 0.03 g wet wt/cm^3 recommended by ASTM (2012) provides useful guidance in this respect. To reduce the number of worms in a crowded culture vessel, either of the following procedures (or some suitable modification thereof) is recommended and should be applied. The first option provides the added advantage of sorting worms into two size classes (i.e., juveniles and adults).

Option 1 (as per ESG International Inc., Guelph, ON) (G.L. Stephenson, Aquaterra Environmental, personal communication, 2001):

---

[17] To prepare dehydrated compost, vegetable/fruit materials (free of meat, dairy, bread, coffee grounds, tea leaves, large pits, and tough peels like banana peels) are placed into a stainless steel pail that is used to collect compostable material from office lunch rooms or households. As needed, this material is taken to the laboratory and pulverized with a food processor. The addition of water (deionized or reverse osmosis) to the food processor containing the compostable material might be necessary if the material is too dry to pulverize; however, this occurs infrequently. If the compost is too wet, it can be sieved prior to use. The pulverized material is then placed onto aluminum trays, and distributed to form a thin layer. The trays with the compost can be further processed using a freeze/thaw cycle in order to eliminate potential contaminating insects. Alternatively, the trays with compost can be placed into a drying oven (90–105 °C) to dry overnight. The next day, the dried compost (with the consistency of dried pabulum) is placed into a food container, and stored in the refrigerator until used (within seven days after preparation). If composted manure is used as a supplement, it should be dried at 60 °C for ≥ 2 days and sieved to 4 mm prior to use (J. Princz, Environment and Climate Change Canada, personal communication, 2020). The dehydrated composted vegetable material or composted manure is sprinkled on the surface of the substrate in each culture vessel. It is not necessary to rehydrate this material if the culturing substrate is sufficiently moist (Stephenson, 2003b).

[18] An alternate approach for redistributing excess water throughout the culturing substrate is to invert culture vessels (with lids in place) at regular intervals, as needed, for a minimum of one hour. This approach is less labour intensive than turning the substrate in each culture vessel, and may be applied for this purpose. A disadvantage of this procedure is that it does not enable concurrent observations of earthworms that are evident when turning the contents of a culture vessel by hand. If this procedure is followed without turning and the contents of the culturing substrate is not turned manually, gentle stirring of the surface of the substrate on a regular basis is recommended to minimize the proliferation of populations of mites. The use of culture vessels with small mesh-covered holes in the bottom will prevent the buildup of any excess water in the bottom of the vessels.

1. Prepare a fresh mixture of culturing substrate (see Section 2.3.5).

2. Thereafter, transfer an aliquot of ~1 L of fresh substrate to each of two temporary holding containers.

3. Place the contents of an old (crowded, too wet, or foul smelling) culture vessel onto a plastic sheet or in a shallow plastic container with sufficient surface area to allow the contents to be sorted.

4. Remove live and apparently healthy juvenile and adult worms, and transfer them to the two temporary holding containers as two size classes (i.e., juveniles in one container, adults in the other).

5. Prepare two new culture vessels, by mixing in a portion of the old substrate and new substrate in a ratio of 1 part old:3 parts new. After mixing, adjust the moisture content and pH of the substrate in each of these culture vessels as required (Section 2.3.5, paragraph 2).

6. Transfer the juvenile worms in one of the two temporary holding containers to the surface of the substrate in one of these two culture vessels, and the adult worms to the other.

7. Gently distribute individual worms evenly over the surface area, so that they enter the substrate throughout the vessel.

8. Label each culture vessel and record the species, life stage, source of worms, approximate number of individuals per vessel, and the date that the substrate was renewed.

Option 2 (based on ASTM, 2012):

1. Set up a culture vessel with new (freshly prepared) substrate but no worms, and place half of its contents onto a plastic sheet.

2. Transfer the contents of the crowded culture vessel to a separate plastic sheet.

3. Then, carefully remove the worms (including cocoons, juveniles, and adults) from the substrate, and transfer equal numbers and age classes (approximately) temporarily to each of two suitable transfer containers.

4. Thereafter, transfer half of the old substrate to the new culture vessel, and mix the contents gently using gloved hands or a spatula or plastic spoon.

5. Then, mix the half of the new substrate on the plastic sheet with the remaining half of the old substrate, and transfer the mixture to the previously crowded culture vessel.

6. Thereafter, transfer one of the two groups of worms held briefly in each transfer container to each of the two freshly prepared culture vessels.

The culture bedding should be replenished (by adding additional fresh substrate) or renewed (i.e., replaced) on an as-needed basis, based on the quality of the substrate, and the density and health of the earthworms.[19] An efficient procedure to achieve this (ASTM, 2012) is to prepare a new tray of bedding, and place the contents of the old bedding (including the worms therein) on top of the new bedding. Hold the stacked (old on new) bedding in an uncovered tray at 20 ± 2 °C under continuous illumination for two days, to encourage the worms to burrow into the new bedding. At the end of the two-day period, remove the old bedding from the new bedding and

[19] In an ECCC investigation, it was determined that less frequent bedding changes over the long-term resulted in more consistent *E. andrei* juvenile production in toxicity tests. It was recommended that the need for culture bedding changes be based on the quality of the culture substrate and density and health of the earthworms. Signs of deteriorating substrate include differences in colour between the bottom layer and upper few centimeters, extremely wet or heavy substrate, and/or a strong odour indicative of anaerobic conditions. A lack of cocoons or clitellated adults in the culture, or the presence of worms on the sides and lid of the culture vessels might also be indicative of deteriorating substrate quality (ECCC, 2020b).

discard it.[20] Alternatively, continuous illumination can be used to encourage worms to move away from the top layer of the culture bedding. After a minimum of 2 hours, the top layer of bedding is removed, screened for worms remaining in the bedding, and discarded. Fresh bedding is then added to the surface (P. Boyd, Environment and Climate Change Canada, personal communication, 2021).

The pH, temperature, and moisture content of the bedding in each culture vessel should be monitored at regular intervals, and adjustments made as and if necessary (see Sections 2.3.4 and 2.3.5).

2.3.8 Cultured Worms for Toxicity Tests

To be successful, the culturing procedures used must produce the required number of healthy test organisms of a known developmental stage and similar size. The wet weight of individual worms used to initiate either of the soil toxicity tests described in Sections 4.2 or 4.3 must be within the range identified in Section 2.1. Additionally, the cultured organisms must meet specific health and performance-related indices (Section 2.3.9). Age-synchronized cultures may be used for either the avoidance test or the reproduction test.[21]

It is highly recommended that laboratory-cultured *E. andrei* or *D. rubidus* used to start a 56-day toxicity test for effects on reproduction (Section 4.2) be acclimated in the laboratory to conditions for this toxicity test, for ≥ 7 days (or ≥ 14 days, if worms are supplied by another laboratory's culture for use in a 56-day reproduction test; see Section 2.4.8). If the culturing substrate used is essentially soil (or a mixture of soil and peat moss; see Section 2.3.5), and the food provided to cultures is the same as that used in the reproduction test (i.e., Magic® Worm Food or organic mixed grains; see Sections 2.3.6 and 4.2.4), then acclimation to these test conditions has been achieved and any additional transfer and handling of worms for this purpose is not advised. However, if culturing conditions of substrate and/or food differ appreciably from those/that to which worms in the negative control soil will be exposed during a reproduction test, all worms to be used in the toxicity test should be acclimated for ≥ 7 days (see Section 2.4) in negative control soil. During this acclimation period, lighting and temperature conditions should be the same as those to be used in the 56-day reproduction test, and worms must be fed the same food (i.e., MWF or organic mixed grains) as that to be used in the test (see Sections 2.4., 2.4.3, 2.4.4, 2.4.6, and 4.2.2).

For the avoidance test, it is highly recommended that any laboratory-cultured *E. andrei* or *D. rubidus* used to start a test be acclimated in the laboratory to the substrate and temperature conditions representing those in this toxicity test for ≥ 7 days (see previous paragraph and Sections 2.4.4, 2.4.5, and 4.3.2). Worms to be used in an avoidance test, however, need not be acclimated beforehand to the conditions of complete darkness that occur throughout this test (see Sections 2.4.3 and 4.3.2).

2.3.9 Health and Performance Indices

Each culture vessel should be checked as needed to monitor and record culture performance (see Sections 2.3.1, 2.3.6, and 2.3.7). Procedures and conditions used to maintain each culture should be evaluated routinely, and adjusted as necessary to maintain or restore the health of the culture. Any juvenile or adult worms that appear to be dead, inactive, not burrowing in the bedding substrate, or otherwise unhealthy or atypical should be discarded.

---

[20] This procedure does not recover the cocoons, and some of the juvenile and adult worms will likely remain in the old bedding (ASTM, 2012).

[21] In recent ECCC studies, age-synchronization of cultures was investigated as a possible means for reducing the variability of juvenile production during testing for both *E. andrei* and *D. rubidus* (ECCC, 2020b). Synchronized cultures were created and tested using artificial soil in side-by-side tests with asynchronous cultures. Although age-synchronized cultures produced higher and less variable numbers of juveniles in negative control soil than the standard non-synchronized cultures, the continued use of a given synchronized culture over time led to lower and more variable reproduction results. Therefore, if age-synchronized cultures are to be used for testing, it is recommended that a given age-synchronized culture be used only once to initiate a test or tests set up at a given time, and that any additional worms left over should be returned to the general culture. Also, it was determined that age-synchronized *D. rubidus* cultures could only be used for testing for a limited amount of time (i.e., 180 days after initiation of the age-synchronization). A similar drop in reproduction and increase in variability have been observed for worms used from these "older" age-synchronized cultures (ECCC, 2020b).

If the culture appears unhealthy or atypical during any check, it should then be checked more frequently to make sure that "cascade mortality" (i.e., rate of death increasing exponentially over time) is not occurring. If more than 20% of the juvenile or adult worms in a culture vessel appear to be dead, inactive, or unhealthy during any period of observation, the entire group in the container should be discarded. Also, if the combined number of mortalities and apparently unhealthy worms observed on the surface of the culturing substrate persists or increases over time, the contents of the culture vessel should be discarded.

There are two possibilities for meeting minimum QA requirements for assessing test organisms' sensitivity using a known reference substance (e.g., boric acid) for the reproduction and avoidance tests. The first option is to conduct two multi-concentration reference toxicity tests annually (i.e., once every six months) using worms derived from the same culture(s) of earthworms from which the test organisms are obtained for definitive testing (see Section 4.4). The second option is to include a positive control concentration with each toxicity test using a portion of the adult worms from those used for the definitive toxicity test (see Section 4.4 for details).[22] All tests with the reference toxicant(s) must be performed using the conditions and procedures outlined in Section 4.4. Test-related criteria used to judge the validity of a particular soil toxicity test (and, indirectly, the health of the culture), based on the performance of test organisms in the negative control soil, are given in Sections 4.2.3 and 4.3.3.

A laboratory that routinely performs reproduction toxicity tests (Section 4.2) might find it useful to monitor the data on number of juveniles produced in negative control soil, as a measure of culture health and performance. A plot of such data over time might show problems with respect to reproductive success that are attributable to diet or other conditions to which cultures are exposed (G. Stephenson, Stantec Consulting Ltd., personal communication, 2004).

2.4 Acclimation of E. andrei *and* D. rubidus

2.4.1 General

It is highly recommended that any group of earthworms (*E. andrei* or *D. rubidus*) to be used in either of the soil toxicity tests described herein first be acclimated to the laboratory conditions to which they will be exposed during the test(s). Procedures and conditions for the acclimation of any group of *E. andrei* or *D. rubidus* cultured in-house or transported to the laboratory from another laboratory's culture for their use in either a 56-day reproduction (Section 4.2) or 48-hour avoidance test (Section 4.3) are described here, and summarized in Table 2. Guidance on sources of earthworms to be delivered to a testing laboratory for use in either test is provided in Section 2.2. Refer to Section 2.3 for guidance on conditions and procedures for culturing earthworms (*E. andrei* or *D. rubidus*) to be used in either the avoidance or reproduction tests.

As with initial tests using earthworms cultured in the testing laboratory (see Sections 2.3.1 and 3.2.1), it is the responsibility of each laboratory not experienced with the biological test method(s) described in this document to demonstrate its ability to obtain consistent, precise results using a reference toxicant when initially setting up to perform avoidance or reproduction tests with groups of earthworms (*E. andrei* or *D. rubidus*) obtained from another laboratory's earthworm culture. For this purpose, intralaboratory precision, expressed as a coefficient of variation for the appropriate endpoint data, should be determined by performing five or more tests with different lots (groups) of test organisms from the same supplier, using the same reference toxicant and identical procedures and conditions for each test (see Section 4.4). The laboratory should also confirm its test precision at this time by conducting five or more toxicity tests using negative control soil and different lots of test organisms (EC, 2005b, 2013b, 2014a; ECCC, 2020a). The conditions and

[22] It is highly recommended that laboratories conduct reference toxicity testing relevant to the endpoints being measured in a definitive test. For laboratories that conduct avoidance testing infrequently, however, both of the reference toxicity testing options for the 56-day reproduction test (i.e., the multi-concentration test or the positive control run with each test) may be used to provide information on the health and sensitivity of the culture to satisfy the reference toxicity testing requirements for the 48-hour avoidance test, in addition to the 56-day reproduction test (see Section 4.4).

Table 2 Checklist of required and recommended conditions and procedures for acclimating *E. andrei* and *D. rubidus*, to provide test organisms for use in soil toxicity tests

Source of worms	−	government or private toxicity testing laboratory who has an existing culture of the earthworm species; all from the same source
Life stage and size on receipt	−	depending on timing of toxicity test, worms may be obtained as juveniles or as sexually mature worms with clitellum; individual wet weight within the indicated range (Section 2.1)
Vessel(s) for holding and acclimation	−	6–50-L "breeding" boxes (e.g., plastic trays measuring ~ 30 × 40 × 15 cm or ~ 60 × 40 × 20 cm for *E andrei*, and smaller plastic trays measuring ~ 32 × 17 × 12 cm for *D. rubidus*), covered with perforated lid to allow air exchange and minimize evaporation; sides and/or lid transparent or translucent to enable light to contact surface of substrate; recommended minimum depth, 10 cm
Air temperature	−	recommend adjusting gradually (e.g., ≤ 3 °C/day) for temperature differences upon arrival; thereafter, maintain at a daily average temperature of 20 ± 2 °C and an instantaneous temperature of 20 ± 3 °C; acclimate to these conditions for ≥ 7 days immediately preceding the test
Lighting	−	incandescent, fluorescent, or LED; intensity, 400–800 lux at surface of holding/acclimation vessel; fixed photoperiod (e.g., 16 L:8 D or 12 L:12 D); for acclimation to the avoidance test, worms may be held in the dark; acclimate to these conditions for ≥ 7 days immediately preceding the test
Type of substrate	−	options include: negative control soil (natural or artificial) or a mixture of potting soil, artificial soil, and peat moss
Hydration of substrate	−	hydrated with test water; moisture content sufficient to keep surface of bedding moist but with no standing water in the bottom of the holding/acclimation vessel; soil particles should not adhere to earthworms
pH of substrate	−	near neutral; no adjustment if natural (field-collected) negative control soil; adjusted to range within 6.0–7.5 using reagent-grade calcium carbonate if necessary
Renewal or refreshment of substrate	−	as required; if worms held for an extended period before use in soil toxicity test; sort and transfer worms manually; alternatively, use constant light to move worms out of top layer of old bedding into fresh bedding placed underneath or into bottom half of old bedding; remove and discard top layer of old bedding, and replace with fresh bedding
Duration of acclimation	−	≥ 7 days during the period immediately preceding the test, to laboratory conditions; ≥ 14 days highly recommended for earthworms obtained from an outside source for use in the 56-day reproduction test
Monitoring substrate quality	−	temperature, pH, and moisture content measured at regular intervals in each holding/acclimation vessel

Feeding	–	either Magic® Worm Food or ground and sieved organic mixed grains provided upon establishment of acclimation vessel and then at least once biweekly by placing in a shallow depression of the substrate, hydrating with deionized water, and then covering with a thin layer of substrate; any excess (uneaten) food and any visible mould, fungi, or mites nearby should be removed before feeding
Maintenance	–	examine substrate at regular intervals (e.g., biweekly) during acclimation; rehydrate and gently turn substrate in holding/acclimation vessels manually, as necessary; remove dead, injured, or atypical (lethargic) worms; record apparent condition of substrate and worms; maintain loading density of worms at ≤ 0.03 g/cm^3
Age/size for test	–	clitellated adults for both avoidance and reproduction tests; individual wet wt within the size range of 250–600 mg for *E. andrei* and 50–200 mg for *D. rubidus*
Health indices	–	worms in holding/acclimation vessel(s) considered healthy if (1) they appear to be active when observed, and do not try to leave the substrate, and (2) results for reference toxicity tests or positive controls using worms from the original culture fall within historic warning limits; discard entire group if >20% of juvenile or adult worms are dead, inactive, or unhealthy at any time

[*] The information in this table is for summary purposes only. Definitive requirements and recommendations of this test method are contained in the main body of this document.

procedures used to perform these initial tests with negative control soil should be identical and according to Section 4.2 (if tests for effects of exposure on reproduction are intended) or Section 4.3 (if avoidance tests are intended).

2.4.2 Facilities and Apparatus

Worms should be held and acclimated in a controlled-temperature laboratory facility isolated from any testing, sample storage, or sample-preparation areas. See Section 2.3.2 for further guidance on holding/acclimation facilities and suitable containers (i.e., culture vessels) and lids for acclimating worms to be used in soil toxicity tests.

2.4.3 Lighting

Incandescent, fluorescent, or LED lights should illuminate the vessel(s) used to acclimate worms to be used in 56-day reproduction soil toxicity tests. Photoperiod should be regulated (e.g., 16 h light and 8 h dark, or 12 h light and 12 h dark) and should be the same as that used in the test. Light intensity adjacent to the top of the holding/acclimation vessel(s) should range within 400–800 lux. This range is equivalent to a quantal flux of 5.6–11.2 μmol/(m^2 · s) for cool-white fluorescent, 6.4–12.8 μmol/(m^2 · s) for full-spectrum fluorescent, or 7.6–15.2 μmol/(m^2 · s) for incandescent. Worms being acclimated for use in 48-hour avoidance tests may be held in continuous dark. Worms should be acclimated to these lighting conditions for ≥ 7 days immediately before being used in a test.[23]

2.4.4 Temperature

The air temperature in the holding facility should be 20 ± 2 °C as a daily average throughout the acclimation period. Additionally, the instantaneous temperature should be 20 ± 3 °C throughout this period. An incubator or temperature-controlled room isolated from the testing facility should be used to achieve this. If necessary, worms should be adjusted gradually (e.g., ≤ 3 °C/day) to the acclimation temperature.

Upon the receipt of worms at the testing laboratory, the temperature of the substrate within the transport

[23] This acclimation period is recommended to provide a minimum number of days (≥ 7) for recovery from any stress due to transfer to the testing laboratory, before they are used in a toxicity test.

container should be measured and recorded. For groups of cultured worms transferred to a testing laboratory from another laboratory, the temperature of the substrate and worms therein should be adjusted gradually (e.g., ≤ 3 °C/day) to the acclimation temperature; these worms should be acclimated to the mean test temperature (i.e., 20 ± 2 °C) for ≥ 7 days immediately preceding their use in any toxicity test.

2.4.5 Substrate

Bedding material for acclimating earthworms, in preparation for soil toxicity tests, may be the same substrate as that intended to be used as negative control soil in the test(s). This may be either natural, field-collected soil from an uncontaminated site (Section 3.3.1) or artificial soil (Section 3.3.2). Alternatively, the bedding material recommended for culturing *E. andrei* and *D. rubidus* (Section 2.3.5) may be used for acclimating the earthworms. The moisture content of the substrate should be sufficient to keep the bedding moist, while not causing water to pool in the bottom of the acclimation vessel. Adjustments for moisture content might be necessary[24] (see Section 2.3.5).

The pH of substrate used to acclimate test organisms should be near neutral, so that it is not stressful to them. Adjustments may be made (see Section 2.3.5), as necessary, to bring the pH of the bedding material into a suitable range (i.e., 6.0–7.5).

Worms should be acclimated to these substrate conditions for ≥ 7 days immediately before being used in a test.

2.4.6 Food and Feeding

Worms placed into one or more vessel(s) for acclimating must be fed at that time, using either Magic® Worm Food or ground and sieved organic mixed grains (see Section 2.3.6). Those held for periods of longer than two weeks must be fed a minimum of once every two weeks. Guidance in Section 2.3.6 for food preparation and feeding should be followed.

2.4.7 Handling and Maintaining Organisms

Guidance in Section 2.3.7 applies when handling and maintaining worms held in the laboratory before their use in toxicity tests. If the acclimation period exceeds two weeks, the contents of each acclimation vessel should be manually turned, as needed (e.g., just before each feeding event). At this time, the apparent condition of the bedding substrate and the worms should be observed and recorded. Any dead, injured, or apparently atypical (e.g., lethargic) worms observed should be removed and discarded.

The loading density of worms in each acclimation vessel should be restricted to prevent overcrowding and the resulting adverse effects on worm condition, performance, and health. The maximum loading density of 0.03 g wet wt/cm^3 recommended by ASTM (2012) for cultures of *E. andrei* and *D. rubidus* (see Section 2.3.7) should be used as a guide in this respect. If the period for holding earthworms in the laboratory is extended (e.g., several months), and worm density increases during this time, overcrowding should be prevented by splitting the batch in a holding/acclimation vessel and or refreshed by adding new bedding material (see Section 2.3.7).

At the start of the acclimation period, the temperature, pH, and moisture content of the substrate in each holding/acclimation vessel should be measured and recorded. Regular measurements of each of these soil quality variables should be made if the acclimation period extends beyond two weeks, and adjustments made as and if necessary.

2.4.8 Acclimated Worms for Toxicity Tests

All earthworms used in a soil toxicity test must appear healthy, and be of similar size. Additionally, it is highly recommended that they be held and acclimated according to the procedures and conditions described herein (Sections 2.4.1 to 2.4.7, inclusive). Each worm to be used in an avoidance or reproduction test must have a wet weight ranging within that identified for each species in Section 2.1. Animals used in a toxicity test must satisfy specific health and performance-related indices (Section 2.4.9). Conditions and procedures described in

[24] If soil particles are observed to be adhering to the worms, the soil is too dry and its moisture content should be increased.

Section 2.3.8 apply when acclimating worms for use in an avoidance or reproduction test.

Any earthworms (*E. andrei* or *D. rubidus*) obtained from another laboratory's culture, for initiating a 56-day reproduction test (Section 4.2), must be acclimated for ≥ 7 days immediately before the test, however longer (i.e., ≥ 14 days) is highly recommended. If earthworms to be used in a 48-hour avoidance test (Section 4.3) are obtained from an outside source, they must be acclimated for ≥ 7 days, immediately before the test. Applicable guidance provided in Sections 2.3.8 and 2.4 must also be followed.

2.4.9 Health and Performance Indices

Each holding/acclimation vessel should be checked at regular intervals, during which time the condition of the worms and substrate therein should be monitored and recorded (see Section 2.4.7). Procedures and conditions used to maintain the worms in each holding/acclimation vessel should be evaluated routinely, and adjusted as necessary to optimal levels. Any juvenile or adult worms that appear to be dead, inactive, not burrowing in the bedding substrate, or otherwise unhealthy or atypical, should be discarded. If more than 20% of the juvenile or adult earthworms in a holding/acclimation vessel appear to be dead, inactive, or unhealthy during any period of observation, the entire contents of the container should be discarded.

The QA requirements for assessing test organism sensitivity using a known reference substance described in Section 2.3.9 must also be applied to organisms obtained from another laboratory's culture for use in a test.[25] All tests with reference toxicant(s) must be performed using the conditions and procedures outlined in Section 4.4.

Criteria used to judge the validity of a particular soil toxicity test (and, indirectly, the health of the population of acclimated worms), based on the performance of test organisms in the negative control soil, are given in Sections 4.2.3 and 4.3.3.

[25] If the positive control concentration option is chosen to satisfy the requirements for testing with a reference toxicant (see Section 4.4), it must be conducted in the testing laboratory, concurrently with each definitive test. If, however, the multi-concentration reference toxicity test option is chosen, these tests may be carried out at the laboratory culturing the earthworms and the results provided to the testing laboratory with each shipment of test organisms. It is the responsibility of the testing laboratory to ensure that worms to be used in a test are obtained from a supplier that maintains an ongoing QA/QC program to provide healthy test organisms.

Section 3

Test System

3.1 Facilities and Apparatus

Tests must be performed in an environmental chamber or equivalent facility having acceptable temperature and lighting control (see Sections 4.2.2 and 4.3.2). The test facility should be well ventilated to prevent personnel from being exposed to harmful fumes, and it should be isolated from physical disturbances or any contaminants that might affect the test organisms. The area used to prepare *test soils* should contain a fume hood and be properly ventilated.

The test facility should be isolated from the area where the worms are cultured (Section 2.3) or held and acclimated (Section 2.4), to avoid potential contamination. Additionally, the test facility should be removed from places where samples are stored or prepared, to prevent the possibility of contamination of test vessels and contents from these sources. The ventilation system should be designed, inspected, and operated to prevent air within the testing facility from contaminating the culturing or holding/acclimation facilities. Return air from sample handling and storage facilities or those where chemicals are processed or tested should not be circulated to the area of the laboratory where tests are conducted.

Any construction materials that might contact the organisms, soil, water, or test vessels within this facility must be nontoxic (see Section 2.3.2) and should minimize sorption of chemicals. Borosilicate glass, nylon, high-density polyethylene, high-density polystyrene, polycarbonate, fluorocarbon plastics, Teflon™, Nalgene®, porcelain, fibreglass, and Type 316 stainless steel should be used whenever possible to minimize chemical sorption and leaching. The use of toxic materials including copper, zinc, brass, galvanized metal, lead, and natural rubber must be avoided.

The test facility must have the basic instruments required to monitor the quality (e.g., temperature, pH) of the test soil and associated *test (hydration) water*. Additionally, the laboratory should be equipped to facilitate prompt and accurate analysis of the moisture content of test soils. Equipment requirements include a drying oven that can be set at 105 °C for drying soils, a weighing balance accurate to the nearest 0.1 mg, and a pH meter. Safety apparatus, including a respirator with dust protection, gloves, laboratory clothing, and glasses for eye protection, are required when preparing mixtures and aliquots of test soil.

All test vessels, equipment, and supplies that might contact site soils, test soils, test (hydration) water, *stock solutions*, or test solutions must be clean and rinsed with deionized or distilled water (i.e., test water), before use. All non-disposable materials should be washed after use. The following cleaning procedure is recommended (EC, 2005b, 2013b, 2014a; ECCC, 2020a):[26]

1. soak in tap water (with or without detergent added) for 15 minutes, then scrub with detergent or clean in an automatic dishwater;

2. rinse twice with tap water;

3. rinse carefully with fresh, dilute (10%, v:v[27]) nitric (HNO_3) or hydrochloric acid (HCl) (metal-free grade) to remove scale, metals, and bases;

4. rinse twice with deionized water (or other test water);

5. rinse once with full-strength, pesticide-grade acetone to remove organic compounds and with reagent-grade (e.g., HPLC-grade, ≥ 98.5% purity) hexane for oily residues (use a fume

[26] Steps 1–4 of the cleaning procedure should be used if metal contamination is of concern; steps 1, 2, 5, 6, and 7 should be used if contamination with organics is of concern; and all steps should be followed if both metal and organics contamination is suspected.

[27] To prepare a 10% solution of acid, carefully add 10 mL of concentrated acid to 90 mL of deionized water.

hood);[28]

6. allow organic solvent to volatilize from dishware in fume hood and rewash with detergent (scrub if necessary); and

7. rinse three times with deionized water (or other test water).

Test vessels and apparatus that might contact soil or test (hydration) water should be thoroughly rinsed with test water immediately before being used in the test.

3.2 Initial and Definitive Tests

3.2.1 Initial Tests

Before definitive soil toxicity tests using either of the test methods defined in Sections 4.2 or 4.3 are performed for the first time by a testing laboratory, it is recommended that a minimum of five control performance tests with one or more samples of uncontaminated natural or artificial soil intended (or under consideration) for use in one or more definitive soil toxicity tests as negative control soil (see Section 3.3) be undertaken by laboratory personnel. Additionally, a minimum of five reference toxicity tests should be performed using one or more samples of a candidate artificial or natural negative control soil intended for routine use in conjunction with definitive soil toxicity tests (see Section 4.4). These initial tests are recommended to confirm that acceptable performance of the test species (*E. andrei* or *D. rubidus*) can be achieved in a candidate natural or artificial negative control soil (see Section 3.3) in a specific laboratory and under the culturing or holding/acclimation conditions and procedures specified in this report (see Sections 2.3 and 2.4).

The conditions and procedures used to perform these initial tests with negative control soil should be identical and according to Section 4.2 (if 56-day reproduction tests are intended) or Section 4.3 (if avoidance tests are intended).[29] The conditions and procedures used to perform these initial reference toxicity tests should be identical and according to Section 4.4. Each test with negative control soil or reference toxicant(s) should be performed using a different lot (group) of test organisms of the same species from the same source.

Data from the control performance tests ($n \geq 5$) must show that the criterion or criteria for test validity (see Sections 4.2.3 and 4.3.3) can be met for the intended test species using a natural or artificial soil intended for use as negative control soil in a definitive soil toxicity test. Data from the initial reference toxicity tests ($n \geq 5$) should be compared by calculating and appraising the magnitude of the coefficient of variation (CV) for the respective series of tests and endpoint values (see Section 4.4).

3.2.2 Reproduction Test

Glass jars with a 500-mL capacity for *E. andrei*, and a 250-mL capacity for *D. rubidus*, must be used as test vessels. Wide-mouth glass Mason jars have been successfully used as test vessels for the 56-day reproduction test. Each glass jar (new or used) must be cleaned thoroughly (see Section 3.1) before and after use, and rinsed well with deionized or other test water immediately before use. For *E. andrei* each test vessel should be covered with a lid that is perforated with ≥ 5 small (e.g., ~1–2 mm) holes (to minimize evaporation and allow air exchange) or with a piece of 50 μm Nitex mesh (see guidance in the following paragraph for *D. rubidus*) and secured to the lip of each jar using a rubber band or screw ring tightened onto the test vessel. If the test material is known or thought to contain volatile compounds

[28] Rinsing Plexiglas™ or any plastic equipment or vessels with acetone or hexane is **not** recommended, since plastic can become pitted and etched by these solvents and can turn from transparent to opaque.

[29] Initial tests with negative control soil for the avoidance test may follow the guidance for "dual tests" provided in ISO 17512-1 (ISO, 2008). To assess the performance of a candidate negative control soil (i.e., natural negative control soil) that differs from the laboratory soil that the worms are acclimated to (i.e., artificial soil or culture substrate), three of the avoidance unit compartments are filled with the laboratory soil and the three others with the candidate negative control soil (ECCC, 2020b). At the end of the 48-hour exposure, a relatively homogenous distribution of worms (i.e., 40–60% in each soil type) is indicative of suitable negative control soil. Alternatively, laboratories may use the results of the initial tests for control performance in the 56-day reproduction test as indication that a control soil is suitable for use in a 48-hour avoidance test.

(e.g., PAHs), the use of opaque aluminum foil as covers is recommended, together with side lighting sufficient to achieve the minimal light intensity required at the surface of the soil (see footnote 51 in Section 4.2.1).

For *D. rubidus*, each vessel should be covered with a piece of 50 µm Nitex mesh, or equivalent, and secured to each jar using a screw ring. Each vessel and lid should then be loosely covered to reduce the loss of moisture through evaporation. This can be accomplished by placing the metal cap portion of the Mason canning jar lid on top of the Nitex mesh and screw ring.[30]

3.2.3 Avoidance Test

The recommended test apparatus for performing an avoidance test with earthworms is illustrated in Figure 2.[31] The design of each test unit consists of a circular container with an outer diameter of ~230 mm. Each test unit is partitioned into a central cylinder with an inner diameter of ~54 mm, and six pie-shaped interconnecting compartments each with a capacity for ~350 mL of soil. A series of 1-cm holes drilled in the bottom of the central chamber (two per compartment) and on the sides of each pie-shaped compartment (three per side) enable the free movement of earthworms from the central cylinder (devoid of substrate) to the test compartments containing test soils, and free movement of test organisms between the compartments. A set of six removable side partitions, made of rigid steel sheeting (see Figure 2 for illustration and dimensions) is required for insertion alongside each of the six walls separating compartments at the end of the test (Section 4.3.6).

The apparatus can be constructed of high quality stainless steel sheeting (1–4 mm thick) or Plexiglas™ sheeting (5–6 mm thick), and includes a removable lid (also made of stainless steel or Plexiglas™) that does not seal and enables an exchange of air within the test compartments. Avoidance apparatus constructed of stainless steel is recommended when testing soils contaminated or spiked with organic compounds (particularly petroleum products), since this material sorbs fewer organics than Plexiglas™ and can be rinsed with acetone and/or hexane without damaging it. Apparatus constructed of Plexiglas™ is recommended when testing soil contaminated with heavy metals. The test unit should be used as described above for *E. andrei*.

For *D. rubidus*, however, the test unit must be modified by adding a false back to each segment of the test unit (see Figure 3C) to reduce the volume of soil required for each compartment (Section 4.3.1).[32] A steel U-shaped bracket (85 mm long × 85 mm high × 25 mm long sides bent at 120° angles) can be used for this purpose (see Figures 3A and 3B). Alternatively, the removable side partitions, used for separating compartments at the end of the test, can be used to create a false back for each compartment (see Figure 3C).

---

[30] Test vessel lids were modified for *D. rubidus* to prevent the organisms' tendency to escape from the test vessels when exposed to inhospitable soils. The use of Nitex screening was successful in preventing organism escapes, however allowed significant moisture loss through evaporation. Following a laboratory investigation, it was determined that loss of soil moisture through the mesh top could be mitigated by placing a metal lid on top of the test vessel (i.e., placed loosely on top of the mesh and screw ring) (ECCC, 2020b).

[31] The experimental apparatus used in this test is called the "Kaushik chamber," after the professor responsible for its design (N. Kaushik, University of Guelph, personal communication, 1995). The prototype was used in the early 1960s to investigate the preference of aquatic oligochaetes to sediment with different grain size characteristics. The design was modified to accommodate the larger terrestrial earthworm species.

The dimensions for the apparatus depicted in Figure 2 represent a test unit constructed of stainless steel; similar dimensions apply when constructing the avoidance apparatus using Plexiglas™.

[32] The avoidance test unit described herein was designed for use with worms that are much larger than *D. rubidus* (e.g., *E. andrei* and *L. terrestris).* During the development of the avoidance test for *D. rubidus*, it was determined that smaller volumes of soil used in each compartment facilitated the movement of these smaller worms throughout the test unit and increased the recovery of adult worms at the end of the test (ECCC, 2020b).

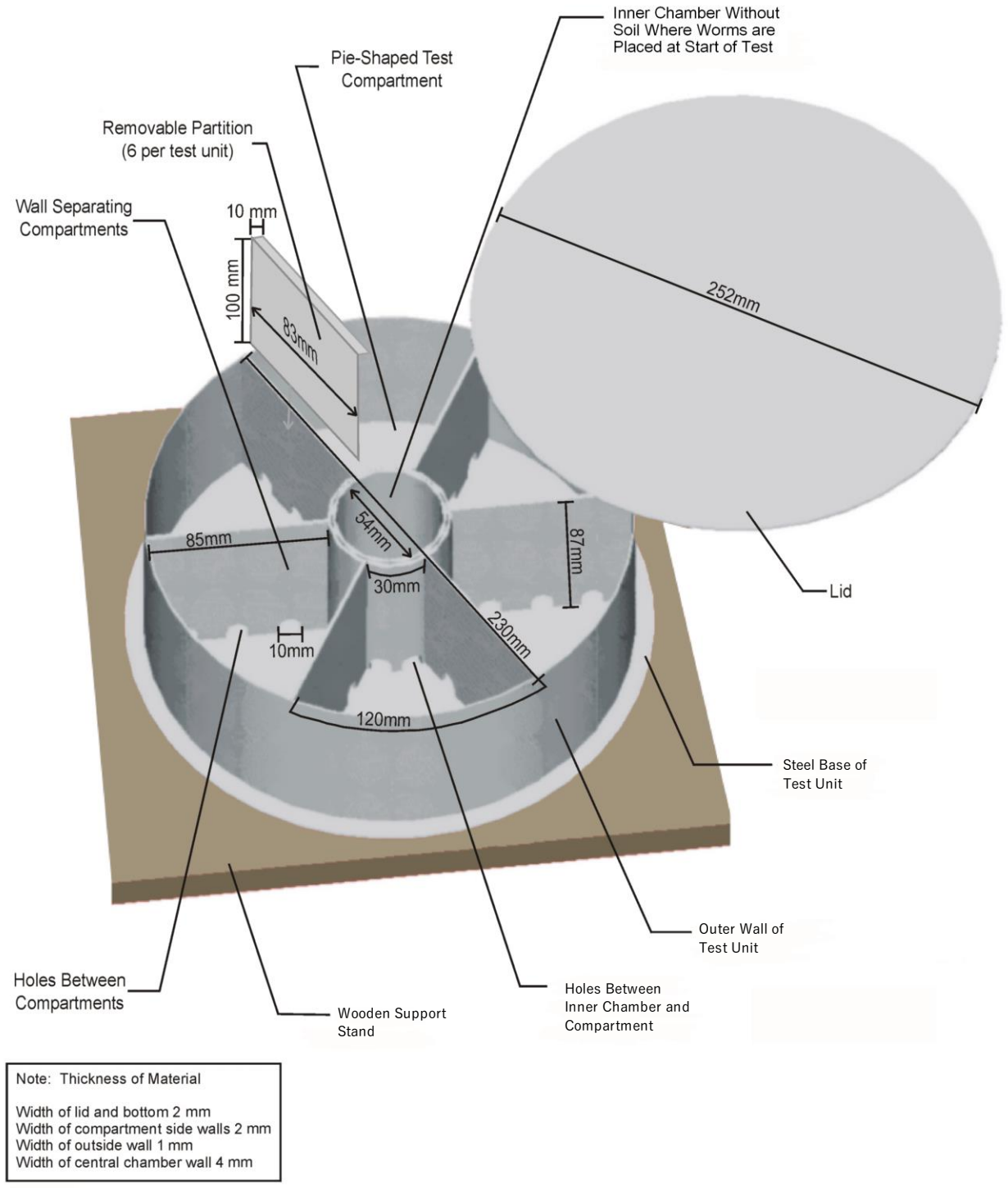

Figure 2 Recommended design of test unit for performing an avoidance test using earthworms (*E. andrei* or *D. rubidus*) and clean or contaminated soil

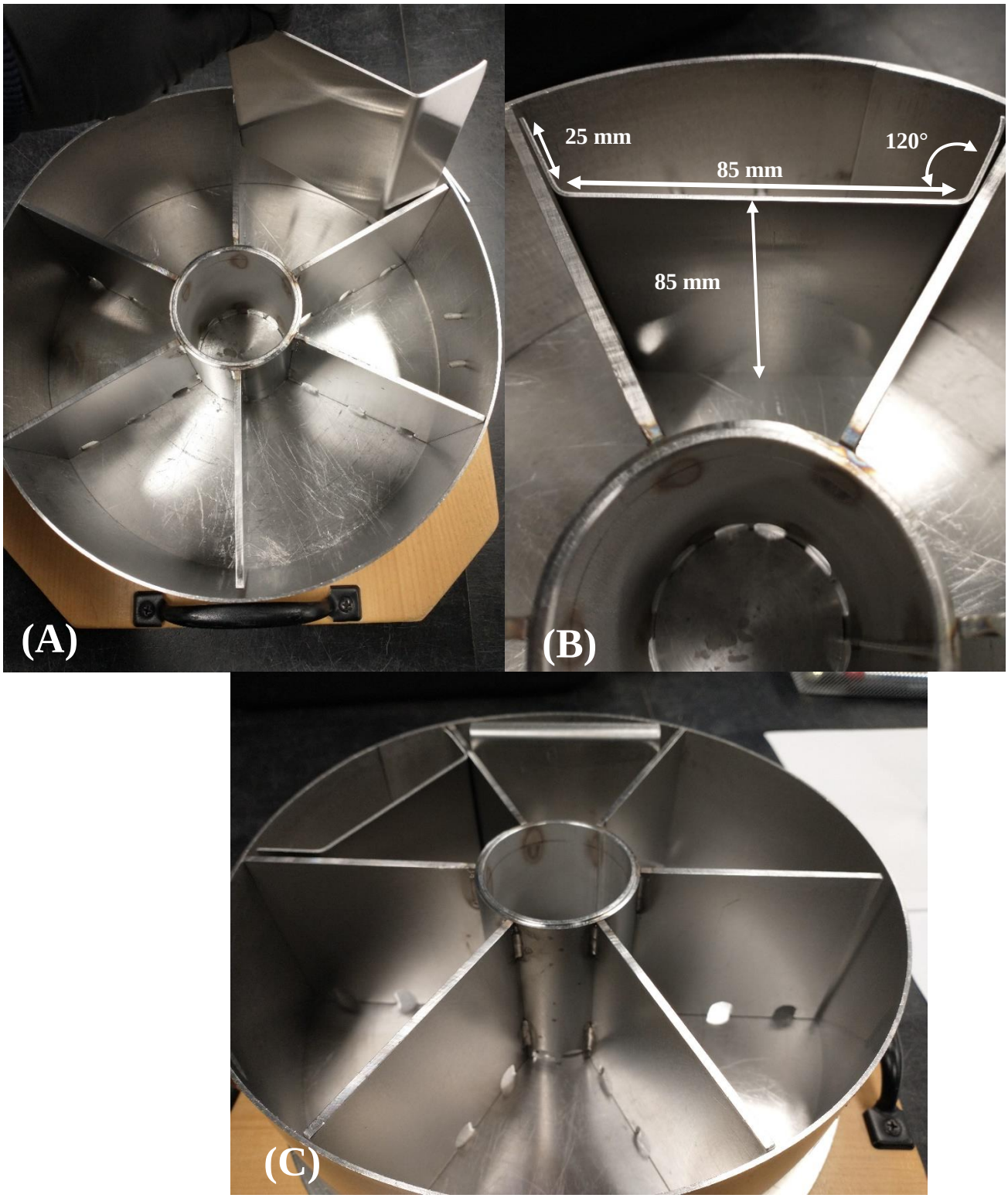

Figure 3 **Modification of the test unit required for performing an avoidance test using *D. rubidus*. (A) False back plate being placed inside a compartment of the test unit. (B) False back plate in position within a compartment of the test unit. (C) Two options for creating a false back in each compartment of the test unit.**

Since these partitions are slightly taller than the top of the test unit, the lid is not flush when placed on top of the unit, thereby creating a small gap between the lid and the top of the test unit. If any gap occurs between the top of the test unit and the lid or the bottom of the test unit and the base (i.e., due to the height and/or warping of the false back inserts), one strip of Parafilm should be placed around the lid and side of the unit and a second around the bottom of the unit and the steel base to seal the test unit. This is important to prevent *D. rubidus* from escaping from the test units.

A minimum of five test units are required for a single-concentration toxicity test (or fewer if the single-concentration test is for screening or range-finding purposes only), and a minimum of 2 test units per test concentration (i.e., ≥ 10 units per test) are required for each multi-concentration test (see Section 4.3.1). Each test unit must be cleaned thoroughly before and after use, and rinsed well with deionized or other test water before use.

3.3 Negative Control Soil

Each soil toxicity test must include negative control soil as one of the experimental treatments. Negative control soil is essentially free of any contaminants that could adversely affect the performance of earthworms during the test. The use of negative control soil provides a measure of test acceptability, evidence of the health and performance of the test organisms, assurance as to the suitability of the test conditions and procedures, and a basis for interpreting data derived from the test soils.

A soil toxicity test may use clean (uncontaminated) natural soil and/or artificial soil as the negative control soil. The selection of an appropriate negative control soil depends on considerations such as the study design, physicochemical characteristics of the test soil(s), and the availability of suitable clean natural soil with acceptable properties.[33] For definitive tests with field-collected boreal forest and taiga soils, it is recommended that uncontaminated natural soil be used as the negative control soil. Regardless of soil type, there must be prior experimental evidence (see Section 3.2.1) that the soil chosen for use as negative control soil with the chosen test species will consistently and reliably meet the criteria for test validity defined herein for each test method (Sections 4.2.3 and 4.3.3).

The first edition of the biological test methods described herein were developed and tested using five negative control soils with diverse physicochemical characteristics (Aquaterra Environmental, 1998; Stephenson *et al.*, 1999a, 1999b, 2000a; Aquaterra Environmental and ESG, 2000; ESG, 2000, 2001, 2002; ESG and Aquaterra Environmental, 2002). These clean soils included one artificial soil and four natural soils (i.e., samples of sandy loam and silt loam agricultural soils from southern Ontario, a clay loam prairie soil from Alberta, and a forest loam soil from the Canadian Shield in northern Ontario). The test methodologies described in this second edition test method document were further developed for new test designs (56-d reproduction and 48-h avoidance) and for testing boreal soils with *D. rubidus*, using a variety of clean soils including: an artificial soil, a standard agricultural soil from Europe (LUFA 2.2), 2 agronomic soils (for *E. andrei* only), and 7 natural soils collected from Canada's Boreal Region (for *D. rubidus* only). These boreal soils include: Gleyed Humo-ferric Podzols from Newfoundland, New Brunswick, and Ontario; a Dark Grey Luvisol, an Ortho Eutric Brunisol, and an Eluviated Dystric Brunisol from Saskatchewan; and a Rego Dark Grey Chernozem from Alberta. These soils differed in composition with respect to the physicochemical characteristics that could potentially influence the fate and effects of contaminants. All of the field-collected soils originated from uncontaminated areas that had not been subjected to any direct application of pesticides in recent previous years and were therefore considered to be "clean." The origin and physicochemical characteristics of these natural soils

[33] The Canadian Council of Ministers of the Environment (CCME) provides a comprehensive website on Canadian Environmental Quality Guidelines including those for soil (www.ccme.ca). This information is useful when reviewing analytical data (e.g., values for metals or PAHs) for samples of field-collected soil from a location under consideration as a source of natural soil suitable for use as negative control soil in toxicity tests. The website and associated links will assist the investigator(s) reviewing the physicochemical characteristics of presumably clean natural soils under consideration for use as negative control soil in soil toxicity tests. The CCME can also be contacted by phone (1-204-948-2090) or email (info@ccme.ca).

are further described in Appendix F. The test validity criteria for *E. andrei* or *D. rubidus* described in Sections 4.2.3 and 4.3.3 are based on the performance data for these earthworms in negative control soil that were generated for each of these diverse soils (EC, 2010; ECCC, 2020b). During the development of the 56-day reproduction test method for *D. rubidus*, there was an observed trend of lower numbers of juveniles produced in some of the *soil horizons*, often, but not limited, to those soils with pH < 4.

3.3.1 Natural Soil

Negative control soil may be natural soil collected from a clean (uncontaminated) site known to have been free of pesticide or fertilizer applications for at least five years. The source of this negative control soil might be the same as that where earthworms were collected to establish a culture or to obtain test organisms (Section 2.2).

It is recommended that all samples of natural soil selected for possible use as negative control soil in soil toxicity tests be checked on an ongoing basis to ensure test organisms continue to meet test validity criteria, as there is potential for some natural control soils to degrade when stored for prolonged periods of time (ECCC, 2020b).[34] In addition all samples selected for used as negative control soil (as well as samples of candidate *reference soil*) must be analyzed for the following physicochemical characteristics:

- particle size distribution (% sand, % silt, and % clay)
- *total organic carbon* content (%)[35]
- organic matter content (%)[35]
- pH
- electrical conductivity
- moisture content (%)
- water-holding capacity (WHC)
- *cation exchange capacity* (CEC)

Additionally, the following analyses should be performed:

- major cations and anions (Na^+, K^+, Mg^{2+}, Ca^{2+}, Al^{3+}, S^{2-}, Cl^-)
- nitrogen as total N, nitrate (NO_3^-), nitrite (NO_2^-), and ammonium (NH_4^+)
- phosphorus as total and/or bioavailable
- potassium as total and/or bioavailable
- C:N ratio

To confirm that the negative control and/or reference soils are not contaminated, the following screening analyses are recommended:

- organophosphorus insecticide suite
- organochlorine insecticide suite
- herbicide suite
- metal suite
- petroleum hydrocarbons (including PAHs)
- other site- or area-specific contaminants of concern

Pesticide and metal concentrations should not exceed the CCME soil quality criteria, if available (see footnote 33). If indigenous organisms are present and/or problematic in the sample(s) of natural soil at any time (i.e., during storage or testing), their presence (e.g., physical description and estimated numbers) should be recorded, and they should be removed manually (e.g., by sieving), if possible. Alternatively, most indigenous organisms can be killed by at least one or more freeze/thaw cycles if it is suspected that they are too small to remove manually (see footnote 110 in Section 5.3 herein, and Section 5.6.6 of EC, 2012). If the results of both the initial biological tests and the physicochemical analyses are satisfactory, a larger sample of this natural soil can be collected, air dried to a moisture content of between 10 and 20%, coarse-screened (e.g., 4–10 mm),[36] transferred to clean, thoroughly rinsed plastic pails, and stored in

---

[34] After long periods of storage, soil can be refreshed by lightly saturating and gently mixing the soil to aerate, before being re-covered and stored. This process should be repeated weekly for a minimum of 6 weeks before use in a test.

[35] Organic matter content can be used to calculate total organic carbon (TOC) by multiplying the organic matter content (OM) of a soil by a soil constant (AESA, 2001). However, the relationship between TOC and OM is slightly different among soils, and the total organic carbon content should also be determined by laboratory analysis.

[36] The more porous sieve sizes (e.g., 6–10 mm) might be needed for soils with a higher organic content. Further guidance on the requirement for sieving, including

darkness at 4 ± 2 °C until required. Plastic pails should not be used for collection and storage of soils if there are concerns about chemical constituents of the plastic leaching into the soil.

3.3.2 Artificial Soil

Negative control soil may be artificial soil formulated in the laboratory. The use of artificial soil offers a consistent, standardized approach and is advantageous when testing the toxicity of chemicals or chemical products spiked in negative control soil (Section 6).

In keeping with the formulation of artificial soil used in four other Environment Canada soil toxicity test methods (EC, 2005b, 2013b, 2014a; ECCC, 2020a), the following ingredients should be used to prepare artificial soil to be used in the biological test methods described herein (based on dry mass):

i) 10% *Sphagnum* sp. peat, air dried and sieved (e.g., through a 2-mm mesh screen);
ii) 20% kaolin clay with particles < 40 μm; and
iii) 70% "grade 70" silica sand

The ingredients (above percentages expressed as dry mass fraction) should be mixed thoroughly in their dry form using a mechanical stirrer and/or gloved hands.[37] Reagent-grade calcium carbonate should be added to the dry mixture in a quantity sufficient to attain a pH for the artificial soil ranging within 6.0–7.5 once it is hydrated.[38] Thereafter, the mixture should be hydrated gradually using test water (i.e., deionized or distilled water) until its moisture content is ~20% (which is ~28% of the soil's water-holding capacity),[39] while mixing further until the soil is visibly uniform in colour and texture. As necessary, reagent-grade calcium carbonate should be added to the hydrated mixture in a quantity sufficient to maintain a pH ranging within 6.0–7.5. Samples of pH-adjusted artificial soil should be stored in darkness at 20 ± 2 °C for a minimum of three days before being used in a toxicity test, to enable adequate time for pH equilibration (see footnote 38). Thereafter, artificial soil can be stored at 4 ± 2 °C. As and when required for a soil toxicity test, a suitable quantity of stored artificial soil should be hydrated further using test water until its moisture content is ~70% of the water-holding capacity or until it has the optimal texture for testing (i.e., a homogeneous crumbly consistency with clumps ~3–5 mm in diameter; see Section 5.3).

Samples of artificial soil selected for possible use as

---

appropriate sieve size selection, is provided in EC (2012).

[37] It is recommended that the dry ingredients initially be mixed (to incorporate the calcium carbonate) using a mechanical stirrer. Mixing should be completed using a gloved hand, to ensure that all of the soil from the corners of the container have been well mixed. Personnel must take the appropriate precautions for protection to prevent the inhalation of and contact with these ingredients.

[38] The amount of calcium carbonate ($CaCO_3$) required to adjust the pH of artificial soil to within this range depends on the nature (i.e., acidity) of the ingredients (and, in particular, that of the *Sphagnum* sp. peat). A quantity of 10–30 g of $CaCO_3$ for each kg of peat might prove adequate. A pH as low as 4.5 might occur when the soil is first formulated without the addition of $CaCO_3$. The initial pH adjustment should attempt to raise pH to range within 7.0–7.5, since the pH of artificial soil typically drops slightly (to 6.5–7.0) during the three-day equilibration period, before it stabilizes. The pH of stored samples of artificial soil should be checked regularly (e.g., once every two weeks) to ensure that it has not changed dramatically; adjustments should be made as necessary by adding additional quantities of $CaCO_3$ (Aquaterra Environmental, 1998; G.L. Stephenson, Aquaterra Environmental, personal communication, 2001). A mixture of formulated artificial soil can also be stored dry, followed by partial hydration to ~20% moisture content, storage at 20 ± 2 °C for a minimum 3-day period, and subsequent hydration to ~70% WHC (or until it has the optimal texture for testing) when required for use in a toxicity test. If storing formulated artificial soil dry, it is necessary to partially hydrate (to ~20% moisture) and equilibrate thereafter (for ≥ 3 days) to provide conditions for pH equilibrium similar to those recommended herein using artificial soil stored partially hydrated. Using this optional approach, the interim storage as partially hydrated artificial soil is necessary to enable the addition of more water (and, in certain instances, the addition of a chemical solution) as required when finalizing the pH and moisture content (i.e., adjusted to ~70% WHC) of artificial test soil. Storage of artificial soil that is partially hydrated, rather than dry, is considered a preferred approach since it enables laboratory personnel to more quickly hydrate to the desired moisture content (i.e., ~70% WHC) while ensuring pH equilibrium, and reduces any further delay in time associated with the dry storage of artificial soil.

[39] The % hydration might need to be adjusted higher or lower depending on the type of peat used in preparing artificial soil.

negative control soil in soil toxicity tests must be analyzed for the following physicochemical characteristics:

- particle size distribution (% sand, % silt, and % clay)
- total organic carbon content (%)[35]
- organic matter content (%)[35]
- pH
- electrical conductivity
- moisture content (%)
- water-holding capacity (WHC)
- cation exchange capacity (CEC)

Additional analyses, such as those described for natural soils (Section 3.3.1) may also be carried out, as necessary.

3.4 Positive Control Soil

The use of one or more samples of *positive control soil* is recommended for inclusion in each series of soil toxicity tests with earthworms, to assist in interpreting the test results. In choosing a positive control soil, the intent is to select a toxic soil that will elicit a response in the test organisms that is predictable based on earlier toxicity tests with this material. The positive control soil may be a sample of negative control soil that is spiked with a reference toxicant for which historic data are available on its toxicity to earthworms using specified test conditions and procedures. For the two biological test methods described herein, one or more reference toxicants must be used in a multi-concentration test or as *replicates* of a positive control soil (i.e., at a specified concentration) when appraising the sensitivity of the test organisms and the precision and reliability of results obtained by the laboratory for that material (see Section 4.4). A test might also include a sample of negative control soil (natural or artificial; see Section 3.3) that has been spiked experimentally (Section 6) with one or more toxic chemicals or chemical products of particular concern when evaluating the sample(s) of test soil, at a concentration toxic to earthworms according to the biological test method to be used. In some instances, a test might include a positive control soil that consists of a highly contaminated sample of field-collected soil or sludge shown previously to be consistently toxic to earthworms according to the biological test method to be used.[40]

3.5 Reference Soil

One or more samples of reference soil might be included in a soil toxicity test using earthworms. The type and nature of the sample(s) of soil used as reference soil in a particular study depend on the experimental design and the study's objectives. If the toxicity of samples of field-collected soil from a contaminated or potentially contaminated site is under investigation, the reference soil included in the study might be one or more samples of field-collected soil taken from a clean (uncontaminated) site where the physicochemical properties (e.g., total organic carbon content, organic matter content, particle size distribution, texture, pH, electrical conductivity) represent the sample(s) of test (contaminated) soil as much as possible. Ideally, the reference soil is collected from the general vicinity of the site(s) where samples of test soil are collected, but is removed from the source(s) of contamination. One or more samples of field-collected clean reference soil from near the test site(s) might also be chosen due to their known lack of toxicity in previous tests with earthworms, and their possession of physicochemical characteristics similar to the test soil samples. Boreal forest and taiga reference soils must be collected as separate soil horizons, where possible. Each soil horizon must then be stored and tested individually (i.e., each horizon is treated as a separate soil sample) (see Section 5.1 and EC, 2012). The sample(s) of field-collected reference soil used in a study could be tested for toxic effects as undiluted soil only, or this soil could be mixed with the sample(s) of test soil to prepare a range of concentrations to be included in a multi-concentration test[41] (see Sections 3.6, 4.1, 5.3, and

[40] If the positive control soil consists of a highly contaminated sample of field-collected soil, it is important that its toxic potential be stable over time (i.e., the sample is old enough that the bioavailability has been stabilized).

[41] Alternatively, the series of test concentrations used in a multi-concentration test could be prepared using negative control soil. The choice might be influenced by whether or not the candidate reference soils are likely known to be nontoxic in the test to which they are to be applied, or a desire to prepare a range of concentrations of test soil using a clean soil with characteristics (e.g., texture, organic matter content) that closely match that of the test soil.

5.6.1). Samples of reference soil should not be collected from sites known to have received applications of pesticides or fertilizers within the past five years or more.

An investigator might choose to include one or more samples of artificial soil as reference soil in a particular test. For instance, these could be used in multi-concentration tests with site soils or chemical-spiked soils to investigate the influence of certain physicochemical characteristics (e.g., a number of artificial reference soils prepared to provide a range of differing values for texture and/or organic matter content (%); Sheppard and Evenden, 1998; Stephenson *et al.*, 2002) on the toxicity of a contaminated site soil or a chemical-spiked soil. Multiple samples of clean field-collected soil collected from various sites, which differ markedly with respect to one or more physicochemical characteristics, might also be used for this purpose. For such a study, a portion of each reference soil used to prepare a series of concentrations of the test soil should be included in the test without dilution (i.e., 100% reference soil).

Each test involving one or more samples of reference soil must include a sample of negative control soil (see Section 3.3). Conversely, certain tests (e.g., one involving a series of concentrations of chemical-spiked soil prepared using artificial or natural negative control soil) need not involve a sample of reference soil. For tests with field-collected site soil, the inclusion of one or more samples of reference soil from a neighbouring site is a preferred approach for comparative purposes (see Section 5.6); the decision to dilute site soil with reference soil (rather than negative control soil) when preparing multiple concentrations for testing depends on the study objectives.

3.6 Test Soil

These biological test methods are intended to measure the toxicity of one or more samples or mixtures of contaminated or potentially contaminated soil (test soil), using earthworms as test organisms. The sample(s) of test soil might be either field-collected soil from an industrial or other site of concern, or industrial or municipal biosolids (e.g., dredged material, municipal sludge from a sewage treatment plant, composted material, or manure) under consideration for possible land disposal. A sample of field-collected test soil might be tested at a single concentration (typically 100%), or evaluated for toxicity in a multi-concentration test whereby a series of concentrations are prepared by mixing measured quantities with either negative control soil or reference soil (see Section 5).

Field-collected soils collected by horizon take into account contamination stratified due, in part, to the different speciation and resultant mobility of contaminants (EC, 2012). Therefore, for soils collected from the boreal or taiga ecozones, both reference and contaminated soils must be collected in separate horizons. Soils collected in horizons must be treated as individual soil samples and tested separately (see Section 4.1). Soils without distinct soil horizons (e.g., where the surface soil horizons have been mixed or disturbed due to human activity) should be collected according to depth (see Section 5.1). The test soil might also be one or more concentrations of a chemical-spiked soil, prepared in the laboratory by mixing one or more chemicals or chemical products with negative control soil, reference soil, or site soil (see Section 6). Guidance on the collection, handling, analyses, and testing of field-collected soils is provided in Section 5.

Section 4

Universal Test Procedures

General procedures and conditions described in this section for each of two biological test methods with earthworms apply when testing the toxicity of samples of soil, particulate waste, or chemicals, and also apply to their associated reference toxicity tests. More specific procedures for conducting tests with field-collected samples of soil or other similar particulate material (e.g., sludge, dewatered mine tailings, drilling mud residue, compost, biosolids) are provided in Section 5. Guidance and specific procedures for conducting tests with negative control soil or other soils spiked (amended) experimentally with chemical(s) or chemical product(s) are given in Section 6. Specific guidance on conducting tests with boreal and taiga soils has been incorporated throughout this test method document.

All aspects of the test system described in Section 3 must be incorporated into these universal test procedures. Those conditions and procedures described in Section 2 for culturing and/or acclimating *E. andrei* and *D. rubidus,* in preparation for soil toxicity tests, also apply.

4.1 Preparing Test Soils

Each test vessel (see Section 3.2.2) or avoidance unit (see Section 3.2.3) placed within the test facility must be clearly coded or labelled to enable identification of the sample and (if diluted) its concentration. For the avoidance test (see Section 4.3), each of the six compartments within each test unit (see Section 3.2.3) must also be coded (e.g., identified by numbers or letters) or otherwise marked to distinguish the test soil therein. The date and time when the test is started must be recorded, either directly on the labels or on separate data sheets dedicated to the test.

The day that earthworms are initially exposed to samples of test materials or substances is designated Day 0. On the day preceding the start of the test (i.e., Day -1), each sample or subsample of test soil or similar particulate material, including negative control soil and, if used, reference soil, should be mixed thoroughly[42] (see Sections 5.3 and 6.2) to provide a homogeneous mixture consistent in colour, texture, and moisture. If field-collected samples of site soil are being prepared for testing, large particles (e.g., stones, thatch, sticks, debris) should be removed before mixing, along with any vegetation or macroinvertebrates observed (see Section 5.3). If there is concern over the volatilization, degradation, or metabolism of contaminants or chemicals in test soils, the test can be initiated immediately after the preparation of the test soil (see Section 6.2).

The quantity of each test soil or soil horizon mixed as a batch should be enough to establish the replicates of a given treatment (see Tables 3 and 4), plus an additional amount for the physicochemical analyses to be performed (Sections 4.2.5 and 4.3.5) and a surplus to account for the unused portion of soil that adheres to the sides of the mixing vessel. The moisture content (%) of each test soil should be known or determined, and adjustments made as necessary by mixing in test water (or, if and as necessary, by dehydrating the sample) until the desired moisture level is achieved (see Sections 5.3 and 6.2). Quantitative measures of the homogeneity of a batch can be made by taking aliquots of the mixture for measurements such as particle size analysis, total organic carbon content (%), organic matter content (%), moisture content (%), and concentration of one or more specific chemicals.

For soils collected as distinct horizons (e.g., boreal or taiga soils), each horizon must be prepared and tested separately in independent definitive tests.[43]

[42] Any liquid that has separated from a sample or subsample of test soil during transport and/or storage must be remixed into the sample.

[43] Initial tests with *D. rubidus* were conducted using various horizons of boreal forest soils layered in test vessels. Results of these tests showed that a majority of the worms (i.e., 94%) were found in the bottom layer of the test vessel, regardless of soil type or how the horizons were layered (i.e., layered as per the profile when

For soils to be assessed in multi-concentration tests, each horizon of the test soil should be mixed with the same horizon of negative control or reference soil (see Section 5.3) at the various test concentrations (e.g., 0%, 6.25%, 12.5%, 25%, etc.). In some cases, it might not be possible to collect the same horizons of negative control soil and test soil. For example, negative control soils might be collected in horizons, but this might not be possible at the site of contamination, i.e.., more than one horizon of test soil might not be present or horizons might be mixed. In this case, test concentrations should be prepared by mixing suitable weights of test soil into the available horizon(s) of negative control soils at the appropriate test concentrations.

For any multi-concentration test to be performed according to the test procedures described herein (see Sections 4.2 and 4.3), concentrations should be chosen to span a wide range, including a low concentration that evokes no adverse effects (similar to that for the *negative control* treatment), and a high concentration that results in "complete" or severe effects. If the anticipated endpoint is bracketed with a closely spaced series of concentrations, all might turn out to be either too low or too high. To keep the wide range of concentrations, and also obtain the important mid-range effects, it might be necessary to use additional treatments to split the selected range more finely. In any case, a consistent geometric series should be used (see Appendix G). See EC (2005a) for additional guidance on selecting test concentrations, which applies here.

In the case of appreciable uncertainty about sample toxicity, a range-finding test might prove worthwhile for selecting, more closely, the concentrations to be used for the definitive test. For a range-finding test, a wide range of concentrations may be tested using fewer replicates (e.g., a single test vessel or avoidance unit) per treatment (see Sections 4.2.1 and 4.3.1).

4.2 Reproduction Test

This biological test method measures the effects of exposure to contaminated soil on the reproductive success of earthworms (laboratory-cultured *E. andrei* or *D. rubidus*).[44]

Table 3 provides a summary checklist of required and recommended conditions and procedures to be universally applied to each test with samples of contaminated or potentially contaminated soil, as well as those for testing specific types of test materials or substances. These could include samples of site soil (including boreal and taiga soils), biosolids mixed into soil (e.g., dredged material, sludge from a sewage treatment plant, composted material, or manure), or negative control soil (or other soil, contaminated or clean) spiked in the laboratory with one or more test chemicals or chemical products. This test method was originally developed using guidance provided by ISO (1991, 1998) and OECD (2000) for the performance of tests for the effects of chemical-spiked soil on the reproduction of *E. andrei*. It has been updated based on feedback received after 16 years of use by Canadian laboratories and research carried out at ECCC for the improvement of variability and efficiency of the method, and for the inclusion of a boreal species (*D. rubidus*) (see Appendix E).

Universal procedures for performing a test for effects on the reproduction of earthworms are described in this section. This is a whole soil toxicity test, with no renewal of test soils during the 56-day test duration (i.e., *static* test). The test begins with

---

collected, or inversed, with upper horizons placed on the bottom of the vessel). It was unclear as to whether or not this effect was due to soil characteristics or an artifact of processing the test soil (i.e., removing the soil in order from top to bottom). These initial results led to the conclusion that for tests involving invertebrates, each soil horizon should be tested separately in independent definitive tests (EC, 2010).

[44] The measurement of juvenile growth is no longer a required endpoint for this test (see footnote 2 in Section 1.3.1). Although reproduction is the preferred endpoint herein, the test does not preclude the measurement of juvenile growth if warranted (see footnote 68 in Section 4.2.6). Care should be taken to consider the possibility of a density-dependent effect or dual effect if growth is to be measured. This is when replicates where large numbers of juveniles are produced show a corresponding decrease in the individual dry weight of the juveniles. This effect can confound the interpretation of the data, and care must be taken to ensure that any lowered individual dry weight observed is not interpreted as a contaminant effect, where the confounding effect of organism density on juvenile growth occurs (MESI, 2014).

Table 3 Checklist of required and recommended conditions and procedures for conducting tests for effects of exposure to contaminated soil on the reproduction of earthworms (*E. andrei* or *D. rubidus*)

Universal		
Test type	−	whole soil toxicity test; no renewal (static test)
Test duration	−	≥ 56 days
Test organisms	−	laboratory-cultured *E. andrei* or *D. rubidus*; sexually mature adults with clitellum; individual wet wt of 250–600 mg for *E. andrei* or 50–200 mg for *D. rubidus*; choose worms as similar in wet wt as possible; acclimate for ≥ 7 days in negative control soil containing the same food as that to be used in the test; 4 worms per replicate (test vessel)
Number of replicates	−	for multi-concentration tests: ≥ 5 replicates/treatment; each replicate consisting of four worms in a test vessel
	−	for single-concentration tests, minimum number of replicates depends on test species and target effect size (see Sections 4.2.1 and 5.6.2); each replicate consisting of four worms in a test vessel
Number of concentrations	−	for multi-concentration tests: ≥7 concentrations, plus control(s); more recommended (≥10)
Negative control soil	−	depends on study design and objectives; clean field-collected soil or artificial soil if testing site soils; recommend artificial soil for tests with chemicals or chemical products spiked in soil
Test vessel	−	clean glass jars; 500 mL for *E. andrei* or 250 mL for *D. rubidus*; perforated cover (e.g., perforated lid or Nitex mesh), secured with a rubber band or screw ring
Amount of soil/test vessel	−	identical wet wt, equivalent to a volume of ~350 mL for *E. andrei* or ~200 mL for *D. rubidus*
Moisture content, test soils	−	for soil preparation, hydrate to the optimal percentage of its WHC if field-collected soil (see Section 5.3), or to ~70% of WHC if artificial soil (see Sections 3.3.2 and 6.2); during test, hydrate as necessary
Air temperature	−	daily average, 20 ± 2 °C; instantaneous, 20 ± 3 °C
Lighting	−	incandescent, fluorescent, or LED; intensity, 400–800 lux adjacent to the surface of the soil in the test vessels; fixed photoperiod (e.g., 16h L:8h D or 12h L:12h D)
Feeding	−	Magic® Worm Food or organic mixed grains; 2 g for *E. andrei* or 1 g for *D. rubidus* per test vessel each feeding; on Days 0, 14, 28, and 42 only; place in a shallow depression in the centre of the soil surface in each test vessel, moisten with deionized water, and cover with a thin layer of soil

Measurements during test	−	air temperature in test facility, daily or continuously; moisture content, pH, and electrical conductivity (if necessary) of soil in each treatment/concentration, at start and end
Observations during test	−	total number of live adult worms in each test vessel on Days 0 and 28; optionally, number of live and dead adult worms on surface of the soil in each test vessel at the start of the test (t = 1 h) and on Day 28; number of live juvenile worms in each test vessel on Day 56; obvious pathological symptoms (e.g., open wounds) or distinct behavioural abnormalities (e.g., lethargy) for worms in each test vessel; any excessive growth of mould or fungi, presence and quantity of any uneaten food, and apparent "wetness" of soil every two weeks, on each feeding occasion
Biological endpoints	−	number of live adult worms in each replicate (i.e., in each test vessel) on Day 28; number of live juvenile worms in each replicate on Day 56; optional biomass measurements of live juvenile worms in each replicate on Day 56
Statistical endpoints	−	mean (± SD) percent survival of adults in each treatment, on Day 28; mean (± SD) number of live juveniles in each treatment, on Day 56; mean (± SD) number of live juveniles per adult in control(s), on Day 56; if multi-concentration test: 28-day LC50 for adult worms (data permitting), 56-day ICp for number of live juveniles produced in each concentration during the test
Test validity	−	invalid if mean 28-day survival of adults in negative control soil <90%; invalid if mean reproduction for adults in negative control soil <3 live juveniles/adult
Test with reference toxicant	−	choose between a positive control concentration or a multi-concentration reference toxicity test: ○ if the positive control option is chosen, it must be performed with every definitive test; use boric acid (H_3BO_3) or similar; prepare and test ≥ 5 replicates of a predetermined concentration, using artificial soil as a substrate; 4 worms per replicate; follow procedures and conditions described in Section 4.4 and Appendix H; determine % reduction in juvenile production (as a percent of the control response) at test end (i.e., Day 56) ○ if the multi-concentration reference toxicity test option is chosen, it must be performed twice per year; use boric acid (H_3BO_3) or similar; prepare and test ≥ 5 concentrations plus a negative control, using artificial soil as substrate; ≥ 5 replicates/concentration and 4 worms/replicate; follow procedures and conditions described in Section 4.4; determine 56-day IC50 for inhibition of number of juveniles (including 95% confidence limits); express as mg boric acid/kg dry wt; validity criteria are the same as those for definitive test

Field-Collected Soil

Transport and storage	−	seal in plastic or other appropriate material, and minimize air space; labelled or coded; transport in darkness (e.g., using an opaque cooler, plastic pail, or other light-tight container); do not freeze or overheat during transport; store in dark at 4 ± 2 °C; test should start within two weeks, and must start within six weeks unless soil contaminants are known to be stable

Negative control soil	−	either natural, uncontaminated field-collected soil or artificial soil, for which previous 56-day reproduction tests with the test species showed that all criteria for test validity could be regularly met; analyzed for at least the following: particle sizes (% sand, % silt, % clay), TOC (%), OM (%), pH, electrical conductivity, moisture content (%), WHC, and CEC
Reference soil	−	one or more samples for tests with field-collected soil; taken from site(s) presumed to be clean but near sites of test soil collection; characteristics (TOC [%], OM [%], particle size distribution, texture, pH, and electrical conductivity) similar to test soil(s); analyzed as described for natural negative control soil
Characterization of test soils	−	must include at least moisture content (%), WHC, pH, electrical conductivity, TOC (%), OM (%), particle sizes (% sand, % silt, % clay), and CEC; should include at least nitrogen, phosphorus, potassium, C:N ratio, major cations and anions; and, optionally, bulk density, total inorganic carbon, total volatile solids, biochemical oxygen demand, chemical oxygen demand, redox potential, soluble salts, metal oxides, sodium adsorption ratio, contaminants of concern (e.g., metals, polycyclic aromatic hydrocarbons, pesticides), and characteristics of the contamination (e.g., odour, staining, debris, presence of fuel or solvent)
Preparation of test soils	−	if necessary, remove debris and indigenous macro-organisms using forceps; if necessary, gently pass through a sieve of suitable mesh size (e.g., 4–10 mm); homogenize; determine percent moisture content and WHC; hydrate with test water (or, if and as necessary, dehydrate) to the optimal percentage of its WHC (see Section 5.3); mix; dilute with control or reference soil if multi-concentration test; ensure homogeneity

Soil Spiked with Chemical(s) or Chemical Product(s)

Negative control soil	−	artificial soil or a clean field-collected soil for which previous 56-day reproduction tests with the test species showed that all criteria for test validity could be regularly met; analyzed for at least the following: particle sizes (% sand, % silt, % clay), TOC (%), OM (%), pH, electrical conductivity, moisture content (%), WHC, and CEC
Characterization of chemical(s) or chemical product(s)	−	information on concentration of active ingredients and impurities, water solubility, vapour pressure, stability, dissociation constants, adsorption coefficients, toxicity to humans and terrestrial organisms, and biodegradability of chemical(s) or chemical product(s) spiked into negative control soil should be known beforehand
Solvent	−	deionized water is the preferred solvent; if an organic solvent is used, the test must include a solvent control soil in addition to a negative control soil
Preparation of mixtures	−	procedure dependent on the nature of the test substance(s) and the test design and objectives; chemical/soil mixtures may be prepared manually or by mechanical agitation; test substance(s) may be added as measured quantities in solution (i.e., in water or an organic solvent) or as a solid material comprised partly or completely of the test substance(s); ensure homogeneity

Concentration within soil mixture of chemical(s) or chemical product(s) added	–	normally measure at beginning and end of test, in high, medium, and low strengths as a minimum

[*] The information in this table is for summary purposes only. Definitive requirements and recommendations of this test method are contained in the main body of this document.

adult earthworms taken from laboratory cultures of *E. andrei* or *D. rubidus* (i.e., cultured in house or obtained from another toxicity testing laboratory's culture and acclimated in the testing laboratory before their use in the test; see Section 2). The experimental design involves multiple replicate test vessels per treatment, with 4 adult worms added to each test vessel. Following a 28-day (four-week) exposure of adult worms, the adults are removed and numbers surviving in each test vessel and treatment are determined and recorded.[45] The test is then continued for a further 28 days without the presence of adult worms to measure effects on progeny production (i.e., number of juvenile worms). Food for the adult worms and their progeny is provided throughout the 56 days of the test.

4.2.1 Beginning the Test

The test is performed using clean and appropriately labelled 500-mL glass jars for *E. andrei* and 250-mL glass jars for *D. rubidus* as test vessels (Section 3.2.2).

All test, negative control, reference, and positive control soils must be prepared as described in Section 4.1. Immediately following the mixing of a batch, an identical wet weight of soil equivalent to a volume of ~350 mL for *E. andrei* or a volume of ~200 mL for *D. rubidus* must be transferred to each replicate test vessel[46] (Section 3.2.2). The volume of soil in each replicate test vessel must be the same. The soil added to each test vessel should be smoothed (but not compressed) using a spoon or spatula or by gently tapping the test vessel on the bench top or with a hand.

For a single-concentration test (e.g., soil tested at 100% concentration only, or a particular concentration of test soil), the minimum number of replicates must be based on the test species chosen and the *target effect size*. For *E. andrei*, the required minimum number is 21 for detection of a 40% effect size or 13 for detection of a 50% effect size. Additional replicates are recommended (see Section 5.6.2). For *D. rubidus*, the required minimum number of replicates is 13 for detection of a 30% effect size, 7 for detection of a 40% effect size, or 5 for detection of a 50% effect size. Additional replicates are recommended (see Section 5.6.2). Decisions on the minimum number of replicates were based on power analysis, and the goal is to achieve 80% *power*. For site soils, replicates should represent *replicate samples* (i.e., field replicates) collected individually from a given sample location (see Section 5.1).[47]

[45] Additional endpoints may be collected for adult worms at this point in the test (e.g., wet or dry weights, or contaminant residues). If body or tissue residues of contaminants are of interest, the worms can be frozen (-20 °C) for analyses at a later date.

[46] The wet weight of soil required to achieve a volume of ~350 mL for *E. andrei* or ~200 mL for *D. rubidus* depends on the moisture content, bulk density, and other characteristics of the soil, and will vary from sample to sample. Accordingly, the wet weight of each sample required to achieve this volume should be determined by transferring that amount of sample required to fill a preweighed (or tared) glass beaker or jar (i.e., a 500 mL jar for *E. andrei* or a 250 mL jar for *D. rubidus*) to a mark scribed on its side (reflecting the appropriate volume for the given test species), after gently smoothing (not compressing) the surface of the soil at this mark. Thereafter, the wet weight of that quantity should be determined and recorded, and an identical wet weight transferred to each replicate test vessel.

[47] Although replicate samples are recommended, the typical practice in a laboratory is to prepare "laboratory" replicates or replicate test vessels (i.e., more than one test vessel containing the same replicate sample). Power analysis was used to support decisions on minimum number of laboratory replicates, and the data for power analysis was summarized from laboratory replicates.

For a multi-concentration test, a minimum of five replicate test vessels per negative control soil and a minimum of five replicate test vessels per treatment must be set up. For any test that is intended to estimate the inhibiting concentration for a specified percent effect (*ICp*) in a definitive multi-concentration test, at least seven concentrations plus the control treatment(s) must be set up, and more (i.e., ≥ 10 plus controls) are recommended to improve the likelihood of bracketing the endpoint sought.[48] If a range-finding test is conducted prior to definitive testing, fewer concentrations may be used in the definitive test since more information on the effect concentration/dilution range will be available (see Section 4.1).

It is recommended that a minimum of one additional test vessel containing negative control soil and one additional test vessel containing reference soil and/or the lowest concentration of test soil (if a multi-concentration test) be included in the test. These extra replicates, for which data are not included in the analyses and no reporting requirements pertain, are useful in providing a preliminary assessment as to whether or not acceptable production of young in these treatments has occurred by Day 28 (see Sections 4.2.3 and 4.2.5).[49] If acceptable production of young in these treatments has not occurred by Day 28, the investigator may choose to extend the duration of exposure of adult earthworms in the definitive test vessels from 28 days to 35 days (see Section 4.2.5), in which instance the test duration would be 63 days rather than 56 days.[50]

Following the addition of a measured aliquot of test soil to each test vessel, an unperforated cover (see Section 3.2.2) should be placed over each test vessel to minimize moisture loss. The test vessels should be held overnight under test temperature and lighting conditions (Section 4.2.2) for chemical equilibration (e.g., of chemical-spiked soil or site soil diluted with control soil) of the test soils. On Day 0 (i.e., when starting the test), each cover should be perforated to allow for aeration.[51]

Test organisms (see Section 2.3.8) are transferred to each test vessel the day after the soil is prepared (i.e., Day 0 of the toxicity test). Four adult (fully clitellated) worms that are within the acceptable size range (i.e., individual wet wt of 250–600 mg for *E. andrei* and 50–200 mg for *D. rubidus*) must be used for each replicate in this test. The adult worms used in the test must be laboratory-cultured and acclimated for a minimum of seven days as described in Sections 2.3.8 and 2.4. A number of test organisms in excess of those required for the test should be removed from a culture vessel (or vessels) established to yield the appropriate number of organisms required for a test. Worms chosen for use in the test should be as similar in size (i.e., initial wet wt) as possible, based on the range of individual wet weights within the culture from which they are

[48] The use of 10 or more concentrations (plus the controls) can be used to better show the shape of the concentration-response relationship and to choose the appropriate linear or nonlinear regression model (see Section 6.4.2.1). Use of 10 or more concentrations is particularly prudent if the investigators wish to determine a 28-day LC50 for the adult worms, as well as an ICp for reproductive inhibition (see Section 4.2.7). In certain tests, the investigators might wish to focus on the sublethal endpoint and not derive a 28-day LC50, in which instance 7–9 test concentrations (plus the controls) might prove adequate for this purpose.

[49] If there is concern that the heat-extraction procedure used at the end of the test would modify the physicochemical properties of the test soil, extra replicates (with or without test organisms, depending on the objectives) should be prepared for each test concentration for the sole purpose of conducting physicochemical measurements at test end (see Section 4.2.5).

[50] With the use of Magic® Worm Food or organic mixed grains for culturing and testing, the number of juveniles has been consistently adequate with the removal adult worms at Day 28 (ECCC, 2020b). Adults left in test vessels for 35 days have led to unnecessary large numbers of juveniles to be counted at the end of the test.

[51] For a test involving a sample of contaminated soil with volatile compounds, it is recommended that opaque non-reactive covers (e.g., aluminum foil) be used as covers for the test vessels. These covers should not be perforated during the first week of the test to minimize gaseous emissions and to increase the exposure of worms to these volatile compounds. In this instance, the covers should be perforated on Day 7 (Stephenson *et al.*, 2001). If opaque (e.g., aluminum foil) covers are used in a test, the use of side lighting as well as overhead lighting is recommended to ensure that the minimal light intensity required at the surface of the soil in each test vessel is achieved (see Section 4.2.2). All test vessels, including those containing negative control soil, must be treated identically.

selected. Only those worms appearing healthy, similar in colouration, and active when removed from the bedding substrate should be selected. Earthworms should be selected from a culture vessel, removed by gloved hand or using the blunt arm(s) of rounded forceps, and transferred briefly to a clean, shallow dish or tray where they are quickly rinsed in clean test water (i.e., deionized or distilled water). Thereafter, these worms are placed into a transfer container (e.g., a glass or aluminum tray measuring ~ 10 × 10 cm) lined with paper towel dampened with test water. A final observation should be made of the worms in this container to confirm that their appearance is normal. Any atypical worms should be discarded. Thereafter, individual worms of as similar size as possible should be carefully selected while confirming that they are within the acceptable size range, and then transferred individually (by hand or using the blunt arm[s] of rounded forceps) to the surface of the soil in each test vessel. The transfer of earthworms to each test vessel should be random across replicates and treatments.

Worms are placed onto the surface of the test soil in each test vessel; four per vessel. The number of worms not burrowed into the soil in each vessel after 1 h following their introduction should be noted and recorded, for each test vessel.[52] The test vessels should be positioned such that observations and measurements can be made easily. Treatments should be positioned randomly within the test facility and the position of test vessels within the test facility should be changed regularly during the test (i.e., once every two weeks, randomly) (EC, 2005b, 2013b, 2014a; ECCC, 2020a). The dates and times test and control soils are prepared and organisms are added to the test vessels must be recorded and reported.

Individual wet weights for a minimum of twenty worms must be measured and recorded when the worms are introduced to the test vessels, to determine the variability in initial size of worms used in the test. These weights may either be based on the weights of individual worms representing the various treatments as they are weighed and transferred to the test vessels, or on surplus worms that are from the group selected for use in the test. The mean (± SD) weight for these worms must be calculated and reported (Section 7).

4.2.2 Test Conditions

- This is a 56-day whole soil toxicity test,[53] during which the soil in each test vessel is not renewed.

- The test vessel is a 500-mL (for *E. andrei*) or 250-mL (for *D. rubidus*) glass jar, and its contents (i.e., a 350-mL volume of test soil for *E. andrei* or a 200-mL volume of test soil for *D. rubidus*) are covered (Section 3.2.2).

- For a single-concentration test, the minimum number of replicates must be based on the earthworm species and the chosen effect size (Section 5.6.2). For a multi-concentration test, a minimum of five replicate test vessels per test concentration and five replicate test vessels per control soil must be set up.

- For a multi-concentration test, at least seven concentrations plus the appropriate control treatment(s) must be used, and more concentrations (i.e., ≥ 10 plus controls) are recommended.

- The test must be conducted at a daily mean temperature of 20 ± 2 °C. Additionally, the instantaneous temperature must always be 20 ± 3 °C.

- Test vessels must be illuminated with a fixed daily photoperiod (e.g., 16 h light and 8 h dark, or 12 h light and 12 h dark), and should use incandescent, fluorescent, or LED lights. The photoperiod chosen should be the same as that to which the worms are acclimated before the test

[52] A lack of burrowing might reflect an avoidance response by the worms. It could also indicate their poor condition at the start of the test. A comparison of the mean (± SD) percentage of worms burrowing in negative control soil (and, if used, reference soil) during the first hour of the test, versus percentage of worms burrowing in each test soil at that time (or thereafter; see Section 4.2.5), would provide insight into the possibility that the worms are showing an avoidance response to one or more of the test treatments.

[53] The investigator may choose to extend the duration of the test to 63 days (see Sections 4.2.1 and 4.2.5).

(see Section 2.3.3 and 2.4.3). Light intensity adjacent to the surface of the soil in each test vessel should be 400–800 lux, and must be at least 400 lux as a minimum.

- Worms in each test vessel must be fed an identical quantity of food (see Section 4.2.4), on Days 0, 14, 28, and 42 only.

4.2.3 Criteria for a Valid Test

For the results of this biological test method to be considered valid, each of the two following criteria must be achieved:[54]

i) the mean survival for the adult worms held in negative control soil for 28 days (or 35 days) must be ≥ 90%,[55] and

ii) the reproduction for adult worms in negative control soil must average ≥ 3 live juveniles per adult at the end of the test.

4.2.4 Food and Feeding

During a toxicity test, earthworms in each test vessel must be fed the same food to which they were acclimated for the 7 days (or 14 days) prior to testing (Section 2.4.6). Magic® Worm Food or organic mixed grains must be used for the 56-day reproduction test described herein (see Section 2.3.6).[56] Worms in each test vessel must be fed a measured quantity of food on each of the following days of the test, only: Day 0, Day 14, Day 28, and Day 42. The same quantity of food must be added to each test vessel at a given feeding. For Magic® Worm Food or organic mixed grains, a 2-g portion per *E. andrei* test vessel and 1-g portion per *D. rubidus* test vessel is recommended.

On Day 0 (i.e., when starting the test), before adding the earthworms to the test vessels, a small hollow should be made in the centre of the soil surface within each vessel. An appropriate volume of food should be placed into this depression, hydrated with a small amount of deionized water, and covered with a thin layer of surrounding soil to reduce fungal growth. The adult earthworms should then be added, four per vessel (see Section 4.2.1), and the vessels covered with perforated lids (see Section 3.2.2). On Day 14, the cover of each test vessel is removed and an additional aliquot of food added (as per the procedure for Day 0). On Day 28, following the removal of adult worms and the return of the remainder of the contents of the jar to each test vessel (Section 4.2.5), another aliquot of food should be added to each test vessel (for development and growth of their progeny) in the same manner as before. A final aliquot of food should be added to each test vessel on Day 42. When adding food to the test vessels on Days 14, 28, and 42, any old food evident in the surficial layer of the soil within each test vessel should be left undisturbed (since hatchling worms are frequently found in and around the food). Less food may be used if a large amount of uneaten food remains following the previous feeding; however, all test vessels must be treated equally (i.e., all test vessels are fed the same amount of food at a given feeding).[57]

---

[54] ISO (1998) and OECD (2000) used this (or an equivalent value) as a criterion for a valid test for effects of chemicals on the reproduction of *E. andrei*. For *D. rubidus*, the test validity criteria presented here are based on control data generated in many studies carried out during the development of these methods for this species (EC, 2010; ECCC, 2020b). Clean soils included in the development of the test validity criteria included an artificial soil, one agricultural soil, and seven boreal soils (including thirteen different horizons in total; see Appendix F). The validity criteria were based on a calculation of the 5th percentile of survival and reproduction data for these clean soils (ECCC, 2020b).

[55] If the duration of exposure of adult worms is extended to 35 days (see Sections 4.2.1 and 4.2.5), this survival criteria must still be applied.

[56] In the first edition of this test method document, hydrated, cooked oatmeal was the food required for use in the earthworm survival, reproduction, and growth test. Subsequent research carried out at ECCC indicated that lower variability, higher reproduction, and fewer test failures were achieved by using Magic® Worm Food or organic mixed grains for culturing and testing with *E. andrei* and *D. rubidus* (ECCC, 2020b).

[57] If, on Days 14 and/or 28 only, uneaten food is evident within the surficial layer of the soil in certain or all test vessels representing any treatment, the amount of food provided to all replicate test vessels (and treatments) in the test should be reduced. The discretionary practice of reducing the amount fed at these times is advisable to avoid the risk of overfeeding and the risk of excessive mould or sorption of toxic contaminants caused by uneaten food. Feeding should without exception be

4.2.5 Observations and Measurements During the Test

The biological endpoints for this test are the number of surviving adult worms at Day 28, and the number of progeny produced in each test vessel at the end of the test (Day 56). *Biomass* metrics (i.e., wet and/or dry mass) of progeny are optional (see footnote 68 in Section 4.2.6). The condition, appearance, and number of live worms transferred to each test vessel on Day 0 must be observed and recorded. At 1 h following their transfer, the number of worms on the surface of the soil in each jar, or against the glass on the inner sides or bottom of each test vessel, should be noted and recorded (see Section 4.2.1).

At the time of each feeding (see Section 4.2.4), the cover should be removed from each test vessel, and observations and records made of the number of live or dead worms on the surface of the soil therein. Thereafter, the surficial layer of the soil in each test vessel should be examined to appraise the presence and quantity of any uneaten food and to reach a decision regarding the need to reduce the amount of food provided (see Section 4.2.4). Observations and records should also be made at this time of the number of worms seen inside each test vessel on its glass sides or bottom (this might be an indication of an avoidance response to the soil therein).

On Day 28 of the test, the covers of any "extra" test vessels used to determine if acceptable production of progeny in these treatments has occurred by this time (see Section 4.2.1) should be removed.[58] The contents of these "extra" test vessels should be examined for the presence of cocoons or juvenile worms. If any cocoons and/or juveniles are observed in each of these treatments, the cover of each of the definitive test vessels should be removed and its contents examined (see next paragraph). If cocoons or juvenile worms are not observed in the "extra" test vessels representing each of these treatments, it is recommended that the definitive test vessels be left undisturbed for an additional seven days before their examination for and removal of adults (ESG, 2001, 2002; ESG and Aquaterra Environmental, 2002). In this instance, the contents of the "extra" test vessels (including all cocoons, and any live juvenile and adult worms) should be returned to the test vessels and held under test conditions until they are re-examined. Thereafter (i.e., on Day 28 or, in some instances, Day 35; see preceding paragraph), the cover of each definitive test vessel must be removed, as should the covers of each "extra" test vessel. The number of live and dead adult worms on the surface of the soil or against the glass on the inner sides or bottom of each test vessel should be observed and recorded. Thereafter, the contents of each test vessel must be transferred to a sorting tray or plastic sheeting, and the number of live and dead adult worms counted and recorded. Adults appearing to be dead should be touched gently on their anterior end with a glass rod or spatula; absence of any response is defined as death. Missing adults must be counted as dead. The appearance (e.g., normal or signs of discolouration or lesions, de-clitellation, or pinching) and behaviour (e.g., normally active, coiling, or lethargic) of each surviving adult should be noted and recorded.[59] Immediately after this evaluation, and having removed the adults, the test soil is returned to the jar together with any cocoons and juvenile worms therein. The cover of each test vessel must be replaced. Observations of the number of juvenile worms produced during the test, appearance, and behaviour must be made 28 days later, at the end of the test (i.e., on Day 56 or, in some instances, on Day 63).

Air temperature in the test facility (Section 3.1) must be measured daily (e.g., using a maximum/minimum thermometer) or continuously (e.g., using a continuous chart recorder).

For each treatment, the contents of one or more replicate test vessels must be examined at least once

---

provided on each of the scheduled days (i.e., on Days 0, 14, 28, and 42); however, the amounts might vary between feedings based on observations of food that has not been consumed.

[58] These "extra" test vessels might include one or more additional jars containing negative control soil, one or more jars containing reference soil (if included in the definitive test), and one or more jars containing the lowest concentration of a test soil if a multi-concentration test.

[59] The de-clitellation or pinching of adult worms is not always affiliated with lower juvenile production (ECCC, 2020b).

every two weeks for apparent "wetness."[60] Moisture loss should be determined by weighing test vessels. All test vessels can be weighed at the beginning of the test. The weight of each test vessel can then be checked at least once per week, and test water added to compensate for weight loss (i.e., due to water loss) if the loss is >10% of the initial water content (ISO, 1999). For a large number of test vessels, the average amount of water lost can be calculated by weighing a random sample of 10–20% of the test vessels at the beginning of the test and once per week thereafter. This amount of test water can then be added to all of the test vessels. Soil should be moistened by spraying it with test water using a fine-spray mister that dispenses about 1 mL of water per spray. The apparent "wetness" of each test soil can be judged at the same time food is added to test vessels (i.e., on Days 0, 14, 28, and 42), and soils hydrated as necessary.

The pH and moisture content of the test soil or soil horizon representing each treatment (including the negative control soil and, if used, reference soil) must be measured and recorded at the beginning and end of the test. Additionally, it is recommended that electrical conductivity be measured at the beginning and end of the test in instances where the test soil is anticipated to have a high salt content. The initial (Day 0) measurements should be made using a *composite sample* made up of subsamples of each batch of test soil or soil horizon used to set up replicates of a particular treatment (see Section 4.2.1).[61] The final (i.e., Day 56 or, in certain instances, Day 63) measurements should be made using subsamples of the replicates of each treatment to which worms were exposed, following the end-of-test observations of worm distribution within the test vessel, number of live juveniles, and their appearance and behaviour (see Section 4.2.6).

Soil pH should be measured using a $CaCl_2$ slurry method (modified from Hendershot *et al*., 1993; as recommended by Becker-van Slooten *et al*., 2004).[62] For these analyses, 4 g of hydrated soil[63] are placed into a 30-mL glass beaker (~3 cm in diameter and ~7 cm high) with 20 mL of 0.01 *M* $CaCl_2$.[64] The suspension should be stirred intermittently for 30 min (e.g., once every 6 min). The slurry should then be left undisturbed for ~1 h. Thereafter, a pH probe is immersed into the supernatant and the pH recorded once the meter reading is constant.

---

[60] The apparent "wetness" of a soil is affected by the degree of activity of the worms in the soil, as well as by the nature of the soil and the amount of water lost from test vessels due to evaporation. Typically, as much as 3 mL per test vessel might be lost weekly due to evaporation. However, biweekly additions of water frequently result in the soil being too wet at test end. Soils might appear too dry when the WHC has been underestimated (see Section 5.3). Any decision as to whether or not to spray water onto the surface of the soil in each test vessel can also be made based on "apparent wetness" of the test soil during each period of observation (G.L. Stephenson, Aquaterra Environmental, personal communication, 2001). To assess this, a qualitative "squeeze test" (see Section 5.3) can be applied to a small quantity (i.e., a "pinch") of the surficial soil within the test vessel. If no water appears, the soil is likely too dry. In this instance, the surface of the soil in the test vessel should be lightly misted.

[61] On the day before the start of the test (Day -1), one or more additional replicates of each test soil can be placed into a test vessel within the test facility. These replicates can be reserved for physicochemical analyses of Day 0 conditions to which the worms are exposed. A separate set of replicates can also be set up on Day -1, for physicochemical analyses of test end conditions. These additional replicates might or might not have worms added on Day 0.

[62] The method by Hendershot *et al*. (1993) includes a step that involves air drying the sample for 48 hours before its analysis for pH. The experience by Environment Canada investigators is that this step is needlessly time consuming (K. Doe, Environment Canada, personal communication, 2004; J. Princz, Environment Canada, personal communication, 2004), and does not appreciably modify the pH relative to that for hydrated (i.e., as per the toxicity test) soil (Courchesne *et al.*, 1995; J. Princz, Environment Canada, personal communication, 2004).

[63] It might be necessary to use a lower soil:$CaCl_2$ solution ratio (e.g., 2 g of soil with 20 mL of $CaCl_2$) for soils with a high organic matter content (i.e., for soils where the slurry does not yield a supernatant).

[64] To prepare 0.01 *M* $CaCl_2$, dissolve 2.940 g of calcium chloride dihydrate ($CaCl_2 \cdot 2H_2O$) with distilled water, in a 2000-mL volumetric flask. The electrical conductivity of the $CaCl_2$ solution should be between 224 and 240 mS/m at 25 °C, and the pH should range within 5.5–6.5 at 25 °C (Hendershot *et al.*, 1993). If the pH is outside this range, it should be adjusted to the ranges using a hydrogen chloride (HCl) or calcium hydroxide [$Ca(OH)_2$] solution. If the electrical conductivity is not within the acceptable range, a new solution must be prepared.

The moisture content of each test soil or soil horizon should be measured by placing a 3–5-g subsample of each test soil or soil horizon into a pre-weighed aluminum weighing pan, and measuring and recording the wet weight of the subsample. Each subsample should then be placed into a drying oven at 105 °C until a constant weight is achieved; this usually requires a minimum of 24 hours. The dry weight of each subsample should then be measured and recorded. Soil moisture content must be calculated (on a dry-weight basis) by expressing the moisture content as a percentage of the soil dry weight:

$$\text{Moisture content (\%)} = \frac{\text{wet weight (g)} - \text{dry weight (g)}}{\text{dry weight (g)}} \times 100$$

It is important that the calculation of moisture content (%) be based on dry weight (**not** on wet weight), since the results of these calculations are used with calculations of water-holding capacity (also calculated on dry weight) to express the optimal moisture content in test soils (see Section 5.3).

Depending on the nature of the test and the study design, concentrations of chemical(s) or chemical product(s) of concern might be measured for test soils or selected concentrations thereof, at the beginning and end of the test. For a test using a sample of field-collected site soil, the chemical(s) or chemical product(s) measured will depend on the contaminant(s) of concern (see Section 5.5). For a multi-concentration test with chemical-spiked soil, such measurements should be made for the high, medium, and low concentrations tested, as a minimum (see Section 6.3). Aliquots for these analyses should be taken for each soil or soil horizon as described previously for pH, electrical conductivity, and moisture content;[65] analyses should be according to proven and recognized analytical techniques (e.g., SPAC, 1992; Carter, 1993; Carter and Gregorich, 2008).

4.2.6 Ending the Test

The test must be terminated 28 days after the adults were removed (i.e., on Day 56 if adults were removed on Day 28, and on Day 63 if adults were removed on Day 35; see Section 4.2.5). To terminate this soil toxicity test, the number of live juvenile worms observed on the surface of the soil in each definitive test vessel, or adjacent to the glass on its sides or bottom, should be first be determined and recorded. Thereafter, the number of live juveniles within each test vessel must be counted and recorded.

There are two approaches to recovering juvenile worms from test soil at the end of the test: 1) manual sorting, and 2) heat extraction. For manual sorting, which is the preferred option for juvenile recovery, the entire contents of the soil within each test vessel must be carefully sorted while recovering and counting the number of live juvenile earthworms. For this procedure, the soil is carefully removed from the jar, and placed onto a sorting tray or plastic sheeting. The test soil is carefully sorted using blunt-nosed forceps or other appropriate tools, while recovering all surviving juvenile worms. This process can be repeated two or three times for each replicate soil to ensure that all worms have been recovered. Using this procedure, the number of hatched and unhatched cocoons in the test soil can also be determined. Observations of numbers of unhatched or hatched cocoons at test end, although not included as a test endpoint (see Section 4.2.7), might prove useful for certain tests in discerning adverse effects on (delayed) development of earthworms or early survival of young produced. To enable these (optional) observations, the number of hatched or unhatched cocoons found within the contents of each test vessel should be counted and recorded.[66]

---

[65] The amount of soil collected for analyses of the chemicals or contaminants of concern depends on the method of analyses. Different volumes or weights are required for different methods and might range from 1 g to 50 g of soil or more for poor extraction efficiency. The amount of soil required for analyses should be determined a priori in consultation with the analytical laboratory, and collection volumes adjusted as necessary (G.L. Stephenson, Aquaterra Environmental, personal communication, 2021).

[66] Hatched cocoons are hollow with a translucent appearance, and are easily dented or collapsed by pressing gently with forceps. Conversely, unhatched cocoons usually are turgid, have an opaque appearance, and their exterior covering is not easily dented with gentle forceps pressure, although they will pop open with greater pressure.

The second approach to recovering juveniles from the soil in each test vessel at the end of the test is a novel "heat-extraction" procedure, designed by Stantec and Aquaterra Environmental as part of the developmental studies associated with the standardization of the first edition of this biological test method document. This heat-extraction procedure proved efficient and effective for the recovery of *E. andrei*, and is provided as an alternative to manual sorting for use with this test method.[67] To perform this recovery procedure, the test jars are transferred sequentially to a heated (40–45 °C) water bath, while ensuring that the level of the water in the bath does not rise higher than half of the height of the soil in each jar. Each test jar is left in the water bath for no more than 15 minutes. Thereafter, the surficial 2-cm layer of soil in the jar is carefully removed, and placed onto a sorting tray or plastic sheeting. This subsample of the test soil is then manually sorted, as described above, while recovering all surviving juvenile worms. Laboratories that are not experienced with the heat-extraction procedure described must initially validate and document the efficiency of their heat-extraction system (i.e., demonstrate and record data that show that a significant number of test organisms are not being left in the soil following heat extraction). This can be accomplished by further processing the heat-extracted soil for test organisms by manually sifting through the soil to check on the efficiency of the heat-extraction technique. The heat-extraction process is considered acceptable if there are < 5% of the total number of test organisms remaining in the soil (i.e., extracted from the soil using manual sorting of the soil, following heat extraction). If the heat-extraction efficiency is not acceptable, all treatments must be processed in a similar matter (i.e., using manual sorting following heat extraction). Once laboratory personnel are experienced with heat extraction and have demonstrated the efficiency of their system, they should continue *monitoring* the efficiency periodically. The heat-extraction procedure for recovering test organisms is not suitable for any test involving the recovery of cocoons.

Any worms appearing to be dead should be touched gently on their anterior end with a glass rod or spatula; absence of any response is defined as death. Dead juvenile worms, if observed, are recorded, but they must not be included in the juvenile count. For each test vessel, the appearance (e.g., normal or signs of discolouration or lesions) and behaviour (e.g., normally active or lethargic) of the surviving worms should be noted and recorded.[68]

---

[67] The use of heat to improve the efficiency of recovering juvenile earthworms from test vessels at the end of a reproduction test was recommended by Dr. Kees van Gestel (Institute of Ecological Science, Amsterdam, Netherlands). This procedure causes the surviving worms to move to the cooler soil above the height of the water in the water bath, enabling their efficient recovery. The advantage of using this procedure is that the amount of soil to sort through is appreciably less than the full quantity (~350 mL) in the test vessel. Stantec and Aquaterra Environmental (2004) found that the time required to recover surviving earthworms from each test vessel, while manually sorting the soil, was reduced from about 30–40 minutes per test vessel (if the entire contents were sorted) to only 15–20 minutes per test vessel using the heat-extraction procedure. The percent recovery of earthworms using this procedure was typically 100%, and in all instances ≥ 94% of the surviving worms were recovered by the heat-extraction technique (Stantec and Aquaterra Environmental, 2004). At the time of publication, this heat-recovery system had not been tested for recovery of *D. rubidus* from test soils, so it is not recommended for use with this test species.

[68] If the measurement of juvenile dry weights is warranted, the following procedure is recommended: Separate weighing pans, each containing the group of surviving juveniles recovered from each test vessel, are placed into an oven and dried at 90 °C until a constant weight is achieved (this usually takes a minimum of 48 hours) (Aquaterra Environmental and ESG, 2000). It is important that the system used to process juveniles for dry weight measurement is as standardized as possible. Inconsistencies that have been identified include: the length of time worms are being held while replicates are being processed at the end of the test, how the soil is removed from the worms before being dried and weighed, and resolution of balance stabilization inconsistencies and of the potential loss of worms from weigh boats due to static electricity during weighing (MESI, 2014). Upon removal from the oven, the weighing pans are moved immediately to a desiccator. Following cooling, each weighing pan should be individually and randomly removed from the desiccator, and weighed immediately to the nearest 0.1 mg on a balance that measures accurately to this limit. Mean dry weight per surviving juvenile worm is calculated for each group (see footnote 74 in Section 4.2.7).

During the series of dry weight determinations for the groups of surviving juveniles from a test, the first weighing pan should be replaced in the desiccator and weighed again

Test vessels, irrespective of concentration levels, should be processed in a random manner since counting might become more or less accurate. Following the recovery of juvenile worms from each test vessel, subsamples of each test soil (including the negative control soil and, if included in the test, reference soil) should be taken for determinations of pH and moisture content (Section 4.2.5). Analyses for other chemical constituents (i.e., concentrations of contaminants) should also be made at this time using representative subsamples of each test soil (Section 4.2.5).

4.2.7 Test Endpoints and Calculations

For each test, the percent survival of adult worms in each test vessel exposed to each treatment for 28 days must be calculated. The mean (± SD) percent survival for all adult worms exposed to each concentration (including the negative control soil, and if used, reference soil) for 28 days must be calculated and reported.[69]

The reproductive endpoint for this test is based on the number of surviving progeny (i.e., juveniles) produced in each replicate and each treatment during the 56-day test period. A statistically significant reduction in this number is considered indicative of an adverse toxic effect of the treatment on the reproductive success of the adult worms. The mean (± SD) number of surviving juveniles in the test soil on Day 56 (or Day 63) must be determined and reported for each treatment (including reference and all control soils [negative, solvent, and positive control soils]).[70] In addition, the mean (± SD) number of surviving juveniles produced by each adult worm in the control(s) on Day 56 (or Day 63, if applicable) must be calculated and reported.

The two most common possibilities for a typical test design involve:

i) Soil from multiple *sampling locations*, in which responses at one or more test site sampling locations are compared with those at a *reference site* sampling location,[71] with other test sampling locations, or with the control soil (i.e., single-concentration test). Hypothesis testing is frequently used in the statistical assessment, and the common outcome is that a response at a sampling location is either "different" or "not different" from another sampling location (Section 5.6.1).

ii) Multiple concentrations of a test soil, achieved by mixing a test soil with reference or control soil (Section 5.3), or by *spiking* a soil with various concentrations of a chemical or chemical product (Section 6.2). For a multi-concentration test, the 56-day (or 63-day, if applicable) ICp for reproductive inhibition must be calculated and reported (data permitting).[72]

---

at the end of all weighings. This serves as a check on any sequential gain of water by the weighing pans in the desiccator over time, which might occur when each weighing pan is removed for its weight determination. The change in weight of the first weighing pan over time should not be >5%; if it is, redrying of all weighing pans for ≥ 2 h and reweighing should be carried out.

[69] These calculations are made when the adults are removed from the test vessels (i.e., after 35 days if adults are left in the soil for an additional 7 days; see Sections 4.2.1 and 4.2.5).

[70] These calculations are made at the end of the test (i.e., at Day 63 if adults are left in the soil for an additional 7 days; see Sections 4.2.1 and 4.2.5).

[71] Throughout this document, reference site is used to describe an area in which there is clean soil uninfluenced by the contaminant under study (i.e., reference soil). A reference soil should be collected for these comparisons, as described in Section 5. However, in the absence of a reference soil, a negative control soil may be substituted.

[72] Historically, investigators have analyzed quantitative sublethal data from multi-concentration tests by calculating the *no-observed-effect concentration* (*NOEC*) and the *lowest-observed-effect concentration* (*LOEC*). Disadvantages of these statistical endpoints include their dependence on the test concentrations chosen and the inability to provide any indication of precision (i.e., no 95% or other confidence limits can be derived) (NERI, 1993; EC, 2005a). Given these disadvantages, ICp is the required statistical endpoint for reproduction data derived in a multi-concentration test using *E. andrei* and *D. rubidus*. Contrary to recent criticism blaming the continued generation and publication of NOEC/LOEC data on the failure of governments and international organizations to formally discredit and cease recommending these approaches (van Dam *et al.*, 2012), it is evident that Environment and Climate Change Canada has fully adopted regression-based methods in aquatic-, sediment-, and soil-based environmental toxicity testing (EC, 2005a, 2005b, 2007a, 2007b, 2011a, 2011b, 2013b, 2014a; ECCC, 2020a; Van der Vliet *et al.*, 2012).

In a scenario where there are multiple sampling locations, an understanding of the strengths of various study designs is critical for the successful application of statistical tests. The study objectives should be clearly defined before data are collected, with an appreciation both for the power (ability to detect an effect) of the test design and the ease of interpretation of the results. In general, it is advantageous to limit the number of comparisons made, and this is typically done by choosing a test design and statistical tests that compare test sampling locations with a reference sampling location. Further gains in power can be made if a gradient can be assumed (i.e., samples collected in sequential order away from the point source; see Section P.4 in EC, 2005a). In some cases, study objectives and test design might not have been given adequate attention before the collection of the data, and to compensate, investigators will perform a comparison among all possible sampling locations, maximizing the number of comparisons made. This is strongly discouraged, particularly when large numbers of sampling locations are involved, because undesirable effects on *Type I* and *Type II error* rates might occur; interpretation of results is often more difficult; and, unwarranted focus might be given to particular comparisons after data have been collected. Detailed statistical guidance on hypothesis testing for the number of progeny at test end is provided in Section 5.6 and EC (2005a).

For a single-concentration test (see Sections 5.3 and 6.2), the mean (± SD) value for the percent survival of adult worms on Day 28 (or Day 35, if applicable), as determined for each treatment, is compared with that for the sample(s) of reference soil or, as necessary and appropriate, compared with that for the negative control soil. For a multi-concentration test (see Sections 5.3 and 6.2), the 28-day (or 35-day, if applicable) LC50 (including 95% confidence limits) for adult survival must be calculated and reported if sufficient data are available to enable this. Environment Canada (2005a) provides guidance for calculating LCps, which should be followed; Section 6.4.1 gives further guidance in this regard.

For a single-concentration test (see Sections 5.3 and 6.2), the mean (± SD) value for number of surviving juveniles in the test soil at the end of the test (i.e., Day 56 or Day 63) is determined and compared to that for the sample(s) of reference soil or, as necessary and appropriate, compared to that for the negative control soil. For a multi-concentration test (see Sections 5.3 and 6.2), the 56-day (or 63-day, if applicable) ICp for reproductive inhibition (including 95% confidence limits) must be calculated and reported (data permitting). Environment Canada (2005a) provides direction and advice for calculating ICp endpoints, which should be followed, and Section 6.4.2 herein give further guidance in this regard. Initially, regression techniques (see Section 6.4.2.1) must be applied to multi-concentration data intended for calculation of an ICp.[73] In the event that the data do not lend themselves to calculating the 56-day (or 63-day, if applicable) ICp for the reproductive inhibition using the appropriate regression analysis, linear interpolation of these data using the program ICPIN should be applied in an attempt to derive an ICp (see Section 6.4.2.2).[74]

---

[73] Regression is the method of choice for estimating an ICp. It involves fitting the data mathematically to a selected model and then calculating the statistical endpoint using the model that best describes the exposure-concentration response relationship. Nonlinear regression techniques were originally recommended by Stephenson *et al.* (2000b) for several reasons including: the relationship that exists between exposure concentration and earthworm reproduction responses is typically nonlinear; the *heteroscedasticity* of the data is rarely reduced by transformation; the more standard bootstrap simulation technique has several limitations for these types of data; and nonlinear regression can fit effect distributions showing *hormesis*. By using standard mathematical techniques, a regression can be well-described in terms that convey useful information to others, effects at high and low concentrations can be predicted, and confidence intervals can be estimated. Deficiencies of the smoothing and interpolation method can be largely remedied (EC, 2005a).

[74] If measured, the growth endpoint for this test is based on the mean dry weight of individual juvenile worms produced in each treatment that survived the 56-day test period. A significant reduction in this weight is considered indicative of an adverse toxic effect of the treatment on the growth of surviving juveniles produced by the adult worms used to start the test. For a single-concentration test (see Sections 5.3 and 6.2), the mean (± SD) value for dry weight of individual juveniles surviving in the test soil on Day 56 is determined and compared to that for the sample(s) of reference soil or, as necessary and appropriate, compared to

An initial plot of the raw data (percent adult mortality and number of surviving progeny) against the logarithm of concentration is highly recommended, both for a visual representation of the data and to check for reasonable results by comparison with later statistical computations. Any major disparity between the approximate graphic LCp and ICp and the subsequent computer-derived LCp and ICp must be resolved. The graph would also show whether a logical relationship was obtained between log concentrations (or, in certain instances, concentration) and effect, a desirable feature of a valid test (EC, 2005a).

4.3 Avoidance Test

This biological test method uses adult earthworms (*E. andrei* or *D. rubidus*) as test organisms, and measures avoidance of test soils[75] as the biological endpoint.

Table 4 provides a summary checklist of required and recommended conditions and procedures to be universally applied to each avoidance test with samples of contaminated or potentially contaminated soil (including boreal and taiga soils), as well as those for testing specific types of test materials or substances (e.g., samples of biosolids, or negative control soil spiked in the laboratory with one or more test chemicals or chemical products).

Universal procedures for performing an avoidance test are described in this section. Test organisms are laboratory-cultured adult *E. andrei* or *D. rubidus* (i.e., cultured in house or obtained from another toxicity testing laboratory's culture and acclimated in the testing laboratory before their use in the test; see Section 2).

Test duration is 48 hours.[76] The test organisms are not fed during the test, and the test soils are not renewed (i.e., static).

4.3.1 Beginning the Test

The test is performed using specially designed avoidance test units described in Section 3.2.3. A minimum of five replicates (i.e., 5 test units) is required for a single-concentration test, with each unit containing the same two treatments (i.e., a single site soil or a single concentration of test soil, plus a negative control soil or clean reference soil) in alternating compartments. For contaminated site soils, replicates should ideally represent replicate samples (i.e., field replicates) collected individually from a given sample location (see Section 5.1). However, the degree of replication is dependent on the purpose of testing. If the avoidance test is being used to screen soils for positive effects to inform decisions on further toxicological testing (e.g., 56-day reproduction test), fewer replicates (e.g., 1 test unit) may be used. For a multi-concentration test, a minimum of two replicates (i.e., 2 test units) per test concentration is required, with each test unit containing the same two treatments (i.e., a single test concentration together with aliquots of a clean soil) in alternating compartments. The use of more replicates (e.g., $\geq$ 3 test units) per concentration in a multi-concentration test could provide a more

---

that for the negative control soil. For a multi-concentration test (see Sections 5.3 and 6.2), the 56-day (or 63-day, in some instances) ICp for growth inhibition (i.e., decreased mean dry weights of individual juveniles) can be calculated and reported, data permitting. The direction and advice in Section 6.4.2 for calculating ICps is applicable and should be followed. In this regard, the approach described herein for calculating a 56-day ICp for reproductive inhibition applies equally here, when calculating a 56-day ICp for growth inhibition.

[75] Using the recommended test apparatus defined and illustrated in Section 3.2.3 (Figures 2 and 3), groups of earthworms are given a choice between negative control soil or reference soil and a test soil (e.g., a field-collected soil from a contaminated or potentially contaminated site, or a particular concentration of a chemical or chemical product spiked into negative control soil). Each worm (n = 10) within a test unit is free to move between the clean soil (i.e., negative control soil or reference soil) held in three compartments and the test soil held in three alternating compartments (six compartments/test unit) therein. At the end of a defined exposure period (i.e., 48 hours), the total numbers of worms in the clean soil and the test soil are determined and compared statistically (see Section 4.3.7).

[76] A test duration of 48 hours was initially chosen for *E. andrei* to harmonize with the avoidance test using *E. fetida* or *E. andrei* published by ISO (2003). (Aquaterra Environmental, 1998; Stephenson *et al.*, 1998). More recently, ECCC's Soil Toxicology Laboratory investigated reducing the avoidance test duration for both *E. andrei* and *D. rubidus* to 24 hours. Results of this investigation in a boric-acid-contaminated soil showed that 24 hours was insufficient for a trend in avoidance to become apparent, but there were clear dose-response effects at 48 hours (ECCC, 2020b).

Table 4 Checklist of required and recommended conditions and procedures for conducting tests for effects of exposure to contaminated soil on the avoidance behaviour of earthworms (*E. andrei* or *D. rubidus*)

Universal		
Test type	−	whole soil toxicity test; no renewal (static test)
Test duration	−	48 hours
Test organisms	−	Laboratory-cultured *E. andrei* or *D. rubidus*; sexually mature adults with clitellum; individual wet wt of 250–600 mg for *E. andrei*, or 50–200 mg for *D. rubidus*; choose worms as similar in wet wt as possible; acclimate for ≥ 7 days in negative control soil that is to be used in the test; 10 worms per test unit
Negative control soil	−	depends on study design and objectives; clean field-collected soil or artificial soil if testing site soils; recommend artificial soil for tests with chemicals or chemical products spiked in soil
Test unit	−	circular container with central chamber (inner diameter ~5.4 cm) and six pie-shaped interconnecting compartments, with fitted lid; constructed of high-quality stainless steel or Plexiglas™; outer diameter ~23 cm, height ~9 cm, 1.0-cm holes in bottom of central chamber (two/compartment) and sides of compartments (three/side) for worm movement; modified with false back plates for use with *D. rubidus*
Amount of soil per central chamber	−	none
Amount of soil per test compartment	−	identical wet wt, equivalent to a volume of ~350 mL for *E. andrei* or ~200 mL for *D. rubidus*
Moisture content, test soils	−	hydrate to the optimal percentage of its WHC if field-collected soil (see Section 5.3), or to ~70% of WHC if artificial soil (see Sections 3.3.2 and 6.2)
Number of compartments per test unit with same treatment	−	three (negative control soil or reference soil in each of three compartments, with a single test soil or concentration thereof in each of three alternate compartments)
Number of treatments per test unit	−	two (negative control soil or reference soil, plus a single sample or concentration of a test soil (e.g., a field-collected test soil at 100% or lower concentration, or a single concentration of a chemical-spiked soil)); alternate treatment in each neighbouring compartment
Number of replicate test units per test soil or concentration	−	≥ 2 test units per test soil or concentration for a multi-concentration test; ≥ 5 test units per test soil or concentration for a single-concentration test; fewer replicate test units (e.g., 1 test unit) for screening or range-finding tests

Number of concentrations	−	for multi-concentration test: ≥5 concentrations, plus control(s); more recommended (≥7)
Air temperature	−	daily average, 20 ± 2 °C; instantaneous, 20 ± 3 °C
Lighting	−	continuous darkness (test units wrapped in aluminum foil if made of transparent or translucent Plexiglas™)
Feeding	−	do not feed
Measurements during test	−	air temperature in test facility, daily or continuously; moisture content, pH, and electrical conductivity (if necessary) of each test soil/concentration, at start and end
Observations during test	−	compartment (treatment) entered by each worm at start of test; number of live worms in each test compartment at test end; number of dead worms in each test compartment at test end; number of live or dead worms seen on surface of soil in each test compartment at test end; obvious pathological symptoms (e.g., open wounds) or distinct behavioural abnormalities (e.g., lethargy) for surviving worms in each test compartment at test end
Biological endpoint	−	number of live worms per treatment in each test unit (i.e., total number of live worms in the three compartments containing the same test soil, for each of the two treatments) at test end
Statistical endpoints	−	percent survival of all earthworms in each test unit at test end; for more than one replicate test unit, mean percent survival of all earthworms per test unit for each test soil or test concentration at test end
	−	for single-concentration tests: mean (± SD) number of surviving worms recovered from the test soil and the clean soil in each of the replicate test units; percent avoidance, if calculated
	−	for multi-concentration tests: percent avoidance per treatment in each test unit at test end; 48-hour EC50 (and any other ECp)
Test validity	−	invalid if percent survival of worms in any test unit <90% at test end; for more than one replicate test unit per test soil or concentration, invalid if mean percent survival of worms per test unit is <90% for each test soil or concentration

Test with reference toxicant	− choose between a positive control concentration or a multi-concentration reference toxicity test:
	  o If the positive control option is chosen, it must be performed with every definitive test; use boric acid (H_3BO_3) or similar; prepare and test ≥ 3 replicate units of a predetermined concentration plus a negative control, using artificial soil as a substrate; 10 worms per replicate; follow procedures and conditions described in Section 4.4 and Appendix H; determine % avoidance at test end (i.e., 48 hours)
	  o If the multi-concentration reference toxicity test option is chosen, it must be performed twice per year; use boric acid (H_3BO_3) or similar; prepare and test ≥ 5 concentrations plus a negative control, using artificial soil as substrate; ≥ 2 replicates units/concentration and 10 worms/replicate; follow procedures and conditions described in Section 4.4; determine 48-hour EC50 for avoidance response (including 95% confidence limits); express as mg boric acid/kg dry wt; validity criteria are the same as those for definitive test
	− alternatively, under certain circumstances (i.e., when the test is conducted infrequently; see Section 4.4), the positive control and/or multi-concentration reference toxicity testing options for the 56-day reproduction test may be applied to satisfy the reference toxicity testing requirements for the avoidance test
	− no reference toxicity testing is required if the avoidance test is used for screening or range-finding purposes only

Field-Collected Soil

Transport and storage	− seal in plastic or other appropriate material, and minimize air space; labelled or coded; transport in darkness (e.g., using an opaque cooler, plastic pail, or other light-tight container); do not freeze or overheat during transport; store in dark at 4 ± 2 °C; test should start within two weeks, and must start within six weeks unless soil contaminants are known to be stable
Negative control soil	− either natural, uncontaminated field-collected soil or artificial soil, for which previous avoidance tests with the test species showed that the criterion for test validity could be regularly met; analyzed for at least the following: particle sizes (% sand, % silt, % clay), TOC (%), OM (%), pH, electrical conductivity, moisture content (%), WHC, and CEC
Reference soil	− one or more samples for tests with field-collected soil; ideally taken from site(s) presumed to be clean but near sites of test soil collection; characteristics (TOC [%], OM [%], particle size distribution, texture, pH, and electrical conductivity) similar to test soil(s); analyzed as described for natural negative control soil

Characterization of test soils	−	must include at least moisture content (%), WHC, pH, electrical conductivity, TOC (%), OM (%), particle sizes (% sand, % silt, % clay), and CEC; should include at least nitrogen, phosphorus, potassium, C:N ratio, major cations and anions; and, optionally, bulk density, total inorganic carbon, total volatile solids, biochemical oxygen demand, chemical oxygen demand, redox potential, soluble salts, metal oxides, sodium adsorption ratio, contaminants of concern (e.g., metals, polycyclic aromatic hydrocarbons, pesticides), and characteristics of the contamination (e.g., odour, staining, debris, presence of fuel or solvent)
Preparation of test soils	−	if necessary, remove debris and indigenous macro-organisms using forceps; if necessary, gently pass through a sieve of suitable mesh size (e.g., 4–10 mm); homogenize; determine percent moisture content and WHC; hydrate with test water (or, if and as necessary, dehydrate) to the optimal percentage of its WHC (see Section 5.3); mix; dilute with negative control or reference soil if multi-concentration test; ensure homogeneity

Soil Spiked with Chemical(s) or Chemical Product(s)

Negative control soil	−	artificial soil or a clean field-collected soil for which previous tests with the test species have shown that all criteria for test validity could be regularly met; analyzed for at least the following: particle sizes (% sand, % silt, % clay), TOC (%), OM (%), pH, electrical conductivity, moisture content (%), WHC, and CEC
Characterization of chemical(s) or control chemical product(s)	−	information on concentration of active ingredients and impurities, water solubility, vapour pressure, stability, dissociation constants, adsorption coefficients, toxicity to humans and terrestrial organisms, and biodegradability of chemical(s) or chemical products(s) spiked into negative soil should be known beforehand
Solvent	−	deionized water is the preferred solvent; if an organic solvent is used, the test must include a solvent control soil in addition to a negative control soil
Preparation of mixtures	−	procedure dependent on the nature of the test substance(s) and the test design and objectives; chemical/soil mixtures may be prepared manually or by mechanical agitation; test substance(s) may be added as measured quantities in solution (i.e., in water or an organic solvent) or as a solid material comprised partly or completely of the test substance(s); ensure homogeneity
Concentration within soil mixture of chemical(s) or chemical product(s) added	−	normally measure at beginning and end of test, in high, medium, and low concentrations as a minimum

[*] **The information in this table is for summary purposes only. Definitive requirements and recommendations of this test method are contained in the main body of this document.**

accurate representation of the concentration-response curve, and therefore greater confidence in the test results and their interpretation. For any test that is intended to estimate the ECp in a definitive multi-concentration test, at least five concentrations (with the negative control soil in alternating compartments) must be used. More test concentrations (e.g. ≥ 7) may be used to provide for testing of a broader range of test concentrations and to increase the likelihood of achieving the statistical endpoint sought for this test (see Section 4.3.7). Another option for a multi-concentration test is to conduct a preliminary range-finding test for avoidance responses to a wide range of concentrations using 1 test unit per test concentration, followed by the replication of fewer test concentrations (i.e., since more information on the effect concentration/dilution range would be available) with three test units per concentration. This would also assist with the selection of test concentrations in certain instances of highly contaminated soil, or soil for which the toxicity (e.g., LC50) is unknown.

All test, negative control, reference, and positive control soils must be prepared as described in Section 4.1. Immediately following the mixing of a batch, an identical wet weight of negative control soil (natural or artificial; see Section 3.3) or reference soil (see Section 3.5), equivalent to a volume of ~350 mL for *E. andrei* or ~200 mL for *D. rubidus*,[77] must be transferred to every second test compartment (three/test unit) within each test unit included in an avoidance test. Thereafter, a weighed quantity (sample dependent; equivalent to a volume of ~200 or ~350 mL, depending on the test species used) of test soil from the same batch of mixed test material must be transferred to the other three test compartments within a test unit. Depending on the nature or purpose of the avoidance testing (e.g., a single-concentration test with five replicate test units per test soil or concentration; a single-concentration test with fewer replicate test units per test soil or concentration for the purpose of screening soils; or a multi-concentration test with two replicate test units per test soil or concentration), the same test material (i.e., test soil from the same batch or site) must be placed into each of three alternating compartments within one or more test units. The volume of soil in each section of the test unit must be the same. The soil added to each test compartment should be smoothed (but not compressed) using a spoon, to evenly distribute the soil therein.

Following the addition of clean soil (i.e., negative control soil or reference soil) and a test soil (i.e., test concentration, or contaminated or potentially contaminated soil from the same batch or site) to each of three alternating compartments within each test unit, each unit should be covered with a lid (Section 3.2.3) to minimize moisture loss. The test units should be held overnight at the test temperature (Section 4.3.2) to enable chemical equilibration of the test soils therein.

Ten test organisms (Section 2.3.8) are transferred to each test unit the next day (i.e., on Day 0). At that time, a number of worms in excess of those required for the test should be removed from a culture (Section 2.3) or acclimation (Section 2.4) vessel. Adult (fully clitellated) worms within the acceptable size range (i.e., wet wt of individual worms; 250–600 mg if *E. andrei*, and 50–200 mg if *D. rubidus*) should be selected from this vessel, removed by gloved hand or by using the blunt arm(s) of rounded forceps, and transferred briefly to a clean, shallow dish or tray where they are quickly rinsed in clean test water (i.e., deionized or distilled water). Worms chosen should be similar in size, and only those appearing healthy, similar in colouration, and active when removed from the bedding substrate should be selected. Thereafter, these worms are placed into a transfer container (e.g., a glass or aluminum tray measuring ~ 10 × 10 cm) lined with paper towel dampened with test water. A final observation should be made of the worms in this container to

[77] The wet weight of soil required to achieve a volume of ~350 mL for *E. andrei* or ~200 mL for *D. rubidus* depends on the moisture content, bulk density, and other characteristics of the soil, and will vary from sample to sample. Accordingly, the wet weight of each sample required to achieve this volume should be determined by transferring that amount of sample required to fill a preweighed (or tared) glass beaker or jar (i.e., a 500 mL jar for *E. andrei* or 250 mL jar for *D. rubidus*) to a mark scribed on its side (reflecting the appropriate volume for the given test species), after gently smoothing (not compressing) the surface of the soil at this mark. Thereafter, the wet weight of that quantity should be determined and recorded, and an identical wet weight transferred to each of three (alternate) compartments within a test unit.

confirm that their appearance is normal. Any atypical worms should be discarded. The group of worms transferred to each test unit should be randomly allocated with respect to test soil or concentration. Thereafter, individual worms of as similar size as possible should be carefully selected while confirming that they are within the acceptable size range, and then transferred one at a time to the central chamber of each test unit (devoid of soil substrate). After the first worm is placed into this chamber, it is observed until it has moved into a compartment containing soil. The second worm is then added, and observed until it has disappeared from the central chamber into a neighbouring compartment. This procedure is repeated sequentially until the full complement of 10 worms per test unit has been added.

The test compartment (and its contents) entered by each worm should be noted and recorded.[78] Any worm that does not enter a test compartment within 30 minutes should be removed and discarded, and replaced with another worm from the transfer container.[79] Once a group of 10 worms has been added to a test unit, and all worms have moved into clean or test soil within the compartments, the time must be recorded (t = 0 h) and the lid placed on the test unit. For tests with *D. rubidus*, the test units can be wrapped with Parafilm to prevent the worms from escaping (see Section 3.2.3). Any test units constructed of transparent or translucent Plexiglas™ (see Section 3.2.3) must either be wrapped with opaque sheeting (e.g., aluminum foil) or held in a darkened facility throughout the test period.[80] The test units should be positioned randomly within the test facility. The dates and times test and control soils are prepared and organisms are added to the test units must be recorded and reported.

A minimum of 10 worms, taken randomly from the group selected for use in the test, must be weighed individually prior to being placed in a test unit to estimate the variability in size of worms used in the test. These individual weights must be recorded, and the mean (± SD) weight calculated and reported (Section 7).

4.3.2 Test Conditions

- This is a 48-hour sublethal test for avoidance of test soils by adult earthworms, during which the test soils are not renewed and the worms in each test unit are left undisturbed to choose between clean soil (i.e., negative control soil or reference soil) and a single test soil (e.g., a field-collected test soil at 100% or lower concentration, or a single concentration of a chemical-spiked soil).

- Each test unit is comprised of a central chamber devoid of soil and six identical pie-shaped test compartments that are interconnected and enable movement of worms from compartment to compartment. A "false back" must be place into each compartment if testing with *D. rubidus*, to reduce the volume of soil used for this smaller species. Three of the test compartments in each test unit must contain clean soil from the same batch, and three must contain a single test soil from the same batch or site. The location of clean and test soil in the six compartments within a test unit must alternate between compartments (i.e., each neighbouring compartment contains an alternate treatment).

- For a single-concentration test to quantify the estimate of effect, at least five replicate test units must be used. Each test unit has three compartments containing clean soil from the same batch, and three compartments containing a single test soil from the same batch or field replicate (if applicable). Identical aliquots of

[78] Records of entry to test compartments provide useful information as to whether the worms enter randomly and are initially randomly distributed within the test compartments (see Section 4.3.7), or whether they show a preference for clean soil (i.e., negative control soil or reference soil) versus the test soil in alternate compartments within a test unit.

[79] Individual worms typically move from the central chamber of a test unit to a test compartment within three to five minutes (Stephenson *et al.*, 1998).

[80] Worms must be held under conditions of continuous darkness during this test, to prevent light from affecting their behavioural response. Use of test units constructed of stainless steel or opaque Plexiglas™ effectively provides darkened conditions; otherwise, testing in darkness or shrouding the test units in aluminum foil or other opaque wrapping material is required.

clean and test soil from the same two batches are placed into alternating compartments within all five test units. If the purpose of the single-concentration avoidance testing is to screen a large number of soils for potential positive effects, then fewer replicates (e.g., 1 test unit) may be used and would be more cost effective.

- For a multi-concentration test, at least five test concentrations must be used, and more (i.e., ≥ 7) are recommended. A minimum of two replicate test units must be used for each test concentration. Each test unit must have three compartments containing clean soil from the same batch, and a single concentration of test soil from the same batch. Identical aliquots of clean soil from the same batch must be placed into alternating compartments within each of these test units. The concentration of test soil in the three alternating compartments of a single test unit must be identical (from the same batch); however, concentrations of test soil differ from unit to unit.

- The test must be conducted at a daily mean temperature of 20 ± 2 °C. Additionally, the instantaneous temperature must always be 20 ± 3 °C.

- Test organisms are held in continuous darkness throughout the test period.

4.3.3 Criterion for a Valid Test

The avoidance test is designed to detect sublethal effects (ISO, 2008). Therefore, for the results of this biological test method to be considered valid, the percent survival of all earthworms in each test unit must be ≥ 90% at test end. Where more than one replicate test unit is used in a test, the mean percent survival of all earthworms per test unit must be ≥ 90% for each test soil or test concentration at test end.

4.3.4 Food and Feeding

No supplementary feeding is to be provided during the test.

4.3.5 Observations and Measurements During the Test

The biological endpoint for this test is the number of live worms residing in each test compartment at the end of the test (see Section 4.3.6). When adding test organisms to the central chamber of each test unit at the start of the test, the compartment entered by each worm should be observed and recorded (see Section 4.3.1). Observations are not possible once the test begins (t = 0). Care should be taken to not move or otherwise disturb the test units throughout the test period (or until after side partitions have been inserted at the end of the test to segregate worms within each compartment; see Section 4.3.6).[81]

Air temperature in the test facility (Section 3.1) must be measured daily (e.g., using a maximum/minimum thermometer) or continuously (e.g., using a continuous chart recorder).

The pH and moisture content of at least one replicate of each test soil (including the negative control soil and, if used, reference soil) must be measured and recorded at the beginning and end of the test. Additionally, it is recommended that electrical conductivity be measured at the beginning and end of the test in instances where the test soil is anticipated to have a high salt content. The initial measurements should be made using subsamples of each batch of test soil used to set up replicates of a particular treatment (see Section 4.1). The final (i.e., t = 48 h) measurements should be made using subsamples of the replicates of each treatment to which worms were exposed, following the end-of-test observations of worm distribution, survival, appearance, and behaviour (see Section 4.3.6). Measurements of soil pH and moisture content should be made according to the guidance in Section 4.2.5.

The test soils might be analyzed for concentrations of chemical(s) or chemical product(s) of concern. Guidance in Section 4.2.5 applies here. Sections 5.5 and 6.3 should be consulted for further advice.

---

[81] Banging, jarring, or other related activities (such as moving the test units during the test or upon its completion but before the insertion of side partitions) that disturb the worms might cause them to start moving from compartment to compartment, and might result in spurious findings (G.L. Stephenson, Aquaterra Environmental, personal communication, 2001).

4.3.6 Ending the Test

The test must be terminated after 48 hours of exposure. To terminate this soil toxicity test, the lid of each test unit is removed without any other movement of or disturbance to the test apparatus. Then, a side partition (see Section 3.2.3, including Figure 2) is quickly inserted adjacent to the side of each test compartment to confine the test organisms to each compartment. Immediately thereafter, the number of live and dead worms on the surface of the soil in each test compartment must be observed and recorded. The contents of each test compartment should then be gently removed with a spoon or spatula, placed into a sorting tray or onto a plastic sheet, and the number of live and dead worms counted and recorded.[82] Worms appearing to be dead should be touched gently on their anterior end with a glass rod or spatula; absence of any response is defined as death. Dead worms are discarded. Missing worms must be counted as dead. The appearance (e.g., normal or signs of discolouration or lesions) and behaviour (e.g., normally active or lethargic) of each surviving worm should be noted and recorded.

Immediately after this evaluation, subsamples of each test soil (including the negative control soil and, if included in the test, reference soil) should be taken for determinations of pH and moisture content (Section 4.2.5). Analyses for other chemical constituents (i.e., concentrations of contaminants) should also be made at this time using representative subsamples of each test soil (Section 4.2.5).

4.3.7 Test Endpoints and Calculations

For each test, the total number of surviving worms in the test soil (i.e., contaminated or potentially contaminated soil) and the clean soil (i.e., negative control soil or reference soil) within each test unit at the end of the test must be determined and recorded.

The three most common possibilities for a typical test design involve:

i) Soil from multiple sampling locations, in which responses at one or more test site sampling locations are compared with those at a reference site sampling location,[83] with other test sampling locations, or with the control soil (i.e., single-concentration test). Hypothesis testing is frequently used in the statistical assessment, and the common outcome is that a response at a sampling location is either "different" or "not different" from another sampling location (Section 5.6.1).

ii) Soil from multiple sampling locations or spiked soils, screened (i.e., compared with a reference or control soil in a single-concentration test) for positive avoidance responses in order to identify, and therefore prioritize, further assessment for toxicological effects (i.e., 56-day reproduction test).

iii) Multiple concentrations of a test soil, achieved by mixing a test soil with reference or control soil (Section 5.3), or by spiking a soil with various concentrations of a chemical or chemical product (Section 6.2). For a multi-concentration test, the 48-hour *EC50* for avoidance must be calculated and reported (data permitting).

In a scenario where there are multiple sampling locations, an understanding of the strengths of various study designs is critical for the successful application of statistical tests. The study objectives should be clearly defined before data are collected, with an appreciation both for the power (ability to detect an effect) of the test design and the ease of interpretation of the results. In general, it is advantageous to limit the number of comparisons made, and this is typically done by choosing a test design and statistical tests that compare test sampling locations with a reference sampling location. Further gains in power can be made if a gradient can be assumed (i.e., samples collected in sequential order away from the point source; see Section P.4 in EC, 2005a, and EC, 2012). In some cases, study objectives and test design might not

[82] Rarely, a worm is severed by the insertion of a side partition at the end of the test. If any severed segment of a worm is found in a test compartment, it should be counted and recorded only if that segment is from the anterior end of the worm.

[83] Throughout this document, reference site is used to describe an area in which there is clean soil uninfluenced by the contaminant under study (i.e., reference soil). A reference soil should be collected for these comparisons, as described in Section 5. However, in the absence of a reference soil, a negative control soil may be substituted.

have been given adequate attention before the collection of the data, and to compensate, investigators will perform a comparison among all possible sampling locations, maximizing the number of comparisons made. This is strongly discouraged, particularly when large numbers of sampling locations are involved, because undesirable effects on Type I and Type II error rates might occur; interpretation of results is often more difficult; and unwarranted focus might be given to particular comparisons after data have been collected.

The percent survival of all earthworms in each test unit at the end of the test must be calculated and reported. For tests where more than one replicate test unit is used in a test, the mean percent survival of all earthworms per test unit must be calculated and reported for each test soil or test concentration at the end of the test test.

For a single-concentration test, the mean (± SD) number of surviving worms recovered from the test soil and the clean soil in each of the replicate test units must be calculated and reported. These values should be compared statistically using an appropriate statistic for pairwise comparisons (see Section 5.6). Results showing a significantly lower mean number of surviving worms in the test soil, relative to those in the clean soil, indicate an avoidance response to the test soil (or a preference response to the clean soil).

For a multi-concentration test, the percent avoidance of surviving worms for each concentration must be calculated and reported. The presence and extent of an apparent avoidance response to each test concentration is determined based on the (lesser) number of worms in the test soil relative to a neutral (no avoidance, no preference) response. A neutral response is defined as the presence of an equal number of worms in the test soil and the clean soil (i.e., negative control soil or reference soil) at the end of the exposure period. Using this definition, the total number of worms determined to be in a particular test soil within a test unit is converted to a value indicative of percent avoidance as follows:[84]

$$\text{avoidance (\%)} = \frac{\text{no. in clean soil} - \text{no. in test soil}}{\text{total number of worms}} \times 100$$

where:

"no. in clean soil"	is the number of live worms found in all compartments containing clean soil, at the end of the test;
"no. in test soil"	is the number of live worms found in all compartments containing test soil, at the end of the test; and
"total number of worms"	is the total number of live worms found in all compartments, at the end of the test.

Data permitting, the *median effective concentration* (EC50; including 95% confidence limits) and, if desired, any other ECp (e.g., EC20 or EC25) causing avoidance must then be estimated and reported (see Section 6.4) based on the percent avoidance responses determined for each test concentration. Environment Canada (Section 4 of EC, 2005a) provides direction and advice for calculating the ECp endpoint, which should be followed, and Section 6.4.1 herein gives further guidance in this regard. An initial plot of the raw data (percent avoidance) against the logarithm of concentration is highly recommended, both for a visual representation of the data and to check for reasonable results by comparison with later statistical computations. Any major disparity

[84] Using this equation, and assuming that numbers are based on the distribution of each group of 10 worms within a single test unit, the avoidance response to each test soil is calculated as follows: (a) if 5 or more worms in test soil, (5 − 5) ÷ 10 × 100 = 0% avoidance; (b) if 4 worms in test soil, (6 − 4) ÷ 10 × 100 = 20% avoidance; (c) if 3 worms in test soil, (7 − 3) ÷ 10 × 100 = 40% avoidance; (d) if 2 worms in test soil, (8 − 2) ÷ 10 × 100 = 60% avoidance; (e) if 1 worm in test soil, (9 − 1) ÷ 10 × 100 = 80% avoidance; and (f) if 0 worms in test soil, (10 − 0) ÷ 10 × 100 = 100% avoidance. If the experimental design includes two test units per concentration (see Section 4.3.1), with 10 worms per unit (i.e., n = 20), the same equation applies when calculating percent avoidance for each concentration. For instance, if 10 or more worms were found in test soil within these two units, (10 − 10) ÷ 20 × 100 = 0% avoidance; if 9 worms were found in the test soil, (11 − 9) ÷ 20 × 100 = 10% avoidance; if 8 worms were found in the test soil, (12 − 8) ÷ 20 × 100 = 20% avoidance, etc.

between the approximate graphic ECp and the subsequent computer-derived ECp must be resolved. The graph would also show whether a logical relationship was obtained between log concentrations (or, in certain instances, concentration) and effect, a desirable feature of a valid test (EC, 2005a).

An investigator might wish to analyze the data showing numbers of worms entering each test compartment at the start of the test (see Sections 4.3.1 and 4.3.5), to test for randomness of this response. Chi-square analysis (EC, 2005a) is suitable for this purpose. A significant difference due to treatment (i.e., clean soil versus test soil) indicates an initial detection and response (i.e., avoidance or preference) of earthworms to the test soil. A significant difference among compartments suggests a lack of random movement of earthworms into the test compartments at the start of the test.

4.4 Tests with a Reference Toxicant

The routine use of a reference toxicant is used to assess, under standardized test conditions, the relative sensitivity of a portion of the population of adult earthworms within a particular culture (Section 2.3.9) or from a particular batch of acclimated worms (i.e., for earthworms transported from another laboratory for testing; Section 2.4.9) from which test organisms are selected for use in one or more definitive soil toxicity test (i.e., 56-day reproduction test or 48-hour avoidance test). Tests with a reference toxicant also serve to demonstrate the precision and reliability of data produced by the laboratory for that reference toxicant, under standardized test conditions, as well as the technical proficiency of the laboratory staff conducting the test (EC, 1995). Testing with a reference toxicant, conducted according to the procedures and conditions described herein, must be performed according to one of the following two regimes:

i) a multi-concentration reference toxicity test at least twice per year[85] using organisms taken from the population of earthworms that is being cultured for use in the definitive test(s) (Section 2.3); or

ii) a positive control concentration run concurrently with each definitive test using worms from the same batch as those used in the definitive test (Section 2.3.9 and Appendix H).

If the avoidance test is used for contaminated site soil screening or substance range-finding purposes only (versus quantifying the estimated effect in a definitive test), testing with a reference toxicant is not required.

A laboratory that chooses to monitor the sensitivity of its culture(s) to a reference toxicant in a multi-concentration reference toxicity test should conduct these tests at least once every six months. Reference toxicity tests may be run concurrently with a definitive soil toxicity test using organisms from the same culture (Section 2.3) or the same acclimated batch (Section 2.4) as those used in the definitive test, if the number of available test organisms allows.

Described herein are the procedures and conditions to be followed when performing multi-concentration reference toxicity tests in conjunction with a definitive soil toxicity test using *E. andrei* or *D. rubidus*. These procedures also apply to tests for assessing the acceptability and suitability of cultures to be used in soil toxicity tests. They should be applied to assess intralaboratory precision when a laboratory is inexperienced with the biological test methods defined in this document, and during initial test setup (see Sections 2.3.1 and 2.3.9).

For the first option of testing with a reference toxicant in conjunction with the 56-day reproduction test, a reference toxicity test must be conducted as a definitive, static multi-concentration test using a 56-day (or 63-day, if applicable) IC50 for inhibition of reproduction as the endpoint (see Section 4.2.7). The test conditions and procedures described herein for

[85] Environment and Climate Change Canada previously included monthly reference toxicity tests as the option for routine testing (EC, 2004b); however, due to the effort required for test organism production and the lack of relevance of the 14-day acute reference toxicity test (required in the first edition test method document) to the endpoints measured in both the reproduction test and the avoidance test, the requirements for testing with a reference toxicant have changed and are as described herein.

performing a reproduction test must be applied to each of these reference toxicity tests. Additional conditions and procedures described in Section 4.2 for performing a multi-concentration reproduction test with samples of test soil apply equally to each reference toxicity test.

For the first option of testing with a reference toxicant in conjunction with the avoidance test, it is highly recommended that the reference toxicity test be a definitive, static multi-concentration test using a 48-hour EC50 for avoidance as the endpoint (see Section 4.3.7). The test conditions and procedures described herein for performing a 48-hour avoidance test must be applied to each of these reference toxicity tests. Additional conditions and procedures described in Section 4.3 for performing a multi-concentration avoidance test with samples of test soil apply equally to each reference toxicity test. Alternatively, the results of a reproduction test using a reference toxicant (described in the previous paragraph) may be applied to satisfy the reference toxicity testing requirements for the avoidance test.[86] Procedures given in Section 6 for the preparation and testing of chemicals spiked in negative control soil also apply to all reference toxicity tests, and should be referred to for further information. Environment Canada's guidance document on using negative control sediment spiked with a reference toxicant (EC, 1995) provides useful information that is also applicable when performing reference toxicity tests with negative control soil spiked with a reference toxicant.

The multi-concentration reference toxicity test must be performed using the same test vessels or test units as those used for definitive tests (Section 3.2.2 and 3.2.3), with the same volume of soil (i.e., ~350 mL for *E. andrei* and ~200 mL for *D. rubidus*; Section 4.1) at optimal moisture content. The number of replicate test vessels/units per reference toxicant concentration and negative control soil must be as described for each test (i.e., ≥ 5 for the reproduction test, see Section 4.2.1; and ≥ 2 for the avoidance test, see Section 4.3.1). The number of earthworms per test vessel must be 4 for the reproduction test as described in Section 4.2.1. The number of earthworms per test unit must be 10 for the avoidance test as described in Section 4.3.1.

Procedures for starting and ending a multi-concentration reproduction reference toxicity test must be consistent with those described in Sections 4.2.1 and 4.2.6. Test conditions described in Section 4.2.2 must be applied. Test organisms must be fed as described in Section 4.2.4. Test observations and measurements given in Section 4.2.5 must be followed.

Procedures for starting and ending a 48-hour multi-concentration avoidance reference toxicity test must be consistent with those described in Sections 4.3.1 and 4.3.6. Test conditions described in Section 4.3.2 must be applied. Test organisms are not fed during the test, as described in Section 4.3.4. Test observations and measurements given in Section 4.3.5 must be followed.

The validity criteria for reference toxicity tests are the same as those described for definitive toxicity tests (see Section 4.2.3 for the reproduction test and Section 4.3.3 for the avoidance test). Results for a reference toxicity test should be expressed as mg reference chemical/kg soil dry weight.

Appropriate criteria for selecting the reference toxicant to be used in conjunction with a definitive reproduction test using earthworms include the following (EC, 1995):

- chemical readily available in pure form;
- stable (long) shelf life of chemical;
- can be interspersed evenly throughout clean substrate;
- good concentration-response curve for test organism;
- stable in aqueous solution and in soil;
- minimal hazard posed to user; and
- concentration easily analyzed with precision.

Any multi-concentration reference toxicity test requires a minimum of six treatments (i.e., negative control soil and five concentrations of reference

[86] This option is available for laboratories that are conducting avoidance tests infrequently, but regularly conducting the 56-day reproduction test and the reference toxicity testing associated with it.

toxicant). Reagent-grade boric acid is recommended for use as the reference toxicant when performing soil toxicity tests with earthworms, although other chemicals may be used if they prove suitable.[87] Each test concentration should be made up according to the guidance in Sections 4.1 and 6.2, using artificial soil (Section 3.3.2) as the substrate.

Routine reference toxicity tests (e.g., those performed twice per year) using boric acid (or another suitable reference chemical) spiked in negative control soil should consistently apply the same test conditions and procedures described herein. A series of test concentrations should be chosen based on preliminary tests, to enable calculation of the required endpoint (i.e., a 56-day IC50 for inhibition of reproduction or a 48-hour EC50; see Section 6.4).[88, 89]

Environment and Climate Change Canada's Biological Assessment and Standardization Section is introducing the use of positive control replicates, included with each definitive test, as an alternative to routine multi-concentration reference toxicity testing. As such, the second option for testing with a reference toxicant offered herein is to include replicates of a single concentration of a known *toxicant*, which elicits a consistent partial response, with each definitive reproduction and avoidance test to serve as a positive control. It is highly

---

[87] Aquaterra Environmental (1998) initially evaluated the performance of various chemicals as candidate reference toxicants for use in conjunction with acute lethality tests for measuring soil toxicity to *E. andrei*. Subsequent testing by Stantec and Aquaterra Environmental (2004) demonstrated the sensitivity of *E. andrei* to boric acid in 56-day tests for effects on their survival, reproduction, and growth, and showed similar findings for a number of 56-day tests with this chemical performed according to Section 4.2 herein (EC, 2010). Further investigations by ECCC (ECCC, 2020b) confirmed the value of boric acid as a suitable reference toxicant for use with *D. rubidus* as well.

[88] Results for a number of 56-day (or, in some instances, 63-day) reference toxicity tests with boric acid, performed by Stantec and Aquaterra Environmental (2004) according to the biological test method described herein in Section 4.2, demonstrated similar findings for *E. andrei*. In two tests initiated using adult earthworms from asynchronous or synchronous cultures, the 35-day LC50s for these worms were 2706 or 3207 mg boric acid/kg dry wt of artificial soil, respectively (Stantec and Aquaterrra Environmental, 2004). Data for number of live progeny generated during these and two additional tests performed with worms from asynchronous or synchronous cultures according to Section 4.2 yielded IC50s ranging from 270 to 568 mg boric acid/kg dry wt of artificial soil, and IC20s ranging from 163 to 425 mg/kg. Data for dry weights of surviving progeny generated during these four tests yielded IC50s ranging from 147 to 948 mg boric acid/kg dry wt, and IC20s ranging from 23 to 414 mg/kg. Results for side-by-side tests performed using worms from asynchronous or synchronous cultures showed that the differences among respective statistical endpoints were, in each instance, not large, and the 95% confidence limits overlapped (Stantec and Aquaterra Environmental, 2004). The following concentrations of boric acid were used by Stantec and Aquaterra Environmental (2004) to calculate both lethal and sublethal endpoints during reproduction tests with *E. andrei* for this reference toxicant: 0, 7, 14, 28, 56, 113, 225, 450, 900, 1800, and 3600 mg/kg soil dry wt. An expanded range (based on a logarithmic series of concentrations; see Appendix G) that includes one or two higher test concentrations is recommended for future tests intended to calculate both lethal and sublethal endpoints. For tests restricted to sublethal endpoints, the following concentrations of boric acid proved adequate when calculating IC50s and IC20s for number of live progeny for *E. andrei* at test end: 0, 10, 16, 30, 50, 100, 300, 560, and 1000 mg/kg soil dry wt (Stantec and Aquaterra Environmental, 2004).

Results for 56-day reproduction tests with boric acid performed by ECCC using *D. rubidus* and the test method for a multi-concentration reference toxicity test described herein yielded IC50s ranging from 199 to 390 mg/kg for the number of live progeny at test end. For reference toxicity tests measuring only the reproduction endpoint, the following concentrations of boric acid proved adequate when calculating IC50s for the number of live *D. rubidus* progeny at test end: 0, 50, 100, 175, 250, 325, 400, and 600 mg/kg soil dry wt (P. Boyd, Environment and Climate Change Canada, personal communication, 2021).

[89] A 48-hour test for avoidance by *E. andrei* of multiple concentrations of boric acid spiked in a field-collected reference soil (Alberta Black Chernozem soil) was performed by four laboratories to validate this test method. Each of these laboratories was able to achieve valid test results (see Section 4.3.3). The mean 48-hour EC50 for boric acid spiked in this reference soil was 874 mg H_3BO_3/kg soil dry wt, with values for individual laboratories ranging from 757 to 979 mg/kg. The interlaboratory CV for these EC50s was 11%, which is considered to be well within an acceptable level of precision between laboratories (EC, 2004a). In an ECCC investigation (2020b), a 48-hour test for avoidance by *D. rubidus* of five concentrations of boric acid spiked in artificial soil (75, 131, 230, 402, and 703 mg/kg soil dry wt) showed a clear dose-response effect. Survival was >90% in all concentrations and avoidance was 70% at the highest boric acid concentration (703 mg/kg soil dry wt).

recommended that the reference toxicity testing conducted is relevant to the definitive tests being carried out; however, for laboratories that frequently conduct reproduction tests, and infrequently conduct avoidance tests, reproduction results of the 56-day positive control concentrations may also be applied to satisfy the reference toxicity testing requirements for the 48-hour avoidance test. Positive controls are defined as an exposure of test organisms to conditions similar to a negative control (i.e., same number of replicates, number of organisms per replicate, vessels/units, test conditions, etc.) except exposed to a single concentration of a known toxicant. This option could be more feasible and practical for longer term sublethal- and life-cycle-type toxicity tests, such as the 56-day reproduction test with *E andrei* and *D. rubidus*, described in this test method document.

If chosen, the traditional multi-concentration reference toxicity test is required to be conducted twice per year. The alternative, however is to run a positive control concentration concurrently with every definitive test conducted. This approach could have several advantages: it is economical (reduced effort and resources); it reflects a response by organisms subsampled from the lot (group) used for testing; and it can measure the same endpoint(s) in the same matrix and duration as the definitive test, especially for longer, sublethal soil toxicity tests.

The choice of toxicant for the positive control concentration should be made using the same selection criteria as those used for a multi-concentration reference toxicity test and reagent-grade boric acid (H_3BO_3) is recommended herein. A single concentration known to elicit a consistent partial response must be used (as compared with traditional reference toxicity tests conducted using multiple concentrations to capture a range of effects, e.g., complete lack of reproduction to no effect on reproduction). The positive control replicates must be prepared using the same test vessels/units as those used for definitive tests (Sections 3.2.2 and 3.2.3), with the same volume of soil (Section 4.1) at optimal moisture content. The number of replicate test vessels or test units per positive control sample must be ≥ 5 for the reproduction test and ≥ 3 for the avoidance test. The number of worms per test vessel must be 4 for the reproduction test as described in Section 4.2.1 and 10 per test unit for the avoidance test as described in Section 4.3.1. The positive control concentration should be made up according to the guidance in Sections 4.1 and 6.2 using artificial soil (Section 3.3.2), and the procedures and conditions for testing must be consistent with those used in the definitive test, as described in Sections 4.2 and 4.3. For the positive control option, the required endpoint is the percent response. For the reproduction test, the mean response (i.e., number of progeny produced) in the positive control concentration is subtracted from the mean in the negative control, divided by the mean negative control response, and multiplied by 100 to provide a percent response (see Appendix H). For the avoidance test, the percent response is calculated as described in Section 4.3.7.

If selecting this option, the positive control response (i.e., target effect size) must be defined and include acceptability limits for each endpoint. Acceptability limits for the purposes of this method document are synonymous with *warning limits* and must be operationally defined at each laboratory with variability limits that are fit for purpose. For example (see Appendix H), a laboratory might define for its positive control that boric acid (e.g., 245 mg H_3BO_3/kg dry soil) must produce a 72% inhibition of progeny production (i.e., target effect size) that falls in between calculated warning limits (i.e., $\geq 60\%$ and $\leq 84\%$), with a coefficient of variation (CV) of response over time of $\leq 30\%$. Keeping in line with currently required multi-concentration reference toxicity test results, the results of an individual positive control test are not to be used to determine the acceptability of the corresponding test result (i.e., as test validity criteria), but rather can be used to monitor consistency over time (i.e., similar means among positive control tests) and precision over time (i.e., overlapping ranges among positive control tests). Identifying outliers in test organism response or extreme variability in response for individual tests must be used to trigger investigations into potential causes such as culture sensitivity, culture health, environmental/facility conditions, and technician performance. Data obtained from negative controls, positive controls, and culture health data should be monitored over time (i.e., by trend analysis) to proactively indicate changes in the organism

response.[90] Appendix H provides an example of how to choose a positive control concentration for the reproduction test and how to derive warning limits.

For both multi-concentration reference toxicity tests and positive controls, once sufficient data are available (EC, 1995, 2005a), all comparable endpoints (i.e., EC50s or IC50s for a particular reference toxicant derived from multi-concentration reference toxicity tests, or percent reduction of progeny production or percent avoidance relative to control for a single concentration of reference toxicant tested as positive controls) must be plotted successively on a *warning chart*. For multi-concentration reference toxicity tests, the warning chart should plot logarithm of concentration on the vertical axis against date of the test or test number on the horizontal axis. For positive control concentrations, the warning chart should plot the percent reduction in reproduction or percent avoidance on the vertical axis against the test date or test number on the horizontal axis (Appendix H). Each new data point for the reference toxicant should be examined to determine whether it falls within ± 2 SD of values obtained in previous comparable tests using the same reference toxicant and test procedure (EC, 2005b, 2013b, 2014a; ECCC, 2020a; Appendix H). A separate warning chart must be prepared and updated for each dissimilar procedure (e.g., differing test types, differing species of test organism, or differing reference toxicants) and endpoint. Each new data point for the reference toxicant should be compared with established limits of the chart; the reference toxicant result is acceptable if it falls within the warning limits.

For multi-concentration reference toxicity tests, the logarithm of concentration (including EC50 and IC50) must be used in all calculations of mean and standard deviation, and in all plotting procedures. This represents continued adherence to the assumption by which each EC50 or IC50 was estimated based on logarithms of concentrations. The warning chart can be constructed by plotting the mean and ± 2 SD as logarithms, or by converting them to arithmetic values and plotting them on a logarithmic scale of concentration. Different approaches to creating a warning chart (e.g., Levey-Jennings, moving average) are acceptable. For positive control concentrations, the warning chart can be constructed by plotting the mean and ± 2 SD for percent reduction in reproduction or percent avoidance relative to the control on an arithmetic scale.

The mean of the available endpoint values, together with the upper and lower warning limits (± 2 SD), should be recalculated with each successive endpoint for the reference toxicant until the statistics stabilize (EC, 1995, 2005b, 2013b, 2014a; ECCC, 2020a; Appendix H). Warning charts can be used to detect trends over time. Examples of trends that might be observed include an increasing or decreasing trend, several successive points on one side of the mean, changes that are observed at different times of the year, and successive data points outside the ± 2 SD warning limits. If a particular data point fell outside the warning limits, the sensitivity of the test organisms and the performance and precision of the test would be suspect. Since this might occur 5% of the time due to chance alone, an outlying data point would not necessarily indicate abnormal sensitivity of the culture or batch of earthworms from an outside source being held in the laboratory, nor unsatisfactory precision of toxicity data. Rather, it provides a warning that this might be the case. A thorough check of all culturing, holding/acclimation, and test conditions and procedures, as well as technical proficiency, is required at this time. Depending on the findings, it might be necessary to repeat the reference toxicity test or positive control concentration, establish a new culture, select worms from an alternate culture, or obtain a new batch of test organisms from an outside source before undertaking further soil toxicity tests.

Results that fall within the warning limits do not necessarily indicate that a laboratory is generating consistent results. A laboratory that produced extremely variable historical data for a reference toxicant would have wide warning limits; a new

[90] Performance charts can be maintained for data obtained from negative control soils in 56-day reproduction tests, which plot the number of progeny produced on the vertical axis against the test date or test number on the horizontal axis. Maintaining these charts are valuable for monitoring the control performance of test organisms over time.

datum point could be within the warning limits but still represent undesirable variation in test results. A coefficient of variation (CV) of no more than 30%, and preferably 20% or less, has been suggested as a reasonable limit by Environment Canada (EC, 1995, 2005a) for the mean of the available values of log(EC50) or log(IC50) (see preceding paragraph). For these biological test methods, the CV for mean historic data derived for reference toxicity tests or positive controls performed using boric acid should not exceed 30%.

If a reference toxicity test or positive control result fell outside the control limits (mean ± 3 SD), it would be highly probable that the test was unacceptable and should be repeated, with all aspects of the test being carefully scrutinized. If endpoints fell between the control and warning limits more than 5% of the time, a deterioration in precision would be indicated, and again the most recent test should be repeated with careful scrutiny of procedures, conditions, and calculations.

Concentrations of reference toxicant (including single concentrations used as a positive control) in all stock solutions can be measured chemically using appropriate methods (e.g., analytical methods involving AES with ICP scan for concentration of boron). Test concentrations of reference toxicant in soil are prepared by adding a measured quantity of the stock solution to negative control soil,[91] and mixing thoroughly.[92] Upon preparation of the mixtures of the reference toxicant in soil, aliquots should be taken from at least the negative control soil as well as the low, middle, and high concentrations, or from the single concentration used for a positive control.[93] Each aliquot should either be analyzed directly, or stored for future analysis (i.e., at the end of the test) if the reference toxicity test or positive control result(s) based on nominal concentrations was found to be outside the warning limits. If stored, sample aliquots must be held in the dark at 4 ± 2 °C. Stored aliquots requiring chemical measurement should be analyzed promptly upon completion of testing with a reference toxicant. The reference toxicity test or positive control result (i.e., 48-hour EC50, 56-day IC50, % reduction in response relative to the control, or % avoidance) should be calculated based on the measured concentrations if they are appreciably (i.e., ≥ 20%) different from nominal ones and if the accuracy of the chemical analyses is satisfactory.

If boric acid is used as a reference toxicant for a reference toxicity test or for a positive control, the following analytical method (OMEE, 1996) is an example of a chemical procedures that can be used to confirm the nominal concentrations. A 1–5-g subsample of soil spiked with boric acid is dried at 105 °C to constant weight. A 1-g aliquot is then extracted using a 0.01 *M* solution of $CaCl_2$ by boiling a slurry of soil in 50 mL of this extraction solution and then re-adjusting the final volume to 50 mL using more extraction solution. The 50 mL extract is then filtered through a #4 Whatman™ filter, and diluted to a final volume of 100 mL. A blank sample is prepared in a similar manner. The filtrate is analyzed for elemental boron using ICP/AES. The boric acid concentration in the soil is then calculated using the following equation:

---

[91] Section 6.2 Preparing Text Mixtures includes an example showing the amounts of test water and boric acid to be added to dry artificial soil to prepare treatments for a reference toxicity test with different concentrations of boric acid in artificial soil. The calculations in this example show the amount of water necessary to adjust the moisture content of the artificial soil to a fixed percentage (i.e., 70%) of the soil's water-holding capacity, while taking into account the volume of the stock solution of boric acid as part of the overall adjustment for soil moisture content.

[92] An accepted procedure is to add a precalculated volume of stock solution (using volumetric and/or graduated pipettes) to a glass Erlenmeyer flask, diluting to a graduated mark using deionized water, and then adding a measured volume to the soil. The flask is then rinsed three times with deionized water, and the rinsate is added to the soil. The mixture of soil and stock solution is then mixed thoroughly (for approximately three minutes) with a mechanical mixer (e.g., a hand-held mixer with revolving stainless-steel beaters) until the soil appears homogeneous in colour, texture, and moisture content. During the mixing process, the soil in the mixing bowl should also be stirred intermittently using a large stainless-steel spoon to facilitate homogenization.

[93] If the endpoint for each reference toxicity test is to be based on measured concentrations, it is recommended that one or more aliquots of the chemical-in-soil mixture representing each test concentration be collected and analyzed. If, however, the endpoint for each test is based on nominal concentrations, sampling and analysis of aliquots from at least the low, middle, and highest test concentrations is recommended.

$$\text{boric acid}\left(\frac{\text{mg}}{\text{kg dry wt}}\right) = \frac{\frac{\mu\text{g B}}{\text{mL}}(\text{measured}) \times \text{final volume (mL)} \times \frac{\text{MW}_{\text{boric acid}}}{\text{MW}_{\text{boron}}}}{1000 \times \text{weight of sample (mg dry wt)}} \times 10^6$$

The analytical limit of detection for boric acid in soil is reportedly 1 mg boric acid/kg soil dry wt in most instances (Stephenson, 2003b).

Section 5

Specific Procedures for Testing Field-collected Soil or Similar Particulate Material

This section provides specific instructions for preparing and testing samples of field-collected (site) soil or similar particulate material, in addition to the procedures discussed in Section 4.

Detailed guidance for the collection, handling, transport, storage, and preparation of field-collected soil is given in Environment Canada's *Guidance Document on the Sampling and Preparation of Contaminated Soil for Use in Biological Testing* (EC, 2012). General procedures are outlined therein for the preparation of collecting soil samples, including: developing study objectives; identifying the study area; collecting background data; conducting site surveys, soil surveys, and ecological land classifications; selecting sampling strategies and locations; determining the size and number of samples to collect; establishing proper *quality assurance* and *quality control* (QA/QC) procedures; considerations for environment, health, and safety; and developing sampling plans. Guidance is also provided for soil collection, including: selecting sampling devices; collecting soil samples by horizon or by depth; handling soil samples on-site; selecting sample containers; and transporting samples. Procedures for personnel receiving, preparing (i.e., drying, wetting, sieving, grinding, homogenizing, reconstituting, and characterizing), and storing soil samples for biological testing at the laboratory are also described in EC (2012). Additional procedures and considerations are included that are specific to the nature of the contaminants (i.e., soils contaminated with volatile or unstable contaminants), biological testing requirements, and study objectives. Specific guidance is provided for sampling, handling, transporting, storing, and preparing soil from boreal forest, taiga, and tundra ecozones, as well as organic and wetland soils. Environment Canada's soil collection guidance document (EC, 2012) should be consulted and the guidance therein followed (in addition to the guidance provided here) when collecting samples of field-collected soil and preparing them for toxicity tests with earthworms using either of the biological test methods described herein.

5.1 Sample Collection

Environment Canada (2012) provides substantial guidance on field-sampling design and appropriate techniques for sample collection. The guidance provided therein assumes that some data on the characterization of the chemical and soil properties of the land under investigation are already available. Field surveys of soil toxicity using biological tests with earthworms and/or other suitable, soil-associated test organisms (e.g., EC, 2005b, 2013b, 2014a; ECCC 2020a) are frequently part of more comprehensive land assessments and remediation (Stephenson *et al*., 2008; EC, 2012). Such assessments often include a *battery of toxicity tests* to evaluate the toxicity of soil using more than one test type and test species in conjunction with tests for bioaccumulation of contaminants, chemical analyses, biological surveys of epifaunal and/or infaunal organisms, and perhaps the compilation of geological and hydrographic data. This integrated approach can provide more accurate information on the risk associated with soil contamination in ecological risk assessments and contaminated land management (EC, 2012). Statistical correlation in these assessments can be improved and costs reduced if the samples are taken concurrently for these tests, analyses, and data acquisitions.

Samples of soil to be used in either of the two biological test methods described herein (Section 4) might be collected quarterly, semi-annually, or annually from a number of contaminated or potentially contaminated sites for monitoring and *compliance* purposes. Soil samples might also be collected on one or more occasions during field surveys of sites for spatial (i.e., horizontal or vertical) or temporal definition of soil quality. Increasingly, biological (toxicity) testing is being used at all levels (i.e., Tiers) of risk assessment. Depending on the specific objectives of the assessment and the conditions at a contaminated site, site-specific toxicity data can be used in a number of ways, including:

- to screen soil at a site to locate highly toxic or sublethally toxic areas;
- to identify site soil (determine concentration of contaminant in a site soil) that has a toxic impact;
- to evaluate contaminated soil for lethal or sublethal toxic effects;
- to identify soil characteristics that modify bioavailability;
- to derive (in part) site-specific standards and/or remedial objectives;
- to identify the efficacy of bioremediation technologies and/or site remediation; and
- for long-term monitoring of a remediated site (EC, 2012).

Further guidance on the application of biological testing in contaminated soil assessment is provided in EC (2012).

Environment Canada (2012) provides extensive guidance on defining study objectives and developing a study plan that incorporates biological testing into contaminated land assessments and management. A study plan provides specific guidance on the methods and strategies for sample collection and the procedures required to ensure that all *data quality objectives (DQOs)* are met. Information incorporated into a study plan includes: identification of DQOs; definition of the study area; collection of background data; selection and location of sampling; selection of sampling strategies; QA/QC; and considerations for environment, health, and safety. The sampling strategy (i.e., the process by which the type, location, and collection method of samples is determined) is driven primarily by the study objectives and secondarily by the site characteristics, and is discussed in detail in EC (2012).

The number of locations to be sampled at a study site and the number of replicate samples per location will be specific to each study. The number of samples to collect depends upon the study objectives, the data quality objectives, the desired level of certainty, and site-specific considerations. The number of sample replicates required further depends on the experimental design of biological tests and, in most cases, a compromise between logistical and budgetary constraints (e.g., time and cost) and statistical considerations. Various types of samples (i.e., *point*, *composite*, and *bulk*) might be collected depending on the study objectives.

The majority of samples collected for biological testing are *unconsolidated samples* in which particles become loosened and separated in the sampling process. *Consolidated samples* are those collected such that the soil particles and pore structure remain unaltered (i.e., *cores*). Guidance on the collection of consolidated samples for biological testing is provided in EC (2012); however, this biological test method document and the guidance provided herein apply primarily to the use of unconsolidated soil samples.

Specific procedures for the collection, handling, and preparation of soils contaminated with volatile or unstable compounds are described in EC (2012) and include modifications to procedures for sample collection, transport, storage, preparation, and contaminant analyses. All of the procedures described therein should be applied in order to minimize the loss of contaminants when sampling and handling soils in the field, transporting soils to the toxicity laboratory, and any further loss of these contaminants in the laboratory prior to testing (i.e., during sample storage, handling, or preparation). Environment Canada's soil sampling guidance document (2012) also addresses issues related to QA/QC.

For certain monitoring and regulatory purposes, multiple replicate samples of soil (i.e., field replicates or separate samples from different point or bulk samples taken at the same location) should be taken at each sampling location, including one or more reference location(s). These replicate samples[94] provide information about the variability of the toxicity/bioavailability of the contaminants at the

[94] Replicate samples are field-replicated samples of soil collected from the same sampling location to provide an estimate of the sampling error or to improve the precision

location and allow for statistical comparisons of soil toxicity among more than one location (EC, 2005a). Each of these "true replicate" samples of soil may be tested for its toxicity to earthworms as a single laboratory replicate (i.e., using only one test vessel/unit per replicate sample) or as multiple laboratory replicates (i.e., using more than one test vessel/unit per replicate sample; see Section 5.6.1). ECCC does not have a recommendation on the minimum number of replicate samples based on statistical considerations. Investigators can set the number of field replicates as part of the study plan and data quality objectives (EC, 2012). The usual laboratory practice is to prepare laboratory replicates, and guidance on the minimum number of replicates described herein is well-supported by power analysis. For certain other purposes (e.g., preliminary screening studies or extensive surveys of the spatial distribution of toxicity), the survey design might include only one replicate sample (i.e., field replicate) from each location, including reference and/or control soils. If the objective is to identify and therefore prioritize soils or sites that require further toxicological assessment, laboratory replicates might not be required; however, if the objective is to quantify effects in these single field replicates (see Section 4.2.7 and 4.3.7), they must be homogenized and split between a number of replicate test vessels/units (i.e., laboratory replicates), depending on the species being used, the type of test being performed, and the chosen effect size (see Sections 4.2.1, 4.3.1, and 5.6.2).[95] The latter approach precludes any determination of mean toxicity at a given sampling location, and completely prevents any conclusion on whether a sampling location is different from the control or reference, or from another location. It does, however, allow a statistical comparison of the toxicity of that particular sample with the reference or control, or with one or more samples from other locations. It is important to realize that any conclusion(s) about differences, which arise from testing single field samples lacking field replication, must not be extended to make any conclusion(s) about the sampling locations.

Regardless of the study objectives, one or more sites should be sampled for reference (presumably clean) soil during each field collection (see Section 3.5).[96] Sites for collecting reference soil should be sought where the geochemical properties of the soil are similar to soil characteristics encountered at the test sites. Some of the most critical soil physicochemical properties that should be matched between the reference and contaminated soils include: particle size distribution, total organic carbon content (%), organic matter content (%), pH, and electrical conductivity. In addition, other properties to match might include CEC, total inorganic carbon, *redox potential*, and water-holding capacity (EC, 2012).

---

of estimation. A single soil sample from a sampling location is treated as one replicate. Additional samples collected at the same sampling location are considered to be additional replicate samples and must be treated identically but stored in separate sample containers (i.e., not composited).

[95] More replicates may also be set up to meet specific study objectives, such as those defined for Phase I (i.e., site soil screening tests) in the recommended framework for toxicity assessments in support of the development of site-specific remediation objectives for petroleum hydrocarbons in soil (ECASG, 2006). This framework for toxicity assessment of contaminated lands is divided into two phases, the first of which includes site soil screening tests using undiluted soil samples representative of the study site. The purpose of the screening tests is to: 1) quickly determine if there is toxicity associated with short-term (acute) exposure of the test organisms to the site soil; and 2) if there is no *acute toxicity*, continue the test to assess for *chronic toxicity* associated with prolonged exposure to the site soil. An investigator, therefore, might choose to expand the test design for the single-concentration tests described in this test method document by setting up extra replicates to look for potential acute responses (i.e., adult mortality) early in the test. This approach serves only to judge the potential of an acute response, but is not suitable for defining remedial or cleanup objectives. Phase II of the proposed framework uses multi-concentration tests to determine the magnitude of the toxicity. As described in Section 4.1 of this test method document, a range-finding test can be useful, and is recommended in the framework, for determining the range of effect concentrations (i.e., narrow the range of concentrations to be used in a definitive sublethal test).

[96] Ideally, a reference soil is collected near the site(s) of concern. It possesses geochemical characteristics (e.g., texture, total organic carbon content, organic matter content, and pH) similar to those of the field-collected test soil(s), but without anthropogenic contaminants. It is not unusual for nearby reference sites to have some degree of contamination from anthropogenic chemicals. In some instances, reference soil might be toxic or otherwise unacceptable for use in a soil toxicity test because of naturally occurring physical, chemical, or biological properties.

Matching of total organic carbon content (%) or organic matter content (%) might not be warranted in cases where *pollution* (e.g., from or within sewage or industrial sludge) is responsible for the high organic carbon content of test soils. Preliminary surveys to assess the toxicity and geochemical properties of soil within the region(s) of concern and at neighbouring sites are useful for selecting appropriate sites at which to collect reference soil. Further guidance on obtaining reference soils for biological testing and procedures to be followed when a site-specific reference soil cannot be located is provided in EC (2012).

Samples of municipal or industrial sludge (e.g., sewage sludge, dewatered mine tailings, or biosolids from an industrial clarifier or settling pond) might be collected for the assessment of their toxic effect(s) on earthworms, and for geochemical and contaminant analyses. Other particulate wastes being considered for disposal to land might also be collected for toxicity and physicochemical evaluation. Environment Canada (2012) provides guidance on additional considerations unique to waste pile sampling.

A sampling plan is an important component of the study plan. The sampling plan is a written description of the detailed procedures to follow when collecting samples, handling and preparing samples on site (if required), packaging, labelling, storing (if necessary), and transporting samples. Prior to extracting soil samples, it is important to obtain a thorough field description of the soil to be sampled. In addition, soils should be described at a detailed site-specific level. In Canada, soils are classified using the Canadian System of Soil Classification (CSSC). Soils collected for biological testing should be classified to the subgroup level according to the CSSC, following the guidance provided in EC (2012). Appendix E in EC (2012) provides detailed information on the CSSC and the basic components of soil taxonomic identification.

Procedures used for sample collection (i.e., point, bulk, or composite) will depend on the study objectives and the nature of the soil or other particulate material being collected. Shovels, augers, or soil corers (preferably stainless steel) are frequently used for collecting soil samples. Shovels, scoops, or trowels are among the most commonly used tools in soil sampling when large volumes of soil are needed; however, care must be exercised to ensure that a representative and unbiased sample is collected (e.g., a constant depth or soil horizon must be removed). More precise sampling devices include soil corers, ring samplers, cutting frames, or soil cylinders, but they are less convenient for extracting large soil sample volumes. If soil samples are collected at a specific depth, an auger can be a more efficient and less labour-intensive tool for soil collection. Descriptions of the more commonly used soil collection devices and the procedures that should be followed for collecting soils are provided in EC (2012).

Most Canadian forest or non-agronomic, ecozone soils are highly stratified into soil horizons. The structure and chemistry of soil horizons are often very different, and this can result in different bioavailability and toxicity of contaminants to soil organisms. The top layer (A horizon) is the most commonly sampled horizon for biological testing. This horizon contains the most organic matter and most of the biological activity in mineral soils. Depending on the study objectives, the forest litter (*L layer*), fulvic/humic (*FH horizon*) (e.g., at a forested site), or surficial organic layer (O horizon) of mineral soils (e.g., at a tundra site) might also be collected when present. Subsurface B horizons and, less commonly, C horizons might also be sampled. Soils from the boreal or taiga ecozones sampled for the assessment of effect(s) on worms, described in this test method document, must be collected as separate soil horizons, where possible. Collection of soil samples according to depth is recommended for soils without distinct soil horizons (e.g., where the surface soil horizons have been mixed or disturbed due to human activity). To sample soil by horizon, the soil profile must first be classified, as described earlier and in EC (2012). Care should be taken when sampling soil horizons that dilution of the soil contamination does not occur. This is particularly important in cases where the vertical contamination extends only partially through a soil horizon. In this situation, the horizon may be sampled only to a certain depth, or collected as two different samples at two sampling depths (EC, 2012).

Guidance on the collection of soil samples for toxicity testing is provided in detail in EC (2012). The first step is to establish the boundaries of the

sample location. The surface of the location where each sample is to be collected should then be cleared of debris such as twigs, leaves, stones, thatch, and litter (unless the L layer is being collected as part of the study design). If the location is an area of grass or other herbaceous plant material, the plants should be cut to ground level and removed before the sample is collected. Removal of the vegetation should be done such that removal of soil particles with the roots is minimal. Dense root masses (e.g., grasses) should be removed and then shaken vigorously to remove soil particles adhering to the roots. The soil sample to be collected for toxicity evaluation and chemistry should be taken from one or more depths that represent the layer(s) of concern (e.g., a surficial layer of soil, or one or more deeper layers of soil or subsoil if there is concern about historical deposition of contaminants). Soils exhibiting distinct horizons (e.g., undisturbed forest soils) must be sequentially collected in separate horizons as a soil pit is excavated (EC, 2012).

The minimum volume or mass of soil required for testing depends upon the study objectives, site conditions, and the test to be conducted. For a given test, the amount of soil required varies and depends on the experimental design of the toxicity test (e.g., single-concentration test versus multi-concentration test), as well as the physical characteristics of the soil (e.g., bulk density, moisture content, amount of debris in the soil), the nature of the chemical analyses to be performed, and the distribution of the contaminants in the soil (e.g., vertical distribution). The required volume of soil per sample should be calculated before commencing a sampling program. This calculation should take into account the quantity of soil required to prepare laboratory replicates for soil toxicity tests, as well as that required for particle size characterization, total organic carbon content (%), organic matter content (%), moisture content (%), and specific chemical analyses. Soil collection volume recommendations for specific biological tests are provided in EC (2012). For the tests described herein, a volume of at least 5–7 L of soil per sample is normally required, although this will depend on the study objectives/design (e.g., single-concentration or multi-concentration test) and the nature of the chemical analyses to be performed, and possibly also on the nature of the soil (e.g., need for removal of excess water and/or debris in the laboratory, which can reduce the sample volume). To obtain the required sample volume, it is frequently necessary to combine subsamples retrieved using the sampling device. Guidance provided in EC (2012) for compositing subsamples in the field should be followed. The same collection procedure should be used at all field sites sampled. For samples collected as distinct soil horizons, each horizon must be placed into and stored in separate containers unless the soil profile has been disturbed through attempts to remediate the site.

The preparation of soil samples might begin in the field before the samples are shipped to a testing laboratory. This might include hand-sorting (to remove debris and/or organisms), air-drying, sieving, and homogenization of soil samples. All of these procedures are described in detail in EC (2012).

5.2 Sample Labelling, Transport, Storage, and Analyses

Containers for transport and storage of samples of field-collected soil or similar particulate material must be made of nontoxic, inert material. The choice of container for transporting and storing samples depends on the sample volume, the potential end uses of the sample, and the type and nature of the soil contamination. The containers must be clean and sealable and should be practical for handling and able to support the weight of the sample (EC, 2012). Thick (e.g., 0.1016 mm or 4 mil) plastic bags are routinely used for sample transport and storage. If plastic bags are used, it is recommended that each be placed into a second clean, opaque sample container (e.g., a cooler or a plastic pail with a lid) to prevent tearing and to support the weight of the sample and maintain darkened conditions during sample transport (ASTM, 2012). Plastic containers or liners should not be used if there is concern about the plastic affecting the characteristics of the soil (e.g., compounds from plastic leaching into the soil), the contaminants adsorbing to the plastic, or the contaminants causing the breakdown of the plastic. Containers recommended for the transport and storage of soils are listed in Appendix H of EC (2012).

Following sample addition, the air space in each container used for sample transport and storage

should be minimized (e.g., by collapsing and taping a filled or partially filled plastic bag). Immediately after filling, each sample container must be sealed, and labelled or coded. Labelling and accompanying records must include at least a code or description that identifies sample type (e.g., point, bulk, composite), sample date and time, sample site, precise location of sampling, sample condition, sample identification number (including replicate number, where applicable), and sample volume. The label information should also include the name and signature or initials of sampler(s). Persons collecting samples of soil should also keep records that describe details of:

- the nature, appearance, and volume of each sample;
- the sampling procedure and apparatus;
- any procedure used to composite or subsample bulk or point samples in the field;
- the number of replicate samples taken at each sampling station;
- the time of sampling;
- the types and numbers of containers used for transporting samples;
- any field measurements (e.g., temperature, pH, soil moisture content, bulk density) of the soil at the collection site;
- soil horizon characterization;
- any in-situ field testing (e.g., litterbag, earthworm exposure, bait lamina) performed;
- procedures and conditions for cooling and transporting the samples;
- observations of environmental conditions at the time of sampling (e.g., raining);
- observations and any field sampling of soil fauna and flora at the collection site;
- sample storage duration and conditions prior to arrival at the laboratory; and
- information on sample transportation.

Additional recommendations for site observations and field measurements are provided in Table 10 of EC (2012).

Soil samples should be kept cool during transport or storage and should not freeze or become overheated. As necessary, gel packs, regular ice, or other means of refrigeration should be used to assure that the temperature of the sample(s) remains cool (e.g., 7 ± 3 °C) during transit. It is recommended that samples be kept in darkness (i.e., held in light-tight, opaque transfer containers such as coolers or plastic pails with lids) during transport, especially if they might contain PAHs or other chemicals or chemical products that could be photo-activated or otherwise altered due to exposure to sunlight. All samples must be shipped with appropriate documentation, including chain-of-custody forms, as well as any specific regulatory documentation for transport of contaminated material (see EC [2012] for further guidance on sample transport).

The date the sample(s) is received at the laboratory must be recorded. Sample temperature and moisture content upon receipt at the laboratory must also be measured and recorded. In addition, each sample of field-collected test soil or each separately collected soil horizon should be inspected and the following qualitative descriptions made and recorded: colour; texture; informal description of moisture content; presence of standing water; presence of indigenous invertebrates, fungi, or plant material; and any strong odours (EC, 2012). Samples to be stored for future use must be held under conditions that maintain the characteristics and quality of the soil for its intended use (EC, 2012). If volatile contaminants are in the soil or are of particular concern, any air "headspace" in the storage container should be purged with an inert gas such as nitrogen before being capped tightly. Samples should not freeze or partially freeze during transport or storage (unless they are frozen when collected), and must not be allowed to dehydrate. If, however, one or more samples are saturated with excess water upon arrival at the laboratory (e.g., sampling occurred during a significant rainfall event), the sample(s) may be transferred to plastic sheeting for a brief period (e.g., one or more hours) to enable the excess water to drain or evaporate. Thereafter, the sample(s) should be returned to the transport container(s) or transferred to one or more airtight containers for storage. It is recommended that samples be stored in darkness at 4 ± 2 °C.[97] These

[97] Air-drying soil is another practical option for preserving natural soils and/or soils containing non-volatile or light-sensitive contaminants, as it allows a fast and more precise rehydration, and allows for the storage of samples at room temperature. Guidance on air-drying soils is provided in Section 3.10.3.1 of EC (2012).

storage conditions must be applied in instances where PAHs or other light-sensitive contaminants are present, or if the samples are known to contain unstable volatiles of concern.

It is recommended that samples of soil or similar particulate material be tested as soon as possible after collection. The soil toxicity test(s) should begin within two weeks of sampling, and preferably within one week. The test must begin within six weeks, unless it is known that the soil contaminants are aged and/or weathered and therefore considered stable. Further considerations for the storage of contaminated soil are provided in EC (2012), and the guidance therein should be followed.

In the laboratory, each sample of field-collected soil or distinct soil horizon should be thoroughly mixed (Section 5.3), and representative subsamples taken for physicochemical characterization. Each sample (including all samples of negative control soil and reference soil) must be characterized by analyzing subsamples for at least the following:

- particle size distribution (% sand, % silt, and % clay)
- total organic carbon content (%)[98]
- organic matter content (%)[98]
- pH
- electrical conductivity
- moisture content (%)
- water-holding capacity (WHC)
- cation exchange capacity (CEC)

Additionally, the following analyses should be performed:

- major cations and anions (Na^+, K^+, Mg^{2+}, Ca^{2+}, Al^{3+}, S^{2-}, Cl^-)
- nitrogen as total N, nitrate (NO_3^-), nitrite (NO_2^-), and ammonium (NH_4^+)
- phosphorus as total and/or bioavailable
- potassium as total and/or bioavailable
- C:N ratio

Other analyses could include:

- bulk density
- total inorganic carbon
- total volatile solids
- biochemical oxygen demand
- chemical oxygen demand
- redox potential
- soluble salts
- metal oxides (iron, manganese)
- sodium adsorption ratio
- contaminants and/or co-contaminants of concern
- characteristics of the contamination (e.g., odour, staining, debris, presence of fuel or solvent)

Unless indicated otherwise, identical chemical, physical, and toxicological analyses should be performed with subsamples representative of each replicate sample of field-collected soil or soil horizon (including reference soil) taken for a particular survey of soil quality, together with one or more subsamples of negative control soil.

5.3 Preparing Sample for Testing

Field-collected soil or similar particulate waste material must not be sieved with water, as this would remove contaminants present in the interstitial water or loosely sorbed to particulate material. Large gravel or stones, debris, indigenous macroinvertebrates, or plant material should normally be removed using forceps or a gloved hand. If a sample contains a large quantity of undesirable coarse debris (e.g., plant material, wood chips, glass, plastic, large gravel) or large macroinvertebrates, these may be removed by gently passing the soil through a coarse sieve (e.g., mesh size of 4 to 10 mm; EC, 2012). Dry sieving might also be desirable to ensure that the sample structure (i.e., aggregation, organic matter, or clay distribution) is amenable for testing with worms. Soils should not be sieved in the laboratory if they were sieved in the field, or if they have the crumbly texture that is optimal for testing (i.e., 3–5-mm clumps). Soil samples consisting of moist clayey subsurface soils are very cohesive and often cannot be directly sieved or homogenized. These soils should first be broken up manually and then dried prior to sieving and homogenization, as described in EC (2012). In

[98] Organic matter content can be used to calculate total organic carbon (TOC) by multiplying the organic matter (OM) content of a soil by a soil constant (AESA, 2001). However, the relationship between TOC and OM is slightly different among soils and the total organic carbon content should also be determined by laboratory analysis.

general, grinding of soil samples should be avoided when possible, but might be necessary with some soils (i.e., clayey soils) or if greater homogeneity of a sample is desired than can be achieved by sieving. As with soil sampling and storage procedures, any soil preparation procedures (i.e., *pretreatment*) should be documented and must be reported.

Reconstitution of soil sample constituents might be required prior to testing if the soil contained standing water that was decanted during preparation, or if portions of the sample were removed during preparation (e.g., thatch, plant root, or other organic material) but need testing along with the soil (EC, 2012). Soil horizons collected as separate components of a soil sample must be tested independently as separate soil samples. If the contaminants of concern have only been confirmed in one soil horizon (e.g., upper organic horizon) based on previous analyses and/or toxicity testing, then, depending on the study objectives, a decision must be made as to whether to conduct toxicity testing on this horizon alone or in the additional soil horizons collected from the sampling location.

Unless research or special study objectives dictate otherwise, each sample or horizon of field-collected unconsolidated test material should be homogenized in the laboratory before use (ASTM, 2012; ISO, 2012).[99] Any moisture that separates from a sample during its transport and/or storage must be remixed into it, if possible. Mixing can affect the concentration and bioavailability of contaminants in the soil, and sample homogenization might not be desirable for all purposes. To prepare a homogeneous sample, transfer the precalculated amounts of test and/or reference soil to a clean, rigid mixing container (e.g., a large stainless steel or plastic bowl) or, for larger volumes of soil, to clean plastic sheets spread out on a flat surface. The sample should be mixed manually (using a gloved hand or a nontoxic device such as a stainless steel spoon) or mechanically (e.g., using a domestic hand-held mixer with beaters at low speed, or a hand-held wire egg beater) until its texture and colour are homogeneous. A number of methods used to homogenize soil samples (e.g., folding, mixing, coning) are described in detail in EC (2012). While mixing, care should be taken to ensure that the impact of mixing on soil structure is minimal and that the structure is not destroyed entirely. As soon as the texture and colour of the sample appears to be homogeneous, mixing should be discontinued.

For each sample or soil horizon included in a test, mixing conditions including duration and temperature must be as similar as possible and reported. If there is concern about the effectiveness of sample mixing, subsamples of the soil should be taken after mixing, and analyzed separately to determine the homogeneity of particle sizes, chemical(s) of interest, etc.

As indicated in Section 3.6, one or more samples or horizons of field-collected test soil may either be tested at a single concentration only (typically 100%), or evaluated for toxicity in a multi-concentration test whereby a series of concentrations are prepared by mixing measured quantities with either negative control soil or reference soil. Guidance on concentration series that might prove suitable is found in Section 6.2, along with that for preparing test mixtures, which might apply equally when performing a multi-concentration test with one or more samples of field-collected soil. Refer to Section 4.1 for additional guidance when selecting test concentrations. In each instance, the test must include a treatment consisting solely of negative control soil (see Section 3.3).

As indicated in Section 4.1, for soils collected as distinct horizons, each horizon must be tested separately in independent definitive tests. For a multi-concentration test, the test soil horizon should be mixed with the same horizon of negative control or reference soil at the various test concentrations (e.g., 0%, 6.25%, 12.5%, 25%, etc.). In some cases, it might not be possible to collect the same horizons of negative control soil and test soil. For example, preliminary remedial action might have already been taken at the test site, resulting in disturbed or mixed natural soil horizons. In these scenarios, the test soil may be tested as a mixed soil where test concentrations are prepared by mixing suitable

[99] One of the reasons for routinely homogenizing samples is to mix into the soil any pore water that rises to the surface during sample shipment and storage. Homogenization is also necessary to redistribute the sample constituents that have compacted and layered according to particle size during transport and storage.

weights of test soil into the available horizon(s) of negative control soils at the appropriate test concentrations. The study objectives must take into account the soil profile of the reference soil and the location and/or mobility of the contaminants in the test soil. The goal is to match equivalent horizons in reference and contaminated soil, if possible.

Soil structure is an important factor influencing the survival and reproduction of earthworms, and moisture content plays an important role in the determination of soil structure. A qualitative procedure, informally known as a "squeeze test," can be useful when determining if the optimal moisture content of a sample of test soil has been achieved. Investigators might find it useful to apply this procedure when adjusting the moisture content of each sample of test soil to a particular percentage of the sample's water-holding capacity (see following paragraphs), in preparation for a toxicity test. To perform this test, a small, representative subsample of the test soil (e.g., a "pinch" of soil) is randomly taken using a gloved hand, and gently compressed between the thumb and forefinger. If a small quantity of water can be squeezed from the soil with gentle pressure, then the soil's moisture content is acceptable. If, however, no water appears, the soil is likely too dry. Conversely, if a substantial amount of water can be squeezed from the subsample of soil, it is likely too wet. As the reproduction test proceeds, test vessels should be weighed to determine water loss (see Section 4.2.5).

The moisture content of a given sample of field-collected test soil should be standardized during its preparation by determining its water-holding capacity (WHC) and then hydrating the soil to an optimal moisture content based on a percentage of this value. The optimal percentage of the WHC for each sample of field-collected soil must be determined prior to sample preparation and test initiation. To do so, the moisture content of each homogenized sample (i.e., each sample of test soil, including the negative control soil) must be determined (Sections 4.1, 4.2.1, and 4.3.1). Thereafter, the WHC of each sample must be determined using a recognized standard procedure (see following three paragraphs). A subsample of each soil sample should then be hydrated (or, if and as necessary, dehydrated) to a homogeneous, crumbly consistency with clumps approximately 3–5 mm in diameter.[100] The moisture content, WHC, and optimal percentage of the WHC of each soil horizon must be determined separately. Soil horizons with higher organic matter content can be expected to have a higher WHC than mineral horizons, so will require greater amounts of water to hydrate to a moist, crumbly texture. Based on the initial moisture content of the sample, the WHC of the sample, and the amount of water added to achieve the desired soil consistency, the sample's optimal moisture content can be calculated and expressed as a percentage of the WHC for each soil.[101] Once this target (or optimal) percentage of the WHC has been determined, the moisture content of each sample of test soil (including the negative control soil) can be standardized to the selected (sample-specific) moisture content. Test water (i.e., deionized or distilled water[102]) should be added to each sample with a moisture content that is less than

[100] An unpublished study, carried out by Environment Canada (J. Princz, Environment Canada, personal communication, 2004), determined the optimal moisture content for each of the diverse types of soil used while developing the biological test methods described herein (see Section 3.3 and Appendix F), based on a percentage of each sample's WHC. The optimal percentage of the WHC of these soils ranged from approximately 45–50% for the silt and sandy loam soils to 60% for the clay loam soil. These values were considered optimal since, at these levels of saturation, the soil mixed well, had an adequate moisture content according to the "squeeze test," and formed an acceptable structure (i.e., the resulting macro-aggregation of soil particles was conducive to healthy earthworms). Experience indicates that the actual moisture content of the test soils hydrated to optimal conditions can vary greatly (e.g., 20% for sandy loam soil to 50% for clay loam soil), depending on the bulk density and the WHC of the sample(s) of field-collected soil being tested (ESG and Aquaterra Environmental, 2002; Becker-van Slooten *et al*., 2003).

[101] For soils with high peat content (i.e., extremely high water-holding capacity), the method for determining the percent WHC described herein might be inaccurate and the results misleading. In such cases, the optimal moisture content may be estimated by eye (i.e., sample hydrated to a homogeneous, crumbly consistency with clumps approximately 3–5 mm in diameter) and the moisture content determined thereafter and reported as such (i.e., as moisture content instead of percent WHC).

[102] The use of purified water (i.e., deionized or reverse osmosis) to hydrate soils avoids the introduction of

the predetermined optimal percentage of its WHC, until this moisture content is achieved[103] (Aquaterra Environmental, 1998). If a sample is too wet, it should be spread as a thin layer on a clean sheet of plastic (e.g., a new plastic garbage bag or vapour-barrier plastic) or a clean, non-reactive (e.g., stainless steel or plastic) tray, and allowed to air-dry by evaporation at ambient (~20 °C) room temperature;[104] rehydration to the predetermined optimal percentage of its WHC might be necessary. Upon completion of adjustment of a sample's moisture content to the desired percentage of its WHC, the moisture content (%) of the hydrated soil must be determined and the percent WHC and percent moisture content recorded and reported.

The WHC (and the percent WHC that is optimal for biological testing) of a particular soil is generally unique to each soil type and/or horizon, and is ultimately the result of the interaction of many variables associated with soil structure (e.g., micro/macro-aggregation, pore space, bulk density, texture, organic matter content, and particle size distribution). There are a number of methods that can be used to determine WHC; however, most of these methods require measurements to be made on an intact soil sample (e.g., soil core) where characteristics (e.g., structural aggregations, pore space, bulk density, texture, and organic matter content) are preserved during collection. The USEPA (1989) has described an appropriate method for toxicity testing using unconsolidated materials (such as samples of field-collected soils that have been dried, sieved, and homogenized; or samples of soil formulated in the laboratory from constituents).[105] This method is outlined here.

For this method, ~130 g (wet wt)[106] of sample is placed in an aluminum pan or petri dish (15 × 1 cm), and dried at 105 °C until a constant weight is achieved (this usually takes a minimum of 24 hours). The soil is then cooled for a minimum of 20 minutes in a desiccator. Thereafter, 100 g of the oven-dried soil is placed into a 250-mL glass beaker with 100 mL of distilled or deionized water. The resulting slurry is mixed thoroughly with a glass stir rod. A folded filter paper (185-mm diameter Fisherbrand P8 coarse porosity, qualitative creped filter paper; catalogue no. 09-790-12G) is placed into a glass funnel (with a top inside diameter of 100 mm and a stem length of 95 mm). The folded filter paper should be level with the top of the glass funnel. Using a pipette, up to 9 mL of distilled or deionized water is slowly added to the filter paper to wet the entire surface. The funnel and hydrated filter paper are then weighed. To obtain the initial weight for the mass of the funnel plus hydrated filter paper plus dried soil (see "I" in the following Equation 1), the weight of the dried soil (100 g) is added to the weight of the funnel and the wet filter paper.

---

cations, anions, or trace metals into the soil (EC, 2012).

[103] An alternate approach sometimes used by certain investigators is to standardize (and adjust) the moisture content of each sample of field-collected soil to a fixed concentration, such as 35–45% of its dry weight (ASTM, 2012). However, a disadvantage of this approach is that certain samples of field-collected soil can appear to be very wet and have standing water on the surface after hydration to only 35–45% of their dry weight, whereas other site soils can appear considerably dryer after the same level of hydration (ASTM, 2012). Accordingly, the use of this alternate approach is not recommended here.

[104] If there is concern about volatilization of potential toxicants and/or changes in the nature of the toxicant of concern due to the drying process, alternative methods of drying the soil and/or the effects of drying the soil on the toxicity of the soil may be investigated.

[105] Some participants at a soil toxicity testing workshop sponsored by Environment Canada in Vancouver, BC (February 2003) considered the determination of WHC and a percentage of that capacity to be the most appropriate way of expressing soil moisture content (EC, 2004c). This led to a testing program to compare two different methods for estimating the WHC of soil (i.e., as per Annex C in ISO, 1999, or according to USEPA, 1989) as well as a somewhat different method for expressing soil moisture content, as a percentage of the soil's water-filled pore space (WFPS). The results of this investigation showed that each method had distinct advantages and disadvantages; however, the USEPA (1989) method for measuring WHC was recommended for use in EC's soil toxicity test methods when predicting an amount of water to be added to the soil and adjusting (if and as necessary) the moisture content of soil samples (Becker-van Slooten, *et al*., 2004).

[106] A larger amount of soil (i.e., for highly organic soils) might be necessary to obtain 100 g of soil (dry wt).

The funnel is then placed into a 500-mL Erlenmeyer flask and the soil slurry is slowly poured onto the hydrated filter paper held in the funnel.[107] Any soil remaining on the beaker and stir rod is rinsed into the funnel with the least amount of water necessary to ensure that all of the solid material has been washed onto the filter. The funnel is then tightly covered with aluminum foil and allowed to drain for three hours at room temperature. After three hours, the funnel containing the hydrated filter paper and wet soil is weighed. This weighing represents the final weight for the mass of the funnel plus hydrated filter paper plus (wet) soil (see "F" in the following Equation 1).

The water-holding capacity for the subsample of soil in the funnel, expressed as percentage of soil dry mass, is then calculated using the following equation:

$$\mathrm{WHC} = \frac{\mathrm{F} - \mathrm{I}}{\mathrm{D}} \times 100 \text{ (Equation 1)}$$

where:

WHC = water-holding capacity (%)
F = mass of funnel + hydrated filter paper + wet mass of soil
I = mass of funnel + hydrated filter paper + dry mass of soil
D = 100 g (i.e., dry mass of soil)

The WHC of each sample of test soil should be determined in triplicate, using three subsamples. The percentage of water (i.e., P_W) that is added to a sample of field-collected soil to achieve the desired hydration (i.e., the optimal percentage of the WHC) can be calculated as follows:[108]

---

[107] In very organic soils, where humic compounds' hydrophobicity delays water uptake, WHC can be underestimated unless the length of the soil saturation period is extended.

[108] The following example provides calculations that pertain to the hydration of samples of a contaminated field-collected soil and a negative control soil, when preparing a multi-concentration test, with concentrations ranging from 0 to 100% contamination, for use in an *E. andrei* reproduction test involving five replicates per treatment.

Assumptions:

Soil #1: Negative Control (nc) Soil

W_{nc}	= 2.3934 g	D_{nc}	= 1.9108 g
WHC_{nc}	= 80.30%	P_{WHCnc}	= 60.00%
MC_{nc}	= 25.26%	P_{Wnc}	= 22.92%

Soil #2: Contaminated (c) Soil

W_c	= 7.0575 g	D_c	= 5.6174 g
WHC_c	= 67.10%	P_{WHCc}	= 40.00%
MC_c	= 25.64%	P_{Wc}	= 1.2%

MC	= [(W − D) / D] × 100	[Equation 1]
P_W	= [WHC × (P_{WHC}/100)] − MC	[Equation 2]
V_W	= (P_W × M) / 100	[Equation 3]
M_W	= (M_D × W) / D	[Equation 4]

W = wet mass of substrate (g)
D = dry mass of substrate (g)
WHC = water-holding capacity (% of dry mass)
P_{WHC} = percentage of WHC desired (%)
MC = initial moisture content of substrate (%)
P_W = percentage of water to add to soil (%)
M_D = total mass of soil required for experiment (expressed as dry wt)
V_W = volume of water to add to soil (mL)
M_W = total mass of soil required for experiment (expressed as wet wt based on initial MC)

Calculations for the preparation of soil dilutions with 0, 2, 3, 6, 13, 25, 50, and 100 % contaminated soil combined with negative control soil:

For a reproduction test using this example, it is assumed that a total mass of 1025.00 g dry weight (wt) of soil is sufficient to satisfy the requirement for each treatment (i.e., 200.00 g dry wt per replicate × 5 replicates + 25.00 g dry wt extra soil for pH and electrical conductivity). To simplify the calculations, this example assumes that 200 g (dry wt) of either type of soil is sufficient to provide the 350-mL aliquot of soil to be added to each of five replicate test vessels per treatment, when performing a reproduction test using *E. andrei* (see Section 4.2.1). The following example provides calculations for a 2% contaminated concentration, but all other % contaminated concentrations would be performed the same way, and a summary of the values are presented at the end of this footnote.

= 1025.00 g dry wt × (2/100)
= 20.50 g dry wt of contaminated soil

And the remainder of the test soil required to prepare this treatment (i.e., 98 %) will consist of the negative control soil:

= 1025.00 g dry wt × (98/100)
[or 1025.00 g dry wt − 20.50 g dry wt]

$P_W = [WHC \times (P_{WHC}/100)] - MC$ (Equation 2)

where:

P_W = percentage of water to add to the soil (%)
WHC = water-holding capacity (%)
P_{WHC} = percentage of WHC desired (%)
MC = initial moisture content of the soil

The volume of water (i.e., V_W) that should be added to a sample of field-collected soil to achieve the desired hydration (i.e., the optimal percentage of the sample's water-holding capacity) can be calculated as follows (see footnote 108):

$V_W = (P_W \times M)/100$ (Equation 3)

where:

V_W = volume of water to add to the soil (mL)
P_W = percentage of water to add to the soil (%)
M = total mass of soil required for test (expressed as dry weight)[109]

= 1004.50 g dry wt of negative control soil

The wet weight of control and contaminated soil is calculated with Equation 4.
For control soil:
= (1004.50 g dry × 2.3934 g wet)/1.9108 g dry wt
= 1258.20 g wet weight

For contaminated soil:
= (20.50 g dry × 7.0575 g wet)/5.6174 g dry wt
= 25.76 g wet weight

Volume of water to add to soil is calculated using Equation 3. The volume of water needed for the control soil is combined with the volume of water needed for the contaminated soil. The equation thus becomes:
$V_{Tot} = V_{Wnc} + V_{Wc}$
$= ((P_{Wnc} \times M_{nc})/100) + ((P_{Wc} \times M_c)/100)$

V_{Tot} = ((22.92 × 1004.50)/100) + ((1.20 × 20.50)/100)
= 230.48 mL

The final total mass of soil required, based on wet weight, is 1514.43 g: 1258.20 g wet wt at the soil's initial moisture content (i.e., M_{Wnc}) for the negative control soil + 25.76 g wet wt at the soil's initial moisture content (i.e., M_{Wc}) for the contaminated soil + 230.48 mL of water.

The final moisture content for each soil would be 48.18% {[(1488.43 − 1004.50)/1004.50] × 100} for the negative control soil, and 26.83% {[(26.00 − 20.50)/20.50] × 100} for the contaminated soil.

The final moisture content of the negative control soil (i.e., 48.18% moisture) represents 60% of that soil's water-holding capacity (48.18 ÷ 80.30 = 0.60). The final moisture content of the contaminated soil (i.e., 26.83% moisture) represents 40% of that soil's water-holding capacity (26.83 ÷ 67.10 = 0.40).

Further values for the remaining test concentrations:

Conc	M_{Dnc} (dry wt)	M_{DC} (dry wt)	M_{Wnc} (wet wt)	M_{WC} (wet wt)
0% (control)	1025.00	0	1283.88	0
2% contam	1004.50	20.50	1258.20	25.76
3% contam	994.25	30.75	1245.36	38.63
6% contam	963.50	61.50	1206.85	77.27
13% contam	891.75	133.25	1116.97	167.41
25% contam	768.75	256.25	962.91	321.94
50% contam	512.50	512.50	641.94	643.89
100% contam	0	1025.00	0	1287.77

Conc	Vol Water (V_{Wnc} + V_{WC})	Total Mass
0% (control)	234.93	1518.81
2% contam	230.48	1514.43
3% contam	228.25	1512.25
6% contam	221.57	1505.68
13% contam	205.99	1490.37
25% contam	179.27	1464.12
50% contam	123.62	1409.44
100% contam	12.30	1300.07

[109] For tests with samples of field-collected soil, the amount of soil added to each test vessel is based on the wet weight of soil that is equivalent to a volume of ~350 mL for *E. andrei* and ~200 mL for *D. rubidus* (see Sections 4.2.1 and 4.3.1). When the optimal percentage of the soil's WHC is determined, the equivalent wet weight (of ~350 mL for *E. andrei* or ~200 mL for *D. rubidus*) should be determined, and the sample analyzed for dry mass. Then, the total mass required per replicate and test concentration can be determined, based on dry mass equivalent. The "M" (i.e., the total mass of soil required for the test) is expressed as dry weight in the formula used to calculate the volume of water to be added to a sample of field-collected soil to achieve the desired hydration (see Equation 3). To calculate the amount of soil required per test vessel on a dry-weight basis, a simple calculation is carried out. For example, assume that (for a given sample) the wet and dry weights of a subsample of soil, previously determined for the purpose of calculating the sample's water-holding capacity, are 4.1507 g and 2.7813 g, respectively. The dry weight equivalent to a 350-mL volume of this sample of soil (which has a wet weight of 270 g) can be calculated as follows:

(270 g × 2.7813 g) ÷ 4.1507 g = 181 g

This mass of soil can be rounded up to 200 g dry weight,

Environment Canada (2012) describes various procedures that can be used to manipulate soil samples to render them testable to meet study objectives or DQOs when the conditions do not occur within the sample as collected. Detailed procedures for soil manipulations are described and include: washing, aging/weathering, adjusting soil pH, conditioning, adjusting soil *fertility*, and reducing indigenous soil microorganisms (EC, 2012). In general, samples of field-collected soil must not be adjusted or manipulated, except for research-oriented toxicity tests intended to determine the influence of a particular soil manipulation on sample toxicity. Soil horizons with high organic levels (e.g., LFH horizons), however, might require at least one or more freeze/thaw cycles in order to remove indigenous invertebrates before testing (see Section 5.6.6 of EC, 2012).[110] Studies intending to investigate the effect of a soil manipulation (e.g., pH adjustment) on sample toxicity should conduct two side-by-side tests, whereby one or more sets of treatments are adjusted, and one or more duplicate sets of treatments are not. Detailed, proper documentation of any soil manipulation procedures carried out must be made and reported.
Immediately following sample hydration (or dehydration) and mixing, subsamples of test material required for the toxicity test and for physicochemical analyses must be removed and placed into labelled test vessels/units (see Sections 4.2.1 and 4.3.1), and into the labelled containers required for the storage of subsamples for subsequent physicochemical analyses. Any remaining portions of the homogenized sample that might be required for additional toxicity tests using earthworms or other test organisms (e.g., according to EC, 2005b, 2013b, 2014a; ECCC, 2020a) should also be transferred to labelled containers at this time. Subsamples to be stored for future toxicity testing should be held in sealed containers with minimal air space, in darkness at 4 ± 2 °C (Section 5.2) until tested. These storage conditions must be applied for subsamples collected for physicochemical analysis. Just before being analyzed or used in the toxicity test, each subsample must be brought to room temperature and thoroughly remixed to ensure that it is homogeneous.

5.4 Special Considerations for the Collection, Handling, and Preparation of Soil from Canada's Ecozones

Specific guidance on sampling, handling, transporting, storing, and preparing soil from various Canadian ecozones is provided in EC (2012).

Previously published Environment Canada soil toxicity test methods (EC, 2005b, 2014a) were developed for the assessment of soils with neutral to near-neutral soil pH and organic matter content ranging from approximately 3% to 12%. These soils are generally characteristic of the Ah horizons of agricultural soils in Canada and soils from deciduous mixed forest ecoregions in the southeastern part of the country (i.e., prairies and mixed-wood plains ecozones). There are many other soil types in Canada with widespread distributions that have properties falling outside the ranges considered typical by EC's previously published standard methods, and therefore require special procedures for sampling, handling, transport, storage, and preparation. These soils include: boreal forest soils, taiga soils, stony/shallow soils, organic soils, cryosolic soils, and wetland soils, and are relevant for use with the test methodologies described in this test method document. Given that these soils cover most of Canada's land mass and that anthropogenic activities in these regions (e.g., mining, forestry, oil and gas production) have created or have the potential to create contaminated lands, specific guidance on sampling, handling, transporting, storing, and preparing soils from these various ecozones is provided in EC (2012). Guidance is also provided on the variability of the soils within each of the described ecosystems and special considerations for selecting the appropriate test species when testing soils from these various ecosystems (EC, 2012).

---

thereby providing a little extra soil, if necessary. Therefore, for the example provided here, the mass of this sample of soil required for each replicate (expressed as dry wt) is 200 g. The total mass ("M") can then be calculated simply by multiplying the dry mass required for each replicate (in this instance, 200 g dry wt) by the number of replicates to be used in the test (i.e., for this example, five replicates).

[110] To initiate a freeze/thaw cycle, the soil sample is placed in the freezer (≤ -20 °C) for a minimum of three days. The soil is then removed from the freezer and allowed to thaw at ≥ 20 °C for a minimum of seven days. The cycle may then be repeated at least once more before testing is initiated (C. Fraser, Environment Canada, personal communication, 2013).

5.5 Test Observations and Measurements

A qualitative description of each field-collected test material should be made at the time the test is being set up. This might include observations of sample colour, texture, and homogeneity, and the presence of plants or macroinvertebrates. Any changes in the appearance of the test material observed during the test or upon its termination should be noted and reported. Photographs of the soils can also be used for maintaining records of soil appearance.

Sections 4.2.5 and 4.3.5 provide test-specific guidance and requirements for the observations and measurements to be made during or at the end of each test. These observations and measurements apply and must be made when performing either of the soil toxicity tests described herein using one or more samples of field-collected (site) soil.

Depending on the test objectives and experimental design, additional test vessels may be set up at the beginning of the reproduction test (Section 4.2.1) to monitor soil chemistry. These would be destructively sampled during and at the end of the test. Test organisms might or might not be added to these extra test vessels, depending on the study's objectives. Measurements of chemical concentrations in the soil within these vessels may be made by removing aliquots of the soil for the appropriate analyses (see Section 5.2).

5.6 Test Endpoints and Calculations

The common theme for interpreting the results of tests with one or more samples of field-collected test soil, is a comparison of the biological effects for the test (site) soil(s) with the effects found in a reference soil. The reference sample should be used for comparative purposes whenever possible or appropriate, because this provides a site-specific evaluation of toxicity (EC, 2005b, 2013b, 2014a; ECCC 2020a). Sometimes the reference soil might be unsuitable for comparison because of toxicity or atypical physicochemical characteristics. In such cases, it would be necessary to compare the test soils with the negative control soil. Results for the negative control soil will assist in distinguishing contaminant effects from non-contaminant effects caused by soil physicochemical properties such as particle size distribution, total organic carbon content (%), and organic matter content (%). Regardless of whether the reference soil or negative control soil is used for the statistical comparisons, the results from negative control soil must be used to judge the validity and acceptability of the test (see Sections 4.2.3 and 4.3.3).

The biological endpoints for the two test methods described herein are 28-day (or 35-day, if applicable) adult survival (a *quantal* measurement), if the data allows, and reproductive success (a *quantitative* measurement) for the 56-day (or 63-day, if applicable) reproduction test; and avoidance (a quantal measurement) at the end of the 48-hour avoidance test. Because of the different nature of the measurements involved, different statistical approaches are needed, and these approaches are further refined to reflect the objectives and design of the experiment. This section provides statistical guidance on data from single-concentration tests (i.e., soil samples from multiple sampling locations tested at full strength only). The simplest testing scenario involves the comparison of one test sampling location with one reference sampling location, whereas more complex designs might include a comparison of several sampling locations with a reference sampling location, or with each other. Only summary guidance is provided here for analyzing the mortality, reproduction, and avoidance endpoints as more extensive statistical guidance is available elsewhere (EC, 2005a). Standard statistical procedures are generally all that is needed for analyzing the results. Section 3 in EC (2005a) should be consulted for guidance when comparing the findings for single-concentration tests from multiple locations using parametric or nonparametric tests. As always, the advice of a statistician familiar with *toxicology* should be sought *a priori* for test design and analysis of test data.

Guidance in Section 6 (including that in Section 6.2 for performing range-finding tests, and that in Section 6.4 for calculating test endpoints) should be followed if a multi-concentration test is performed using one or more samples of field-collected soil diluted with negative control soil or clean reference soil. Section 9 in EC (2005a) should be consulted when comparing such point estimates of toxicity for multiple samples of field-collected soil.

5.6.1 Variations in Design and Analysis

Environment Canada (EC, 2005a) provides detailed statistical guidance on the analysis of quantal data in various test designs that examine multiple sampling locations. Choice of a specific statistical test depends on several considerations, including but not limited to:

- the type of comparison that is sought (e.g., complete series of pairwise comparisons between all sampling locations, or compare the response from each sampling location only with that of the reference site);
- if a chemical and/or biological response gradient is expected;[111] and
- the level and type (laboratory or field) of replication.

Environment Canada has also provided detailed statistical guidance on the analysis of quantitative measurements (EC, 2005a),[112] which can be readily applied to measurements of earthworm reproduction (i.e., number of surviving juveniles at the end of the test) in a multiple sampling location scenario. If test results at a single test sampling location are to be compared with test results at a reference sampling location, a *t*-test[113] is normally the appropriate statistical test (Section 3.2 in EC, 2005a). In situations where more than one test sampling location (treatment) is under study, and the investigator wishes to compare multiple sampling locations with the reference, or compare sampling locations with each other, a variety of ANOVA and multiple comparison tests (and nonparametric equivalents) exist (Section 3.3 in EC, 2005a). Choice of a specific test depends on the three conditions described above for quantal tests, in addition to assumptions of *normality* and *homoscedasticity* being met.

A very preliminary survey might have only one sample of test soil (i.e., contaminated or potentially contaminated site soil) and one sample of reference soil, without replication. Simple inspection of the results might provide guidance for designing more extensive studies. A preliminary evaluation might conceivably be conducted with samples from many stations, but without either field replicates or laboratory (within-sample) replicates. The objective might be to identify a reduced number of sampling stations deserving of more detailed and further study. In this case, opportunities for statistical analysis would be limited (EC, 2005a).

A more usual survey of soils would involve the collection of replicate samples from several places by the same procedures, and their comparison with replicate samples of a single reference soil and/or negative control soil. There are several pathways for analysis, depending on the type and quality of data. In these multi-location surveys, the type of replication would influence the interpretation of results (i.e., field replicates or laboratory replicates, or both). If both replicate samples (i.e., field replicates) and replicate vessels/units (i.e., laboratory replicates) have been tested, a statistician should be consulted for analysis options. If only laboratory replicates and no field replicates were tested, it is difficult to make statistically robust conclusions regarding differences between sampling stations (locations) within a site or between sites (see also Section 5.1). The laboratory replicates would only show any differences in the samples that were greater than the baseline variability in the within-laboratory procedures for setting up and running the test. Sample variability due to location would not really be assessed in the statistical analysis, except that it would contribute to any difference in test results associated with sampling location.

If it were desired to compare the test results for the

---

[111] In this case, the expected gradient is determined during the experimental design phase (a priori), not after the data has been collected. Section 3.3 in EC (2005a) provides guidance on cases where a gradient effect is expected. If necessary, a statistician should be consulted for further guidance on analyses of data where a gradient is expected.

[112] Sections 3.2 and 3.3 in EC (2005a) provide guidance on the analysis of quantitative measurements for a single location and quantitative measurements for multi-locations, respectively, and should be consulted for the analysis of reproduction data. Section 7.5 in EC (2005a) provides additional guidance on multiple-comparison tests for hypothesis testing, and should be consulted for additional detail; however, the calculation of NOEC/LOEC is not recommended herein.

[113] The *t*-test assumes equal variance between groups; however, modification of the *t*-test that can accommodate unequal variance is also available (EC, 2005a).

replicate samples from each sampling location with those for the reference soil, a number of tests are recommended, depending on whether the samples show a gradient and depending on whether there is an even or uneven number of replicates (see Section 3 in EC, 2005a).

In a multi-location survey, an investigator might wish to know which of the samples from various sampling locations showed results that differed statistically from the others, as well as knowing which ones were different from the reference and/or negative control sample(s). Such a situation might involve sampling from a number of locations at progressively greater distances from a point source of contamination, in which instance the investigator might want to know which sampling locations provided samples that had significantly higher toxicity than others, and thus which locations were particularly deserving of cleanup. Sections 3.1, 3.3, and 7.5 in EC (2005a) provide further details, alternate tests, and nonparametric options, and the guidance therein should be followed.

5.6.2 Power Analysis

An important factor to consider in the analysis of the results for toxicity tests with soil is the potential for declaring false positives (i.e., calling a clean site contaminated; Type I error) or false negatives (i.e., calling a contaminated site clean; Type II error). Scientists are usually cautious in choosing the level of significance for tolerating false positive results (Type I error), and usually set it at $p = 0.05$ or 0.01. Commonly, scientists following a specified test design will never consider the relationship between power, variability, and effect size, leaving the Type II error (ß) completely unspecified. There are several factors that influence statistical power, including:

- variability of replicate samples representing the same treatment;
- α (i.e., the probability of making a Type I error);
- effect size (i.e., the magnitude of the true effect for which you are testing); and
- *n* (i.e., the number of samples or replicates used in a test, and in some cases, the allocation of those replicates).[114]

Environment Canada's guidance document on statistical methods for environmental toxicity tests (EC, 2005a) provides further information and guidance on Types I and II errors.

In research-based science, power analysis is most useful as part of a preliminary test design (Hoenig and Heisey, 2001; Lenth, 2007; Newman, 2008). Here, a preliminary experiment is run to determine the approximate standard deviation (variability), and to troubleshoot the execution of the experiment in general. Other factors in power analysis, such as effect size and number of replicates, can then be considered along with the standard deviation so that the final test design is optimized (e.g., number of replicates needed to detect a certain effect size is determined).

In the development of standardized test methods, the purpose of employing power analysis remains the optimization of test design or at least estimating the power of the current test design.[115] However, instead

---

[114] In the current test method, power analysis was limited to a two-sample *t*-test. Ideally, power analysis would be well-aligned with the statistical tests used in the study, and a separate power analysis would be performed for a multi-sample test design. If the experimental design requires the comparison of test samples with the reference sample only (e.g., using Dunnett's test or Williams' test), optimal power for the final reproduction endpoint may be achieved by allocating a higher number of replicates in the reference treatment (Dunnett, 1955; Williams, 1972; OECD, 2006). As a general rule, the number of reference replicates (n_o) can be related to the number of test sampling locations (k) and the number of test replicates (n) using: $n_o = n\sqrt{k}$ for Dunnett's test (OECD, 2006). A modified version is recommended if Williams' test is used, where $\sqrt{k}$ is replaced with a range between $1.1\sqrt{k}$ and $1.4\sqrt{k}$ (Williams, 1972). If the investigator was interested in achieving the target effect size through allocation of replicates among treatments, extra replicates could be allocated to the reference samples to achieve appropriate power for a given target effect size. As an example using Dunnett's formula, consider an experiment with one reference sampling location and four test sampling locations, and five replicates for each location. To maximize power, the optimal number of replicate samples at the reference sampling location would be $n_o = n\sqrt{k} = 5\times\sqrt{4} = 10$ replicates.

[115] In 2010, the USEPA introduced a data analysis approach termed the test of significant toxicity approach (TST; USEPA, 2010). The TST is a hypothesis testing approach based on bioequivalence, which is extensively

of a single estimate for variability and effect size, there would typically be a much richer data set to consider. For example, test method experts could collect a number of estimates of variability across different laboratories and different contaminant scenarios (Thursby *et al*., 1997; Van der Hoeven, 1998; Denton *et al*., 2011, 2019). Standardized tests are often used in monitoring or regulatory programs, which might specify the expected effect size (e.g., 25%) to be detected (AE, 2007).

Data from performance testing at the Soil Toxicology Laboratory at ECCC were used to estimate power for detecting a reduction in the number of surviving progeny in the 56-day earthworm reproduction test. Power analysis was not performed for avoidance testing. For *E. andrei*, power analysis used data from the revised test design, with five replicates and four adults per replicate. Variability estimates were collected from 24 tests, which included artificial soil (n = 11) and three field soils (n = 4 or 5 per soil type; total n = 13). Coefficient of variation (CV) was reasonably constant over these tests,[116] and standard deviation was back-calculated from CV in order to perform the power analysis. For *D. rubidus*, power analysis used data from tests with five replicates and four adults per replicate (i.e., the current test design). Variability estimates were collected from 91 tests, which included artificial soil (n = 42) and eight field soils (n = 1 to 11 per soil type; total n = 49). CV was reasonably constant over tests where average number of juveniles was between 12 and 50 per replicate, and CV decreased in tests where average number of juveniles per replicate was >50. The CV was assumed to be constant,[117] and standard deviation was back-calculated from CV in order to perform power analysis. For both *E. andrei* and *D. rubidus*, variability estimates were only available for replicate test vessels (laboratory replicates) and not among replicate samples (field replicates). Effect sizes of 30% (*D. rudibus* only), 40%, and 50% reduction in number of surviving progeny were used. A one-sided *t*-test was used, with $\alpha = 0.05$. Equal variance was assumed, and was estimated using the pooled estimate of standard deviation in the control and test soil. All power analysis was performed in Power and Precision v4.1.0 and cross-checked with G*Power v3.1.9.7 (L. Van der Vliet and C. Martinko, Environment and Climate Change Canada, personal communication, 2021).

To determine the minimum number of replicates necessary for a single-concentration test, the investigator must decide on and state the target effect size before beginning a test. The target effect size must be reported. The fulfilment of this requirement is the responsibility of the investigator, and must be decided on a project-by-project basis. This is a different and new approach for ECCC, as in other test methods the target effect size has been undefined.

Power analysis was performed for both test species using the 50th (moderate) and 85th (high) percentiles of variability[118] to develop requirements and

---

used in pharmaceutical development and evaluation. It is included in the discussion here because power analysis and the TST share some similar goals (e.g., a priori statement of Type I and Type II error) and because of the similar context (application of standardized testing).

[116] For retrospective analysis, the CV from the previous *E. andrei* test design (n = 10, with two adults per replicate) was compared with that of the revised test design (n = 5, with four adults per replicate). The expected result was a decrease in variability in the revised test design; however, the variability remained unchanged. For the previous test design, the 50th percentile of CV was 62.3%. For the revised test design, the 50th percentile of CV was 62.1%. There were small differences at the 15th and 85th percentiles in comparing the previous and revised test design.

[117] When a laboratory has previous data for *D. rubidus* that indicate a high number of control juveniles (≥ 80) are expected in a test soil and the expected CV is notably lower than those used here (e.g., CV ≤ 25%), there might be a substantial gain in power. In this case, MDAU can be contacted to determine if fewer replicates may be used to prevent a test design that is overpowered; alternatively, a laboratory can perform their own power analysis for this purpose. One risk of using an overpowered test is that an effect of low magnitude (i.e., a small decrease in reproduction) might be considered statistically significant when it is not biologically significant.

[118] For *E. andrei*, the 50th percentile of CV was 62.1% and the 85th percentile of CV was 87%; for *D. rubidus*, the 50th percentile of CV was 34.4% and the 85th percentile of CV was 51.8%. The recommendations for using a higher number of replicates (calculated using the high variability value) than the required minimum (calculated using the moderate variability value) are presented to encourage the use of additional replicates, making it more likely for a

recommendations on number of replicates, respectively, for different effect sizes.[119] For *E. andrei*, the power analysis showed that in order to reliably detect a 40% effect (power ≥ 80%) given the conditions listed, a minimum of 21 replicates must be used; and to reliably detect a 50% effect, a minimum of 13 replicates must be used, and 24 replicates are recommended. The power analysis showed that a 30% effect size cannot be reliably detected using a reasonable number of replicates for *E. andrei*.[120] For *D. rubidus*, the power analysis showed that in order to reliably detect a 30% effect (power ≥ 80%) given the conditions listed, a minimum of 13 replicates must be used, and 28 replicates are recommended; to reliably detect a 40% effect, a minimum of 7 replicates must be used, and 15 replicates are recommended; and to reliably detect a 50% effect, a minimum of 5 replicates must be used, and 9 replicates are recommended. There would be significant cost savings to contaminated site risk assessors or site remedial managers if they first assessed a site with less expensive and more rapid earthworm screening tests to identify and prioritize the main contaminated areas for further definitive 56-day reproduction testing (e.g., for confirmation of remedial success and/or soil quality improvement). The 48-hour earthworm avoidance testing has been shown to be a very effective and relevant screening tool for identifying potentially sublethally toxic soils in order to narrow the number of site soils that would require further toxicological assessment (i.e., 56-day reproduction test) (Hind-Rinke *et al*., 2005; EC, 2012).

The above requirements and recommendations for number of replicates used for a 56-day reproduction test are species-specific because the two test species have different variability properties (i.e., *E. andrei* was observed to be approximately 2 times more variable than *D. rubidus*; see footnote 118), which impacts power. However, the test species used should not necessarily be chosen based on power. The choice between test species might be driven by site-specific information such as soil type, pH, and resident species. Known sensitivity or tolerance of a species to the contaminant(s) present at a site can also be used to select the test species. For example, the same soil sample could cause a 50% decrease in reproduction in *E. andrei* but have no substantial effect on *D. rubidus*, and this information would support the use of *E. andrei* in testing. It is important to recognize that statistical significance and biological significance might not always align.

---

laboratory to reach 80% power consistently. With repetition, intralaboratory variability can be determined for each test species used, and it might be appropriate to reduce the number of replicates used to the required minimum.

[119] For example, in a test where the average control reproduction was 50 juveniles, and the average treatment reproduction was 35 juveniles, the calculated effect size would be 30% (i.e, a 30% decrease in reproduction).

[120] The number of replicates necessary to reach the threshold for a recommendation on replicates (85th percentile of variability) was 41. This is neither reasonable, nor readily achievable. Investigators seeking to improve reliability for detecting a 30% effect with *E. andrei* can instead consider practices to minimize variability.

Section 6

Specific Procedures for Testing Chemical-spiked Soil

This section gives guidance and instructions for preparing and testing negative control soil spiked experimentally with chemical(s) or chemical product(s). These recommendations and instructions apply to both biological test methods described in Section 4. Guidance in EC (1995) on spiking negative control sediment with chemical(s) and conducting toxicity tests with chemical/sediment mixtures is also relevant here, for chemical-spiked soil. Further evaluation and standardization of procedures for preparing chemical-spiked soil provided herein (Section 6.2) might be required before soil toxicity tests with earthworms or other appropriate soil organisms are applied to evaluate specific chemical/soil mixtures for regulatory purposes.

The cause(s) of soil toxicity and the interactive toxic effects of chemical(s) or chemical product(s) in association with otherwise clean soil can be examined experimentally by spiking negative control soil (Section 3.3) with these substances. The spiking might be done with one or more chemicals or chemical products. Other options for toxicity tests with earthworms, performed using the procedures described herein, include the spiking of chemical(s) or chemical product(s) in reference soil (Section 3.5) or test soil (Section 3.6). Soil horizons collected separately must be treated as separate soil samples, as described in previous sections (4.1 and 5.3), and must be characterized and prepared (i.e., hydrated and spiked) separately prior to being tested (Section 6.2). Toxicity tests using soil spiked with a range of concentrations of test chemical(s) or chemical product(s) can be used to determine statistical endpoints based on threshold concentrations causing specific sublethal effects (see Sections 6.4.1 and 6.4.2).

In Section 6.2, procedures are described for preparing test mixtures of chemical-spiked soil. Section 6.3 describes procedures for making observations and measurements during and at the end of the toxicity test, and Section 6.4 (and Sections 4.2.7 and 4.3.7) provides procedures for estimating test endpoints for multi-concentration tests. These procedures also apply to the mixing of multiple concentrations of field-collected test soil (including particulate waste material such as sludge or other dredged material intended for disposal) in negative control soil or reference soil, and to performing multi-concentration tests and determining statistical endpoints for these mixtures (see Section 5, and especially 5.6). Multi-concentration tests with positive control soil (Section 3.4) or one or more reference toxicants spiked in negative control soil (Sections 4.4) are also performed using the procedures and statistical guidance described in this section. Additionally, the influence of the physicochemical characteristics of natural or artificial negative control soil on chemical toxicity can be determined with spiked-soil toxicity tests according to the procedures and statistical guidance described in this section.

6.1 Sample Properties, Labelling, and Storage

Information should be obtained on the properties of the chemical(s) or chemical product(s) to be spiked experimentally in the negative control soil.[121] Information should also be obtained for individual chemicals or chemical products (e.g., pesticides or other commercial formulations) on their concentration of major or "active" ingredients and impurities, water solubility, vapour pressure, chemical stability, dissociation constants, adsorption

[121] Some studies might require the spiking (mixing) of one or more concentrations of chemical(s), chemical product(s), or test soil (e.g., contaminated or potentially contaminated field-collected soil or waste sludge) in either negative control soil or reference soil. Other applications could include the spiking of chemical(s) or chemical product(s) in one or more samples of test soil. For such studies involving samples of contaminated soil or similar particulate material (e.g., domestic or industrial sludge), instructions on sample characterization given in Section 5.2 should be followed. Sample(s) of field-collected negative control soil, reference soil, contaminated soil, or particulate waste to be evaluated in spiked-soil toxicity tests should be collected, labelled, transported, stored, and analyzed according to instructions provided in Sections 5.1 and 5.2.

coefficients, toxicity to humans and terrestrial organisms, and biodegradability. Where aqueous solubility is in doubt or problematic, acceptable procedures previously used for preparing aqueous solutions of the chemical(s) should be obtained and reported. If an acceptable procedure for solubilizing the test chemical(s) in water is not available, preliminary testing for its solubility in test water or a non-aqueous solvent should be conducted and confirmed analytically. Other available information such as the structural formulae, nature and percentage of significant impurities, presence and amounts of additives, and n-octanol:water partition coefficient, should be obtained and recorded. Any pertinent Safety Data Sheets (SDSs) should be obtained and reviewed.

Chemical(s) to be tested should be at least reagent grade, unless a test on a formulated commercial product or technical-grade chemical(s) is required. Chemical containers must be sealed and coded or labelled upon receipt. Required information (chemical name, supplier, date received, person responsible for testing, etc.) should be indicated on the label and/or recorded on a separate datasheet dedicated to the sample, as appropriate. Storage conditions (e.g., temperature, protection from light) are frequently dictated by the nature of the chemical.

6.2 Preparing Test Mixtures

On the day preceding the start of the toxicity test (i.e., Day -1), the mixture(s) of chemical(s) or chemical product(s) spiked in negative control soil should be prepared, transferred to test vessels/units, and held overnight before adding the test organisms the next day (i.e., Day 0) (see Sections 4.1, 4.2.1 and 4.3.1). For some chemicals or chemical products (e.g., those that are very volatile, degrade easily, or might be metabolized), the addition of test organisms may be carried out immediately after preparation of the test soil. For other test substances (e.g., sparingly soluble substances) a prolonged period of contact with the soil (i.e., up to several weeks with periodic mixing) might be required before equilibration is reached. The dates of test soil preparation and test organism addition must be recorded and reported. Each batch of test soil representing a particular treatment (concentration) should be prepared in a quantity sufficient to enable all test replicates of that treatment (concentration) to be set up (see Sections 4.2.1, 4.3.1, and 5.6.2) along with any additional replicates or quantities required for either physicochemical analyses (Section 6.3) and/or the performance of other soil toxicity tests using earthworms or other soil organisms (e.g., those performed according to EC, 2005b, 2013b, 2014a; or ECCC, 2020a).

The use of artificial soil (Section 3.3.2) to prepare each test mixture is recommended since it offers a consistent, standardized approach for comparing results for other chemicals or chemical products tested similarly in the same laboratory or by others (e.g., according to USEPA, 1989; ISO, 2008, 2012; ASTM, 2012; or OECD, 2016). If used, the formulation for artificial soil provided in Section 3.3.2 should be followed. The quantity of artificial soil required for the test(s) should be prepared (based on the dry weight of the constituents; see Section 3.3.2) and hydrated to ~20% moisture content (which is ~28% of the soil's WHC), adjusted if and as necessary to a pH within the range of 6.0 to 7.5,[122] aged for a minimum three-day period, and stored until required (see Section 3.3.2). The final moisture content (including that due to the addition of a measured aliquot of a test chemical or chemical product dissolved in test water, with or without an organic solvent) of any chemical-spiked soil prepared using artificial soil should be ~70% of the water-holding capacity of the final mixture (Section 3.3.2), for each treatment (concentration), or that which produces the optimal soil texture for testing (i.e., a homogeneous crumbly consistency with

[122] If, however, the test chemical(s) or chemical product(s) are anticipated to modify soil pH and the intent of the study is to nullify this influence, the (aqueous) pH of each batch (concentration) should be adjusted to a standard value (e.g., pH 6.5). Studies for determining the extent to which an acidic or basic test substance modifies the toxicity of soil spiked with a range of concentrations of this substance, due to the influence of pH *per se*, should involve two side-by-side tests. One test adjusts the pH of each test concentration to a standard value (e.g., pH 6.5) using the required (differing, depending on concentration) quantity of calcium carbonate, and the other test uses an identical quantity of calcium carbonate for each treatment sufficient to attain the "standard" pH (e.g., pH 6.5) in the negative control treatment.

clumps ~3–5 mm in diameter; see Section 5.3).[123] The final moisture content of each mixture (treatment) included in a test should be as similar as possible.

Investigators may choose to use natural control soil (Section 3.3.1) rather than artificial control soil (Section 3.3.2) as the negative control soil to be spiked with chemical(s) or chemical product(s) and for the corresponding replicates of control soil to be included in the test. Procedures described herein for artificial soil apply equally if natural soil is used. An exception is that the final moisture content of each batch of chemical-spiked soil (including control batches) prepared using field-collected soil should be adjusted to the optimal percentage of its WHC (by hydrating or dehydrating the sample, as the case may be) using guidance in Section 5.3. For natural soils, the weight of soil in each test vessel or test compartment of an avoidance unit might also differ due to differences in bulk density of the various soils that might be used.

The procedure to be used for experimentally spiking soil is contingent on the study objectives and the nature of the test substance to be mixed with

---

[123] The following example provides calculations that show the volume of both water (deionized or distilled) and a stock solution of a reference toxicant (boric acid) to be added to a sample of artificial soil with an existing moisture content, to create a treatment with a moisture content that is 70% of the WHC for the artificial soil. The calculations take into account the volume of a stock solution of boric acid added when preparing the treatment, as part of the overall adjustment for soil moisture content. To simplify the calculations, this example assumes that 165 g (dry wt) of artificial soil (AS) is sufficient to provide the 200-mL aliquot of soil to be added to each test vessel when performing a single-concentration boric acid test with *D. rubidus* involving five replicate test vessels per treatment (see Section 4.3.1).

The equations shown in Section 5.3 for calculating WHC and adjusting soil moisture content to a certain percentage of this value apply equally here. For this example, assume that the following assumptions apply (see Section 5.3 for equations and associated definitions of these terms).

Assumptions:

Wet mass of artificial soil (AS) = 3.2486 g
Dry mass of AS = 2.6924 g
Moisture content (MC)
of AS = [(3.2486 – 2.6924)/ 2.6924] × 100
= 20.66% (initial moisture content)
Water-holding capacity (WHC) of AS = 72.10%
Desired percentage of WHC (P_{WHC}) = 70.00%
Dry mass of AS required for
test (M_D) = [165.00g per rep × 5 reps] + 25.00 g extra
= 850.00 g dry wt
Wet mass of AS required for
test (M_W) = (850.00 × 3.2486) / 2.6924
= 1026 g wet wt

Calculations to prepare a treatment comprised of 200 mg boric acid per kg artificial soil (dry wt):

The stock solution consists of 2.5 g of H_3BO_3 in 1 L of deionized water.

The amount of boric acid required on a dry-mass basis is:

H_3BO_3 = (0.2 g H_3BO_3 / 1000 g soil dry wt) × 850.00 g dry wt
= 0.17 g H_3BO_3

The amount of stock solution required, on a volume basis, is:

H_3BO_3 = 0.17 g H_3BO_3 / (2.5 g H_3BO_3 / 1000 mL of water)
= 68.00 mL stock solution

The percentage of water (P_W) required for addition to this treatment to achieve the desired percentage of WHC (70%) is:

P_W = [WHC × (P_{WHC} / 100)] – MC
= [72.10 × (70.00 / 100)] – 20.66
= 29.81%

The volume of water (V_W) required for addition to this treatment to achieve the desired percentage of WHC (70%) is:

V_W = (P_W × M_D) / 100
= (29.81 × 850.00 g dry wt) / 100
= 253.39 mL of water required

However, as part of this required volume, 68.00 mL of the stock solution is to be added for dosing; therefore, an additional volume of water of only 185.39 mL will be required (185.39 mL of water – 68.00 mL of stock solution).

Accordingly, the final total mass of soil required, based on wet weight, would be 1279.39 g [1026 g wet wt at the soil's initial moisture content (i.e., M_W) + 185.39 mL of water + 68.00 mL of stock solution], and the final moisture content of the soil, based on dry weight, would be 50.51 % {[(1279.39 – 850.00) / 850.00] × 100}.

The final moisture content of this test treatment (i.e., 50.51% moisture) represents 70% of the test soil's water-holding capacity (50.51 ÷ 72.10 = 0.70).

negative control soil or other soil. In many instances, a chemical/soil mixture is prepared by making up a stock solution of the test chemical(s) or chemical product(s) and then mixing one or more measured volumes into hydration water, which is then added to artificial or natural negative control soil (Section 3.3).[124] The preferred solvent for preparing stock solutions is test water (i.e., deionized or distilled water); use of a solvent other than 100% test water should be avoided unless it is absolutely necessary. For test chemical(s) or chemical product(s) that do not dissolve readily in test water, a suitable water-miscible organic solvent of low toxicity (e.g., acetone, methanol, or ethanol) may be used in small quantities to help disperse the test substance(s) in water (OECD, 2016). Surfactants should not be used.

If an organic solvent is used, the test must be conducted using a series of replicate test vessels or compartments of an avoidance unit containing only negative control soil (i.e., 100% artificial or natural clean soil containing no solvent and no test substance), as well as a series of replicate test vessels/compartments containing only *solvent control soil* (ASTM, 2012; ISO, 2012; USEPA, 2012; OECD, 2016). For this purpose, a batch of solvent control soil must be prepared containing the concentration of solubilizing agent that is present in the highest concentration of the test chemical(s) or chemical product(s) in soil. Solvent from the same batch used to make the stock solution of test substance(s) must be used. Solvents should be used sparingly, since they might contribute to the toxicity of the prepared test soil. The maximum concentration of solvent in the soil should be at a concentration that does not affect the avoidance response or reproduction of earthworms during the test. If this information is unknown, a preliminary solvent-only test, using various concentrations of solvent in negative control soil, should be conducted to determine the threshold-effect concentration of the particular solvent being considered for use in the definitive test.

For tests involving the preparation of concentrations of chemical spiked in artificial soil, in which the chemical is insoluble in water but soluble in an organic solvent, the quantity of test substance needed to prepare a required volume of a particular test concentration should be dissolved in a small volume of a suitable organic solvent (e.g., acetone). This chemical-in-solvent mixture should then be sprayed onto or mixed into a small portion of the final quantity of fine quartz sand that is required when preparing each test concentration comprised of a measured amount of a particular chemical-in-solvent mixture spiked in artificial soil (see Section 3.3.2). The solvent can then be removed by evaporation by placing the container under a fume hood for at least one hour, and until no residual odour of the solvent can be detected. Thereafter, the chemical-in-sand mixture (with solvent evaporated) can be mixed thoroughly with the remaining quantity of pre-moistened sand and other ingredients required to make up artificial soil (Section 3.3.2). An amount of test water necessary to achieve a final moisture content of approximately 70% of the maximum water-holding capacity for this artificial soil can then be added and mixed with the soil/sand/peat mixture. The chemical-spiked soil can then be added to the test vessels/units (OECD, 2016).

For tests involving the spiking of natural soil, in which the chemical is insoluble in water, the following procedure may be used (R. Kuperman, US Army Edgewood Chemical Biological Center, personal communication, 2004). The chemical is dissolved in a solvent (e.g., acetone) and pipetted onto a 2.5-cm thick layer of soil to establish each chemical concentration in soil, ensuring that the volume of solution added at any one time does not exceed 15% (v:m) of the dry mass soil. The same total chemical:solvent solution volume at different concentrations is added to every treatment, equalling the volume required to dissolve the chemical at the highest concentration tested. The solvent is allowed to volatilize (usually requires a minimum of 18 h) in a dark chemical fume hood to prevent photolysis.

---

[124] Adding the stock solution to the hydration water and then to the soil assists with homogenization and decreases the risk of having the contaminant bind to a very small area of soil. Often, the addition of the predetermined amounts of hydration water is done incrementally to ensure that the WHC is not exceeded. With this approach, it is preferable to add the test chemical or stock solution to a portion of hydration water that is mixed into the soil before fully hydrating the test soil.

Each amended soil sample is mixed until homogeneous (e.g., transferred into a fluorocarbon-coated high-density polyethylene container and mixed for 18 h on a three-dimensional rotary mixer). Other procedures for dissipation of solvent may be used depending on the nature of the test chemical and/or solvent.

The sample of solvent control soil to be included in the test must be prepared using the same procedure but without the addition of the test chemical. Additionally, the solvent control soil must contain a concentration of solvent that is as high as that in any of the concentrations of chemical-spiked soil included in a test.

If the test chemical to be spiked in artificial soil is insoluble in both water and any suitable (nontoxic) organic solvent, a mixture should be prepared comprised of 10-g finely ground industrial quartz sand and the quantity of the test chemical necessary to achieve the desired test concentration in the soil. This mixture should then be mixed thoroughly with the remaining constituents of the pre-moistened artificial soil. An amount of deionized water necessary to achieve a final moisture content of ~70% of the maximum water-holding capacity is then added and mixed in. The resulting mixture of chemical-spiked soil can then be added to the test vessels/units (OECD, 2016).

If the test chemical to be spiked in natural soil is insoluble in both water and any suitable (nontoxic) organic solvent, the test chemical can be added through dry-mixing. The following procedure may be used (Ritchie *et al.*, 2017; EC, 2014b). A mixture of the natural soil and the quantity of test chemical necessary to achieve the desired concentration in the soil is prepared. This mixture is initially combined using an electric mixer, and then mixed over the course of several hours (e.g., 16 h), using a mechanical stirrer or mixer (e.g., rotary mixer) until homogeneous. The spiked soil can be mixed with test water (e.g., up to 50% of its optimal moisture content), prior to chemical spiking. Each concentration can be dry-mixed independently. Alternatively, a mixture of the test chemical and a portion of clean soil can be prepared at the highest test concentration, in a sufficient volume to meet the requirements of a test through dilution of the spiked soil with clean soil, following the initial spiking and mixing event. These mixtures may be prepared several hours or days prior to test initiation to allow for chemical equilibration. The efficacy of the dry-mixing procedures should be evaluated through chemical analysis of aliquots of soil.

Concentrations of chemical(s) or chemical product(s) in soil are usually calculated, measured, and expressed as mg test substance/kg soil (or µg substance/g soil) on a dry-weight basis (OECD, 1984, 2016; ISO, 1993). The assessment endpoints (e.g., ICps) are similarly expressed on a dry-weight basis (Section 6.4).

Mixing conditions, including solution:soil ratio, mixing and holding time, and mixing and holding temperature, must be standardized for each treatment included in a test. Time for mixing a spiked soil should be adequate to ensure homogeneous distribution of the chemical, which could be for several minutes or as much as 24 hours. During mixing, the temperature should be kept low to minimize microbial activity and changes in the mixture's physicochemical characteristics. Analyses of subsamples of the mixture are advisable to determine the degree of mixing and homogeneity achieved.

For some studies, it might be necessary to prepare only one concentration of a particular mixture of negative control (or other) soil and chemical(s) or chemical product(s), or a mixture of only one concentration of contaminated soil or particulate waste in negative control or other soil. For instance, a single-concentration test might be conducted to determine whether a specific concentration of chemical or chemical product in clean soil is toxic to the test organisms. Such an application could be used for research or regulatory purposes (e.g., "limit test").

A multi-concentration test, using a range of concentrations of chemical added to negative control soil (or other soil) under standardized conditions, should be used to determine the desired endpoint(s) (i.e., LC50, EC50, ICp; see Section 6.4) for the chemical/soil mixtures. A multi-concentration test using negative control soil spiked with a specific particulate waste might also be appropriate. At least five test concentrations plus the control(s) must be prepared for each multi-concentration test performed

to estimate a 48-hour EC50 (and any other ECp) for avoidance (see Sections 4.3.1 and 4.3.2); and more (i.e., ≥7 plus controls) are recommended. For a 56-day reproduction test, at least seven test concentrations plus the appropriate control treatment(s) must be prepared for each multi-concentration test, and more (i.e., ≥ 10 plus controls) are recommended (see Sections 4.2.1 and 4.2.2). When selecting the test concentrations, an appropriate geometric dilution series may be used in which each successive concentration of chemical(s) or chemical product(s) in soil is at least 50% of the previous one (e.g., 10, 5, 2.5, 1.25, 0.63 mg/kg). Test concentrations may also be selected from other appropriate logarithmic dilution series (see Appendix G), or may be derived based on the findings of preliminary "range-finding" toxicity tests. The reader is referred to Section 4.1 for additional guidance when selecting test concentrations.

To select a suitable range of concentrations, a preliminary or range-finding test covering a broader range of test concentrations might prove worthwhile. The number of replicates per treatment (see Sections 4.2.1 and 4.3.1) could be reduced or eliminated altogether for range-finding tests and, depending on the expected or demonstrated (based on earlier studies with the same or similar test substance) variance among test vessels within a treatment, might also be reduced for nonregulatory screening *bioassays* or research studies.

Depending on the test objectives, it might be desirable to determine the effect of substrate characteristics (e.g., particle size or organic matter content) on the toxicity of chemical/soil mixtures. For instance, the influence of soil particle size on chemical toxicity could be measured by conducting concurrent multi-concentration tests with a series of mixtures comprised of the test chemical(s) or chemical product(s) mixed in differing fractions (i.e., segregated particle sizes) or types of natural or artificial negative control soil (Section 3.3). Similarly, the degree to which the total organic carbon content (%) or organic matter content (%) of soil can modify chemical toxicity could be examined by performing concurrent multi-concentration tests using different chemical/soil mixtures prepared with a series of organically-enriched negative control soils. Each fraction or formulation of natural or artificial negative control soil used to prepare these mixtures should be included as a separate control in the test.

Depending on the study objectives and design, certain soil toxicity tests using earthworms might be performed with samples of negative control soil or reference soil to which chemical(s) or chemical product(s) are applied to the soil surface, rather than mixing it with the soil. Surface applications can be applied in the field or the laboratory. Procedures for chemical application include the use of a calibrated track sprayer to achieve a uniform distribution of the chemical over a specific area. Concentration of chemical(s) or chemical product(s) in the soil can be determined based on the penetration depth, the surface area or swathe width, the nozzle size, the pressure, and the speed of coverage of the sprayer (G.L. Stephenson, Aquaterra Environmental, personal communication, 2001). The OECD (2016) provides some guidelines for applying test substances to the soil surface, in preparation for earthworm reproduction tests.

6.3 Test Observations and Measurements

A qualitative description of each mixture of chemical-spiked soil should be made when the test is being established. This might include observations on the colour, texture, and visual homogeneity of each mixture of chemical-spiked soil. Any change in appearance of the test mixture during the test, or upon its termination, should be recorded.

Sections 4.2.5 and 4.3.5 provide test-specific guidance on and requirements for the observations and measurements to be made at the beginning, during, and at the end of the test. These observations and measurements apply and must be made when performing either of the soil toxicity tests described herein using one or more samples of chemical-spiked soil. For soils collected as soil horizons, these measurements must be made in each soil horizon tested.

Depending on the test objectives and experimental design, additional test vessels might be set up on Day -1 of the reproduction test (see Sections 4.2.1) to monitor soil chemistry. These would be destructively sampled during (i.e., on Day 0 and, in

certain instances, other days as the test progresses) or at the end of the test. These monitoring vessels would be set up on Day 0 if the test is initiated (i.e., organisms added to the test vessels) immediately after the preparation of the test soil due to concern over the volatilization, degradation, or metabolism of contaminants or chemicals in test soils (see Section 6.1). Test organisms might or might not be added to these extra test vessels, depending on study objectives. Measurements of chemical concentrations in the soil within these test vessels could be made by removing aliquots of soil for the appropriate analyses, at the beginning of the test, as it progresses, and/or at its end, depending on the nature of the toxicant and the objectives of the test.

Measurements of the quality (including soil pH and moisture content) of each mixture of spiked soil being tested (including the negative control soil) must be made and recorded at the beginning and end of the test, as described in Sections 4.2.5 and 4.3.5. If analytical capabilities permit, it is recommended that the stock solution(s) be analyzed together with one or more subsamples of each spiked-soil mixture to determine the chemical concentrations, and to assess whether the soil has been spiked satisfactorily. These should be preserved, stored, and analyzed according to suitable, validated procedures.

Unless there is good reason to believe that the chemical measurements are not accurate, toxicity results for any test in which concentrations are measured for each spiked-soil mixture included in the test should be calculated and expressed in terms of these measured values. As a minimum, sample aliquots should be taken from the high, medium, and low test concentrations at the beginning and end of the test;[125] in which instance, endpoint values calculated (Section 6.4) would be based on nominal ones. Any such measurements of concentrations of the test chemical(s) or chemical product(s) should be compared, reported, and discussed in terms of their degree of difference from nominal strengths. If nominal concentrations are used to express toxicity results, this must be explicitly stated in the test-specific report (see Section 7.1.6).

6.4 Test Endpoints and Calculations

Multi-concentration tests with mixtures of spiked soil are characterized by test-specific statistical endpoints (see Sections 4.2.7 and 4.3.7). Guidance for calculating an LC50 (Sections 4.2.7) or EC50 (Section 4.3.7) is provided in the following Section 6.4.1, whereas that for calculating an ICp (based on data showing reproductive inhibition; see Section 4.2.7) is given in Section 6.4.2. Section 5.6 provides guidance on calculating and comparing endpoints for single-concentration tests using samples of field-collected soil. This guidance applies equally to single-concentration tests performed with mixtures of spiked soil. For further information on these or other appropriate parametric (or nonparametric) statistics to apply to the endpoint data, the investigator should consult the Environment Canada report on statistics for the determination of toxicity endpoints (EC, 2005a).

For any test that includes solvent control soil (see Section 6.2), the test results for earthworms held in that soil and in negative control soil must be examined to determine whether they independently meet the test validity criteria (see Sections 4.2.3 and 4.3.3). If either of these controls fails to meet the test validity criteria, the test results must be considered invalid. If both controls meet the test validity criteria, the results from the solvent control should be used in statistical analysis.[126] If, however, both controls meet the validity criteria but adult survival or reproduction in the solvent control differs

[125] Certain chemicals might be known to be stable under the defined test conditions, and unlikely to change their concentration over the test duration. In this instance, an investigator might choose to restrict their analyses to samples taken only at the beginning of the test.

[126] Evidence to date with aquatic test organisms (Hutchinson *et al*., 2006) has shown that solvents rarely exert a direct effect on the test organism. However, if there was an effect of the solvent on the test organism, these effects would almost always be additive with the test substance, and the use of the solvent control compensates for this (Green, 2014). In addition, there could be an interaction between the test substance and the solvent that modifies toxicity. It is difficult to definitively show that this interaction is absent or present, because the test substance is not evaluated in the absence of the solvent. For this reason, the solvent control is the appropriate choice for comparisons (OECD, 2006).

significantly from the results for the clean control soil, this might be indicative of a potential solvent interference that would then require additional evaluation to determine the impact on the interpretation of the study. The USEPA (2008) provides guidance on what might be included in such an evaluation: (1) assess the relevance of the solvent control response (i.e., percent change relative to the response in control soil); (2) the degree of statistical significance associated with the difference between the two controls (i.e., highly significant difference versus marginally significant difference); (3) assess the breadth of the interference; (4) assess any other potential cause for the interference observed in the solvent control; and (5) assess the impact of the potential solvent control interference on uncertainty in the risk estimate.

6.4.1 LC50 or EC50

When a multi-concentration test with spiked soil mixtures is conducted (Section 6.2), the quantal mortality data for a specific period of exposure must be used to calculate (data permitting) the appropriate *median lethal concentration (LC50)*, together with its 95% confidence limits. For a reproduction test with exposure to multiple concentrations of spiked soil, the 28-day LC50 for the adult (first generation) earthworms must be calculated and reported, data permitting[127] (see Section 4.2.7). To estimate an LC50, mortality data at the specified period of exposure are combined for all replicates at each concentration.

For a multi-concentration avoidance test using earthworms, the median effective concentration (48-hour EC50) must be calculated (together with its 95% confidence limits) at the end of the test, data permitting. This calculation is based on the percent avoidance responses for each test concentration (Section 4.3.7).

The guidance provided by Environment Canada (2005a) on choosing statistical test methods to be applied to quantal (e.g., LC50 or EC50) data should be consulted when choosing the statistical test to be applied to such data for toxicity tests using earthworms (Section 4 in EC, 2005a).

The optimization of the calculation of the LC50 or EC50 and its 95% confidence intervals is based on the number of partial effects observed (EC, 2005a). In brief, probit and/or logit regression is the preferred method if two partial effects are observed; the *Spearman-Kärber* method is preferred if only one partial effect is observed; and the binomial method is used if no partial effects are observed, and as a general "default" method (EC, 2005a).

Regardless of the calculations used, it is highly recommended that any computer-derived LC50 or EC50 be checked by examining a plot, on logarithmic-probability scales, of percent mortalities or avoidance response at a defined period of exposure for the various test concentrations (EC, 2005a). Any major disparity between the estimated LC50 or EC50 derived from this plot and the computer-derived LC50 or EC50 must be resolved. A hand-plotted graph is recommended for this check (EC, 2005a).

6.4.2 ICp

For a multi-concentration reproduction test with exposure of earthworms to spiked-soil mixtures (or field-collected mixtures; see Section 5.3), the quantitative data representing reproductive inhibition must be used to calculate the ICp (see Section 4.2.7 and Section 6.2). The 56-day (or 63-day, if applicable) ICp is a quantitative estimate of the concentration causing a fixed percent reduction in the mean number of juveniles produced by the adult worms during the test.

The ICp is calculated as a specified percent reduction (e.g., the IC25 and/or IC20, which represent 25% and 20% inhibition, respectively). The desired value of "p" is selected by the investigator, and 25% or 20% is currently favoured. Any ICp that is calculated and reported must include the 95% confidence limits.

In the analyses of reproductive performance, the number of surviving juveniles produced in each

[127] Depending on the study objectives and the associated experimental design, a 56-day test for effects on reproduction of earthworms (*E. andrei* or *D. rubidus*) might be solely focused on sublethal effects. In this instance, the test might not include a sufficient number of high (lethal) concentrations to enable calculation of the 28-day LC50.

replicate is used to calculate the average number of surviving juveniles produced per treatment (concentration) in relation to the average number produced in the negative control replicates. A value of zero is assigned for a number of juveniles in a replicate, if the adult earthworms in that replicate died before producing progeny. If any of the adult worms died during the test, after producing young, the number of juveniles produced is still to be used in the analyses. If there are no surviving juveniles in a replicate (test vessel), it contributes a value of zero to the calculation used to obtain the average number of survivors for that treatment (concentration). If there are no surviving juveniles in all replicates at a given concentration, that concentration is still included in the analysis, using an average value of zero juveniles.

As indicated in Section 4.2.7, an ICp for mean number of surviving progeny produced in each treatment must be calculated and reported (data permitting) upon completion of a multi-concentration reproduction test with *E. andrei* or *D. rubidus*. These calculations must be made using the appropriate linear or nonlinear regression analyses (see the following Section 6.4.2.1). If, however, regression analyses fail to provide a meaningful ICp for the mean number of live progeny produced, the ICPIN analyses described in Section 6.4.2.2 should be applied to the corresponding data. Any procedures applied to the data, details regarding any transformation of the data, and the statistical method used for the calculation of ICp must be reported.

6.4.2.1 Use of regression analysis

Upon completion of a definitive 56-day (or, in certain instances, 63-day) multi-concentration reproduction test with *E. andrei* or *D. rubidus*, an ICp (including its respective 95% confidence limits) for the mean number of surviving progeny produced in each treatment must be calculated using regression analysis, provided that the assumptions below are met. A number of models are available to assess reproduction data (using quantitative statistical tests) via regression analysis. The proposed models for application consist of one linear model, and the following four nonlinear regression models: exponential, Gompertz, logistic, and logistic adjusted to accommodate hormesis[128] (see Section 6.5.8 in EC, 2005a). Use of regression techniques requires that the data meet assumptions of normality and homoscedasticity. The reader is strongly advised to consult EC (2005a) for additional guidance on the general application of linear and nonlinear regression for the analysis of quantitative toxicity data.[129]

The general process for the statistical analysis and selection of the most appropriate regression model (linear or nonlinear) for quantitative toxicity data is outlined in Figure 4. The selection process begins with an examination of a scatter plot or line graph of the test data to determine the shape of the concentration-response curve. The shape of the curve is then compared to available models so that one or more appropriate model(s) that best suits the data are selected for further examination (refer to Figure O.1, Appendix O, in EC, 2005a for an example of five potential models).

Once the appropriate model(s) is (are) selected for further consideration, assumptions of normality and homoscedasticity of the *residuals* are assessed.

---

[128] A hormetic response (i.e., hormesis) might be observed at one or more of the lowest, sublethal concentration(s), i.e., performance at such concentration(s) is enhanced relative to that in the negative control (see Section 10.3 in EC, 2005a). For instance, there might be more progeny produced in soil with low concentrations than in the control treatments. This is not a flaw in the testing. Rather, it is a real biological phenomenon. To calculate the ICp when this phenomenon occurs, the data should be analyzed using the hormesis model. The hormetic effects are included in the regression, but do not bias the estimate of the ICp. An estimated IC25 would still represent a 25% reduction in performance from that of the control.

[129] Some of the specific guidance provided in EC (2005a) refers to the use of a general purpose statistical package (i.e., SYSTAT); however, CETIS (a software package designed for environmental toxicology) contains the models described herein for regression analysis. The latest version of SYSTAT is available for purchase by contacting SYSTAT Software, Inc.; see website https://systatsoftware.com/systat/. The latest version of CETIS is available for purchase by contacting Tidepool Scientific Software; see website https://www.tidepool-scientific.com/Cetis/Cetis.html.

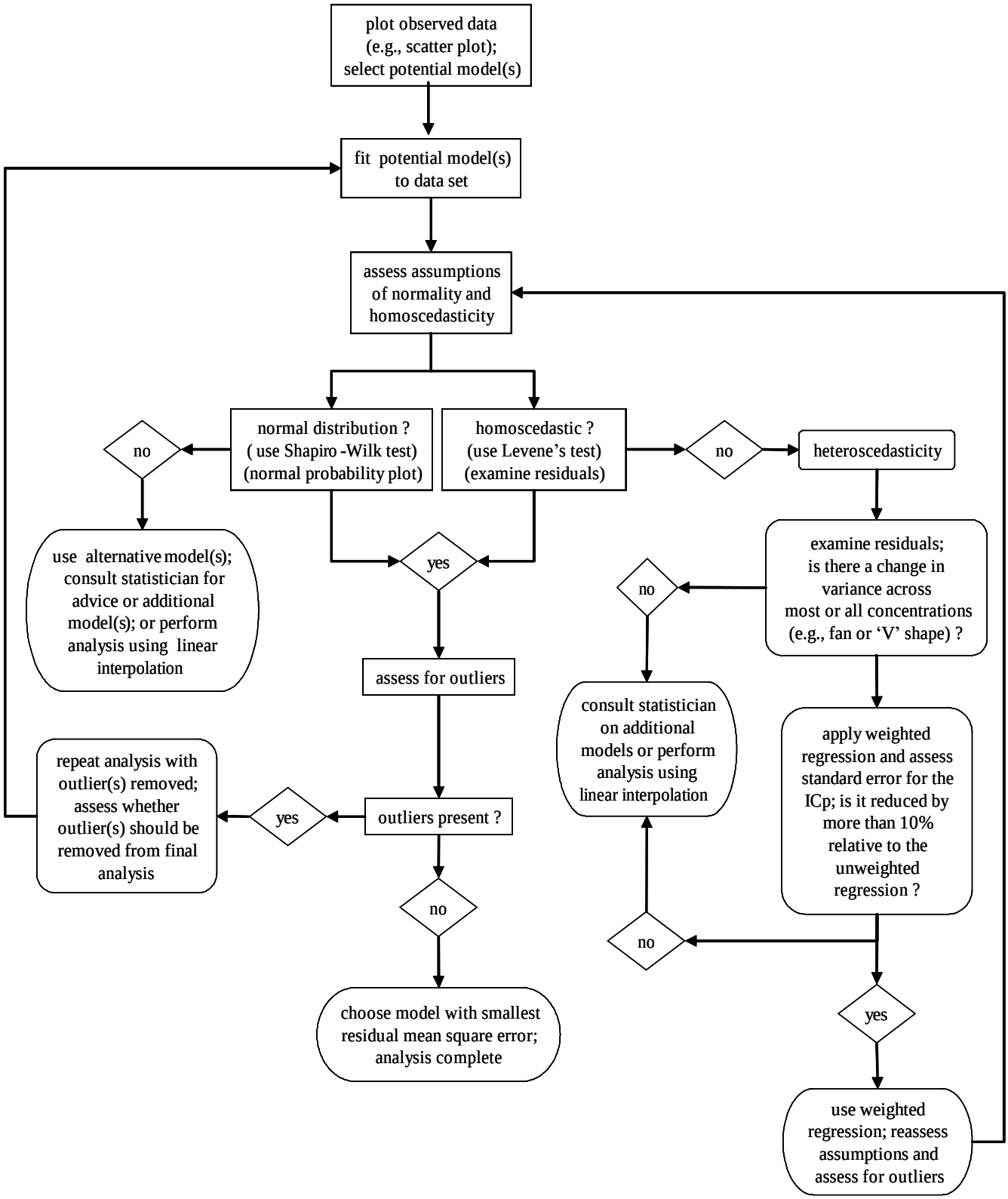

Figure 4 The general process for the statistical analysis and selection of the most appropriate model for quantitative toxicity data (adapted and modified from Stephenson *et al.*, 2000b)

If the regression procedure for one or more of the examined models meets the assumptions, the data (and regression) are examined for the presence of outliers. If an outlier has been observed, the test records and experimental conditions should be scrutinized for human error. If there are one or more outliers present, the analysis should be performed with and without the outlier(s), and the results of the analyses compared to examine the effect of the outlier(s) on the regression. Thereafter, a decision must be made as to whether the outlier(s) should be removed from the final analysis. The decision should take into consideration natural biological variation, and biological reasons that might have caused the apparent anomaly. Additional guidance on the presence of outliers and unusual observations is provided in Section 10.2 of EC (2005a).

If there are no outliers present or none are removed from the final analysis, the model that demonstrates the smallest residual mean square error is selected as the model of best choice.[130] Additional guidance from a statistician familiar with dealing with outlier data is also advised.

Normality should be assessed using the Shapiro-Wilk's test as described in EC (2005a). A normal probability plot of the residuals may also be used during the regression procedure, but is not recommended as a stand-alone test for normality as the detection of a "normal" or "non-normal" distribution depends on the subjective assessment of the user. If the data are not normally distributed, then the user is advised to try another model, consult a statistician for further guidance on model selection or to perform the less-desirable linear interpolation (using ICPIN; see Section 6.4.2.2) method of analysis.

Homoscedasticity of the residuals should be assessed using Levene's test as described in EC (2005a), and by examining the graphs of the residuals against the actual and predicted (estimated) values. Levene's test provides a definite indication of whether the data are homogeneous (e.g., as in Figure O.2A of Appendix O in EC, 2005a) or not. If the data (as indicated by Levene's test) are heteroscedastic (i.e., not homogeneous), then the graphs of the residuals should be examined. If there is a significant change in the variance and the graphs of the residuals produce a distinct fan or "V" pattern (refer to Figure O.2B, Appendix O in EC, 2005a for an example), then the data analysis should be repeated using weighted regression. Traditionally, the data have been weighted by dividing by the inverse of the variance; however, other options are available. Before choosing the weighted regression, the standard error of the ICp is compared to that derived from the unweighted regression. If there is a difference of greater than 10% between the two standard errors[131], then the weighted regression is selected as the regression of best choice. However, if there is less than a 10% difference in the standard error between the weighted and unweighted regressions, then the user should consult a statistician for the application of additional models, given the test data, or the data could be re-analyzed using the less-desirable linear interpolation (using ICPIN; see Section 6.4.2.2) method of analysis. This comparison between weighted and unweighted regression is completed for each of the selected models while proceeding through the process of final model selection (i.e., model and regression of best choice). Some non-divergent patterns might be indicative of an inappropriate or incorrect model (refer to Figure O.2C, Appendix O in EC, 2005a, for an example), and the user is again urged to consult a statistician for further guidance on the application of additional models.

Endpoints generated by regression analysis must be bracketed by test concentrations; extrapolation of endpoints beyond the highest test concentration is not an acceptable practice (EC, 2005a).

6.4.2.2 Linear interpolation using ICPIN

If regression analyses of the endpoint data (see Section 6.4.2.1) fail to provide an acceptable ICp for

[130] The Akaike Information Criterion (or an equivalent, such as the Bayesian Information Criterion) is another option for determining best model fit.

[131] The value of 10% is only a rule-of-thumb based upon experience. Objective tests for the improvement due to weighting are available, but beyond the scope of this document. Weighting should be used only when necessary, as the procedure can introduce additional complications to the modeling procedure. A statistician should be consulted when weighting is necessary.

reproductive inhibition (i.e., assumptions of normality and homoscedasticity cannot be met), linear interpolation using the computer program called *ICPIN* should be applied. This program (Norberg-King, 1993; USEPA, 1995, 2002) is not proprietary, and is included in most computer software for *environmental toxicology*, including TOXSTAT (1996) and CETIS. The original instructions for ICPIN from the USEPA are clearly written and make the program easy to use (Norberg-King, 1993).[132] An earlier version was called BOOTSTRP.

Analysis by ICPIN does not require equal numbers of replicates in different concentrations. The ICp is estimated by smoothing of the data as necessary, then using the two data points adjacent to the selected ICp (USEPA, 1995, Appendix L; USEPA, 2002, Appendix M). The ICp cannot be calculated unless there are test concentrations both lower and higher than the ICp; both those concentrations should have an effect reasonably close to the selected value of p, preferably within 20% of it. If the computer program does not use a logarithmic scale of concentration, users of the program must enter the concentrations as logarithms. Some commercial computer packages have the logarithmic transformation as a general option, but investigators should make sure that it is actually retained when proceeding to ICPIN. ICPIN estimates confidence limits by a special "bootstrap" technique because usual methods would not be valid. Bootstrapping performs many resamplings from the original measurements. The investigator must specify the number of resamplings, which can range from 80 to 1000. At least 400 is recommended here, and 1000 would be beneficial.[133]

If there are several adjacent high concentrations with no surviving juveniles, only the lowest of that string of concentrations should be used in analysis (i.e., the concentration closest to the middle of the series of concentrations used in the test). Normally, there is no particular benefit to including the additional concentrations, because they offer nothing to the analysis (i.e., the data consist only of zero progeny).

Besides determining and reporting the computer-derived ICps for earthworm reproduction at test end, a graph of percent reduction in number of live juveniles produced should be plotted against the logarithm of concentration, to check the mathematical estimations and to provide visual assessments of the nature of the data (EC, 2005a).

If the ICPIN program is used when there is a hormetic effect, an inherent smoothing procedure could change the control value and bias the estimate of ICp. Accordingly, before statistical analysis, hormetic values at low concentration(s) should be arbitrarily replaced by the control value. This is considered a temporary expedient until a superior approach is established (see Option 4, Section 10.3.3 in EC, 2005a). The correction is applied for any test concentration in which the average effect (i.e., the *geometric mean* of the replicate means) is higher ("better") than the average for the control. To apply this correction, replace the observed mean numbers of progeny of the replicates in the hormetic concentration(s), with the means of replicates in the control. The geometric average for that/those concentration(s) will then be the same as that for the control.

---

[132] The instructions in Norberg-King (1993) are sometimes misleading on the identity of "replicates." The term is used in such a way that it would apply to numbers of individual organisms within the same vessel. This slip of wording does not affect the functioning of the program. Some commercial programs have been less user-friendly for entry of data and analysis.

[133] ICPIN has some deficiencies, which is why it is recommended herein only in cases where use of regression fails to provide an acceptable ICp. Its interpolation method is an inefficient use of data, sensitive to peculiarities of the two concentrations used. The program fails to adopt logarithm of concentration, which would introduce a slight bias towards a higher value of ICp. A modification of the bootstrap method has now remedied a problem of overly narrow confidence limits; however, regression analyses provide more accurate methods of estimating the ICp and its 95% confidence limits (EC, 2005a) (see Section 6.4.2.1).

Section 7

Reporting Requirements

Each test-specific report must indicate if there has been any deviation from any of the *must* requirements delineated in Sections 2 to 6 and, if so, provide details of the deviation(s). The reader must be able to establish from the test-specific report whether the conditions and procedures preceding and during the test rendered the results valid and acceptable for the use intended.

Section 7.1 provides a list of items that must be included in each test-specific report. A list of items that must either be included in the test-specific report, provided separately in a general report, or held on file for a minimum of five years, is found in Section 7.2. Specific monitoring programs, related test protocols, or regulations might require selected test-specific items listed in Section 7.2 (e.g., details about the test material and/or explicit procedures and conditions during sample collection, handling, transport, and storage) to be included in the test-specific report, or might relegate certain test-specific information as *data to be held on file*.

Procedures and conditions common to a series of ongoing tests (e.g., routine toxicity tests for monitoring or compliance purposes) and consistent with specifications in this document may be referred to by citation or by attachment of a general report that outlines standard laboratory practice.

Details on the procedures, conditions, and findings of the test, which are not conveyed by the test-specific report or general report, must be kept on file by the laboratory for a minimum of five years so that the appropriate information can be provided if an audit of the test is required (Section 7.2).

7.1 Minimum Requirements for a Test-specific Report

The following items must be included in each test-specific report.

7.1.1 Test Substance or Material

- brief description of sample type (e.g., waste sludge, reference or contaminated field-collected soil, negative control soil) or coding, as provided to the laboratory personnel;
- information on labelling or coding of each sample;
- brief description of soil sampling, storage, and preparation (i.e., pretreatment) procedures;
- information on sample horizons as they were collected (i.e., number, relative depth of each soil horizon), for test, reference and negative control soils, if applicable;
- type of negative control soil (natural or artificial) and, if applicable, reference soil;
- date of sample collection; date and time sample(s) received at test facility; and
- sample temperature and moisture content upon receipt at the test facility.

7.1.2 Test Organisms

- species and source of breeding stock and test organisms;
- wet weight (mean ± SD) of organisms, at start of test; and
- any unusual appearance, behaviour, or treatment of the organisms, before their use in the test.

7.1.3 Test Facilities

- name and address of test laboratory; and
- name of person(s) performing the test (or each component of the test) and verifying results.

7.1.4 Test Method

- citation of biological test method used (i.e., as per this document);

- design and description if specialized procedure(s) (e.g., soil manipulation; preparation of mixtures of spiked soil; preparation and use of solvent and, if so, solvent control) or modification(s) of the standard test method described herein;

- brief description of frequency and type of all measurements and all observations made during test; and

- name and citation of program(s) and methods used for calculating statistical endpoints.

7.1.5 Test Conditions and Procedures

- design and description of any deviation(s) from, or exclusion of, any of the procedures and conditions specified in this document;

- number of discrete samples per treatment; target effect size for single-concentration reproduction tests (if applicable); number of replicate test vessels/units for each treatment; number and description of treatments in each test including the control(s); test concentrations (if applicable);

- volume and/or mass of soil in each test vessel or compartment of avoidance test unit;

- number of organisms per test vessel/unit and treatment;

- dates and times when test and control soils were prepared, test was started (i.e., organisms added to the test and control soils), and test was ended;

- feeding regime and ration, for the reproduction test;

- indication of assessment of soil moisture during the reproduction test;

- date when adults were removed from test vessels, for the reproduction test;

- for each soil sample, any measurements of soil particle size, moisture content, water-holding capacity, pH, TOC, OM, CEC, and electrical conductivity; and

- for each composite sample of subsamples taken at the same time from all replicates of each treatment, all measurements of temperature (air and soil), pH, moisture content, and water-holding capacity.

7.1.6 Test Results

- *for a reproduction test*: mean (± SD) percent survival of adult worms in each treatment on Day 28 (or Day 35, if applicable); mean (± SD) number of surviving juveniles in each treatment on Day 56 (or Day 63, if applicable); mean (± SD) number of surviving juveniles produced by each adult worm in control(s) on Day 56 (or Day 63, if applicable);

- *for an avoidance test*: % survival of all worms in each test unit at test end or mean % survival of worms per test unit where more than one replicate test unit is used for each test soil or test concentration; mean (± SD) number of surviving worms in replicates of each treatment representing clean soil and test soil, at 48 hours for single-concentration tests; % avoidance for each treatment for multi-concentration tests, and for single concentration tests, if calculated;

- any ICp (together with its 95% confidence limits) determined for the data on reproductive success (i.e., number of surviving juvenile worms in each treatment at test end); details regarding any transformation of data, and indication of quantitative statistical method used or procedures applied to the data;

- any LC50 or EC50 (including the associated 95% confidence limits and, if calculated, the slope) determined; any additional LCp or ECp (e.g., LC25 or EC25) calculated;

- for a multi-concentration test with chemical-spiked soil, indication as to whether results are based on nominal or measured concentrations of chemical(s) or chemical product(s); all values for measured concentrations and degree of difference from nominal strength;

- any 56-day IC50 (or 63-day IC50, if applicable) for inhibition of reproduction or 48-hour EC50 for avoidance (including its 95% confidence

limits) for multi-concentration tests; or % reduction in juvenile production relative to the control or % avoidance for positive controls, performed with the reference toxicant in conjunction with the definitive soil toxicity test; geometric mean value (± 2 SD) for the same test type, reference toxicant, and test species, as derived at the test facility in previous tests with a reference toxicant using the procedures and conditions for testing with a reference toxicant described herein; and

- anything unusual about the test, any problems encountered, any remedial measures taken.

7.2 Additional Reporting Requirements

This section provides a list of items that must be either included in the test-specific report or the general report, or held on file for a minimum of five years. Filed information must include the following, if available:

- a record of the chain-of-custody for field-collected or other samples tested for regulatory or monitoring purposes;

- a copy of the record of acquisition for the sample(s);

- chemical analytical data on the sample(s) not included in the test-specific report;

- bench sheets for the observations and measurements recorded during the test;

- bench sheets and warning chart(s) for the reference toxicity tests; and

- information on the calibration of equipment and instruments.

Original data sheets must be signed or initialled, and dated by the laboratory personnel conducting the tests.

7.2.1 Test Substance or Material

- name of person(s) who collected and/or provided the sample;

- records of sample log-entry sheets;

- appearance (e.g., odour, colour) and conditions (e.g., in darkness, in sealed container) of sample upon receipt and during storage; and

- any additional records obtained for field (e.g., field records provided or maintained during sample collection) or chemical samples (e.g., impurities, additives, structural formulae, etc.).

7.2.2 Test Organisms

- records and methods used for taxonomic confirmation of test species;

- history and age of breeding stock, for any culture used to provide test organisms;

- description of culture conditions and procedures for all laboratory cultures, including temperature, lighting, type and amount of substrate and details on its periodic renewal, and methods and records for substrate hydration; measurements and records of substrate quality, density of worms, records of culture condition, and health and performance indices; and any acclimation conditions and procedures (e.g., substrate, food, temperature), including rate of change;

- history of any batch of test organisms obtained from an outside source, including specifics related to the period(s) of holding and acclimation before their use in the test, type and amount of substrate and details on its periodic renewal, methods and records for substrate hydration; measurements and records of substrate quality, density of worms, records of culture condition, health and performance indices, and any acclimation conditions and procedures (e.g., substrate, food, temperature), including rate of change;

- procedures used to count, handle, sort, and transfer animals; and those to determine their mortality, condition, appearance, and behaviour; and

- source and composition of food, procedures used to prepare and store food, feeding method(s),

feeding frequency, and ration.

7.2.3 Test Facilities and Apparatus

- all results for initial tests with negative control soil and reference toxicant, undertaken by the laboratory previously inexperienced with performing the biological test methods described herein in advance of any reporting of definitive test results (see Section 3.2.1);
- description of systems for providing lighting and for regulating temperature within test facility;
- description of test vessels/avoidance units and covers; and
- description of procedures used to clean or rinse test apparatus.

7.2.4 Negative Control Soil or Reference Soil

- procedures for the preparation (if artificial soil) or pretreatment (if natural soil) of negative control soil;
- source of natural soil; history of past use and records of analysis for pesticides or other contaminants;
- formulation of artificial soil, including sources for the constituents and conditions and procedures for hydration and pH adjustment; and
- storage conditions and duration before use.

7.2.5 Test Method

- procedures used for mixing or otherwise manipulating test soils before use; time interval between preparation and testing;
- procedure used in preparing stock and/or test solutions of chemicals; description and concentration(s) of any solvent used;
- details concerning aliquot sampling, preparation, and storage before physicochemical analysis, together with available information regarding the analytical methods used (with citations); and
- use and description of preliminary or range-finding test.

7.2.6 Test Conditions and Procedures

- photoperiod and measurements of light intensity adjacent to surface of the test vessels;
- procedure for adding test organisms to test vessels/units;
- appearance of each sample (or mixture thereof) in test vessels/units; changes in appearance noted during test;
- records of the addition of test water to the surface of the soil in each test vessel throughout a reproduction test, for increasing moisture content;
- record of any growth of mould or fungi, and the presence and estimated quantity of any uneaten food;
- description of procedures used for the removal and counting of earthworms at the end of the test; records of the time and temperatures achieved during heat extraction, if used
- procedures used to assess and validate the efficiency of the heat-extraction procedure and records demonstrating the establishment and ongoing monitoring of the heat-extraction efficiency;
- any other physicochemical measurements (e.g., analyses of aliquots from the same batch to determine homogeneity, contaminant concentration, cations and anions, nitrogen, nitrate, nitrite, ammonia, phosphorus, potassium, C:N ratio, bulk density, total volatile solids, biochemical oxygen demand, chemical oxygen demand, total inorganic carbon, redox potential, soluble salts, metal oxides, sodium adsorption ratio, co-contaminants of concern, characteristics of contamination) made before and during the test on test material (including negative control soil and reference soil) and contents of test vessels/units, including analyses of whole soil and porewater;

- any other observations or analyses made on the test material (including samples of negative control soil or reference soil); e.g., qualitative and/or quantitative data regarding indigenous macrofauna or detritus, or results of geochemical analyses; and

- any chemical analyses of the concentration of chemical in stock solution(s) of reference toxicant and, if measured, in test concentrations.

7.2.7 Test Results

- results for any range-finding test(s) conducted;

- *for a reproduction test*: number of surviving adult worms in each test vessel on Day 28 (or Day 35, if applicable); number of surviving juveniles in each test vessel on Day 56 (or Day 63, if applicable); for regression analyses, information indicating sample size (e.g., number of replicates per treatment), parameter estimates with variance, any ANOVA table(s) generated, plots of fitted and observed values of any models used, and the output provided by the statistical program (e.g., SYSTAT);

- *for an avoidance test*: total number of surviving worms in clean soil and test soil within each test unit at 48 hours;

- warning chart showing the most recent and historic results for reference toxicity tests or positive control concentrations with the reference toxicant; CV for mean historical data derived for reference toxicity tests or positive control concentrations performed using the reference toxicant; and

- graphical presentation of data.

References

AAFC (Agriculture and Agri-Food Canada). 1998. The Canadian System of Soil Classification, 3rd edition, Soil Classification Working Group, Agriculture and Agri-Food Canada Publication 1646, National Research Council (NRC) Research Press, Ottawa. 187 p.

Addison JA. 2009. Distribution and Impacts of Invasive Earthworms in Canadian Forest Ecosystems, *Biological Invasions*, 11:59–79.

AE (Alberta Environment). 2007. Tier 2 Eco-contact Guideline Derivation Protocol, July 13, 2007 (Online Version). 33 p.

AESA (Alberta Environmentally Sustainable Agriculture Program). 2001. AESA Soil Quality Benchmark Study – Soil Organic Matter, Fact Sheet FS2001-ISQ, AESA Soil Quality Program, Conservation and Development Branch, Alberta Agriculture, Food and Rural Development, Edmonton, Alberta. 4 p.

Alves PRL, da Silva EB, Cardoso EJBN, and Alleoni LRF. 2018. Ecotoxicological Impact of Arsenic on Earthworms and Collembolans as Affected by Attributes of a Highly Weathered Tropical Soil, *Environmental Science and Pollution Research*, 25:13217–13225.

Alves PRL, Bandeira FO, Giraldi M, Presotto R, Segat JC, Cardoso EJBN, and Baretta D. 2019. Ecotoxicological Assessment of Fluazuron: Effects on *Folsomia candida* and *Eisenia andrei*, *Environmental Science and Pollution Research*, 26:5842–5850.

Aquaterra Environmental. 1998. Development of Earthworm Toxicity Tests for Assessment of Contaminated Soils, Report prepared for Environment Canada, Ottawa, Ontario.

Aquaterra Environmental and ESG (Ecological Services Group International Inc.). 2000. Assessment of the Biological Test Methods for Terrestrial Plants and Soil Invertebrates: Metals, Report prepared for Environment Canada, Ottawa, Ontario.

ASTM (American Society for Testing and Materials). 1999. Standard Guide for Conducting Laboratory Soil Toxicity or Bioaccumulation Tests with the Lumbricid Earthworm *Eisenia fetida*, ASTM , Philadelphia, Pennsylvania, ASTM E1676-97. 18 p.

ASTM. 2012. Standard Guide for Conducting Laboratory Soil Toxicity or Bioaccumulation Tests with the Lumbricid Earthworm *Eisenia Fetida* and the Enchytraeid Potworm *Enchytraeus albidus*, ASTM International, West Conshohocken, Pennsylvania, ASTM E1676-12. 27 p.

ASTM. 2014. Standard Guide for Use of Lighting in Laboratory Testing, ASTM International, West Conshohocken, Pennsylvania, ASTM E1733-95(2014). 12 p.

Barber I, Bembridge J, Dohmen P, Edwards P, Heimbach F, Heusel R, Romijn K, and Rufli H. 1997. Development and Evaluation of Triggers for Earthworm Toxicity Testing with Plant Protection Products, In: *Advances in Earthworm Ecotoxicology*, Sheppard SC, Bembridge JD, Holmstrup M, and Posthuma L (eds.), SETAC Press, Pensacola, Florida, pp. 269–278.

Bauer C, and Römbke J. 1997. Factors Influencing the Toxicity of Two Pesticides on Three Lumbricid Species in Laboratory Tests, *Soil Biology and Biochemistry*, 29:705–708.

BBA (Biologische Bundesanstalt für Land- und Forstwirtschaft). 1994. Auswirkungen von Pflanzenschutzmitteln auf die Reproduktion und das Wachstum von *Eisenia fetida*/*Eisenia andrei*, *Richtlinien für die Prüfung von Pflanzenschutzmitteln im Zulassungsverfahren*, Teil VI, 2-2.

Becker L, Scheffczyk A, Förster B, Oehlmann J, and Princz J. 2011. Effects of Boric Acid on Various Microbes, Plants, and Soil Invertebrates, *Journal of Soils and Sediments*, 11:238–248.

Becker-van Slooten K, Campiche S, and Tarradellas J. 2003. Research in Support of the Environment Canada Collembolan Toxicity Test Method with *Folsomia candida* for Assessment of Contaminated Soils, Report prepared for Environment Canada, Ottawa, Ontario.

Becker-van Slooten K, Campiche S, and Tarradellas J. 2004. Research in Support of the Environment Canada Soil Toxicity Test Methods: Consideration of Various Methods for Measuring Soil pH, Water Holding Capacity, and Water Filled Pore Space, Report prepared for Environment Canada, Ottawa, Ontario. 29 p.

Bengtsson G, Gunnarsson T, and Rundgren S. 1986. Effects of Metal Pollution on the Earthworm *Dendrobaena rubida* (Sav.) in Acidified Soils, *Water, Air, and Soil Pollution*, 28:361–383.

Berman DI, Meshcheryakova EN, and Leirikh AN. 2010. Egg Cocoons of the Earthworm *Dendrodrilus rubidus tenuis* (Lumbricidae, Oligochaeta) Withstand the Temperature of Liquid Nitrogen, *Doklady Biological Sciences*, 434:347–350.

Bonnell Environmental Consulting. 1994. Assessment of Soil Toxicity Test Species for Canadian Representativeness, Technical Report TS-28, Prepared for Environment Canada, Ottawa, Ontario.

Bouché MB. 1988. Earthworm Toxicological Test, Hazard Assessment and Biomonitoring – A Methodological Approach, In: *Earthworms in Waste and Environmental Management*, Edwards CA and Neuhauser EF (eds.), SPB Academic Publishing, The Hague, The Netherlands, pp. 315–320.

Bouché MB. 1992. Earthworm Species and Ecotoxicological Studies, In: *Ecotoxicology of Earthworms*, Greig-Smith PW, Becker H, Edwards PJ, and Heimbach F (eds.), Intercept Ltd., Andover, United Kingdom, pp. 20–35.

Bouguerra S, Gavina A, Ksibi M, Rasteiro Mda G, Rocha-Santos T, and Pereira R. 2016. Ecotoxicity of Titanium Silicon Oxide ($TiSiO_4$) Nanomaterial for Terrestrial Plants and Soil Invertebrate Species, *Ecotoxicology and Environmental Safety*, 129:291–301.

Bouguerra S, Gavina A, Rasteiro Mda G, Rocha-Santos T, Ksibi M, and Pereira R. 2019. Effects of Cobalt Oxide Nanomaterial on Plants and Soil Invertebrates at Different Levels of Biological Organization, *Journal of Soils and Sediments*, 19:3018–3034.

Bradham KD, Dayton EA, Basta NT, Schroder J, Payton M, and Lanno RP. 2006. Effect of Soil Properties on Lead Bioavailability and Toxicity to Earthworms, *Environmental Toxicology and Chemistry*, 25:769–775.

Brami C, Glover AR, Butt KR, and Lowe CN. 2017. Avoidance, Biomass and Survival Response of Soil Dwelling (Endogeic) Earthworms to OECD Artificial Soil: Potential Implications for Earthworm Ecotoxicology, *Ecotoxicology*, 26:576–579.

Brousseau P, Fugere N, Bernier J, Coderre D, Nadeau D, Poirier G, and Fournier M. 1997. Evaluation of Earthworm Exposure to Contaminated Soil by Cytometric Assay of Coelomocytes Phagocytosis in *Lumbricus terrestris* (Oligochaeta), *Soil Biology and Biochemistry*, 29:681–684.

Button M, Jenkin GR, Bowman KJ, Harrington CF, Brewer TS, Jones GD, and Watts MJ. 2010. DNA Damage in Earthworms from Highly Contaminated Soils: Assessing Resistance to Arsenic Toxicity by Use of the Comet Assay, *Mutation Research*, 696:95–100.

Button M, Koch I, and Reimer KJ. 2012. Arsenic Resistance and Cycling in Earthworms Residing at a Former Gold Mine in Canada, *Environmental Pollution*, 169:74–80.

Callahan CA. 1988. Earthworms as Ecotoxicological Assessment Tools, In: *Earthworms in Waste and Environmental Management*, Edwards CA and Neuhauser EF (eds.), SPB Academic Publishing, The Hague, The Netherlands, pp. 295–301.

Cao Q, Steinman AD, Yao L, and Xie L. 2017. Toxicological and Biochemical Responses of the Earthworm *Eisenia fetida* to Cyanobacteria Toxins, *Scientific Reports*, 7:15954.

Carter, MR (ed.). 1993. *Soil Sampling and Methods of Analysis*, Lewis Publishers, CRC Press Inc., Boca Raton, Florida.

Carter MR, and Gregorich EG (eds.). 2008. *Soil Sampling and Methods of Analysis*, *Second Edition*, Taylor & Francis Group, CRC Press, Boca Raton, Florida.

Cermak J, Stephenson G, Birkholz D, and Dixon DG. 2013. Toxicity and Toxicokinetics of Binary Combinations of Petroleum Hydrocarbon Distillates with the Earthworm *Eisenia andrei*, *Environmental Toxicology and Chemistry*, 32:1016–1026.

Chang LW, Meier JR, and Smith MK. 1997. Application of Plant and Earthworm Bioassays to Evaluate Remediation of a Lead-contaminated Soil, *Archives of Environmental Contamination and Toxicology*, 32:166–171.

Chelinho S, Domene X, Campana P, Natal-da-Luz T, Scheffczyk A, Römbke J, Andrés P, and Sousa JP. 2011. Improving Ecological Risk Assessment in the Mediterranean Area: Selection of Reference Soils and Evaluating the Influence of Soil Properties on Avoidance and Reproduction of Two Oligochaete Species, *Environmental Toxicology and Chemistry*, 30:1050–1058.

Chen Y, Liu X, Leng Y, and Wang J. 2020. Defense Responses in Earthworms (*Eisenia fetida*) Exposed to Low-Density Polyethylene Microplastics in Soils, *Ecotoxicology and Environmental Safety*, 187:109788.

Christensen OM, and Mather JG. 1994. Earthworms as Ecotoxicological Test-Organisms, Technical Report, Danish Environmental Protection Agency, Copenhagen, Denmark. 99 p.

Cikutovic MA, Fitzpatrick LC, Venables BJ, and Goven AJ. 1993. Sperm Count in Earthworms (*Lumbricus terrestris*) as a Biomarker for Environmental Technology: Effects of Cadmium and Chlordane, *Environmental Pollution*, 81:123–125.

Conder JM, and Lanno RP. 2000. Evaluation of Surrogate Measures of Cadmium, Lead, and Zinc Bioavailability to *Eisenia fetida*, *Chemosphere*, 41:1659–1668.

Cooke AS, Greig-Smith PW, and Jones SA. 1992. Consequences for Vertebrate Wildlife of Toxic Residues in Earthworm Prey, In: *Ecotoxicology of Earthworms*, Greig-Smith PW, Becker H, Edwards PJ, and Heimbach F (eds.), Intercept Ltd., Andover, United Kingdom, p. 139–155.

Coulson SJ, Fjellberg A, Gwiazdowicz DJ, Lebedeva NV, Melekhina EN, Solhøy T, Erséus C, Maraldo K, Miko L, Schatz H, Schmelz RM, Søli G, and Stur E. 2013. Introduction of Invertebrates into the High Arctic via Imported Soils: The Case of Barentsburg in the Svalbard, *Biological Invasions*, 15:1–5.

Courchesne F, Savoie S, and Dufresne A. 1995. Effects of Air-Drying on the Measurement of Soil pH in Acidic Forest Soils of Quebec, Canada, *Soil Science*, 160:56–68.

Crépin J, and Johnson RL. 1993. Soil Sampling for Environmental Assessment, Chapter 2, In: *Soil Sampling and Methods of Analysis*, Carter MR (ed.), Lewis Publishers, CRC Press Inc., Boca Raton, Florida, pp. 5–18.

Csuzdi C, Chang CH, Pavlícek T, Szederjesi T, Esopi D, and Szlávecz K. 2017. Molecular Phylogeny and Systematics of Native North American Lumbricid Earthworms (Clitellata: Megadrili), *PLoS One*, 12:e0181504.

Curry JP. 1988. The Ecology of Earthworms in Reclaimed Soils and their Influence on Soil Fertility, In: *Earthworms in Waste and Environmental Management*, Edwards CA and Neuhauser EF (eds.), SPB Academic Publishing, The Hague, The Netherlands, pp. 251–261.

Deitzer G. 1994. Spectral Comparisons of Sunlight and Different Lamps, In: *Proceedings of International Lighting in Controlled Environments Workshop*, Tibbits TW (ed.), Madison, Wisconsin, pp. 197–199.

Delgadillo V, Verdejo J, Mondaca P, Verdugo G, Gaete H, Hodson ME, and Neaman A. 2017. Proposed Modification to Avoidance Test with *Eisenia fetida* to Assess Metal Toxicity in Agricultural Soils Affected by Mining Activities, *Ecotoxicology and Environmental Safety*, 140:230–234.

de Lima e Silva C, de Rooij W, Rudo Verweij RA, and van Gestel CAM. 2020. Toxicity in Neonicotinoids to *Folsima candida* and *Eisenia andrei*, *Environmental Toxicology and Chemistry*, 39:548–555.

de Santo FB, Guerra N, Vianna MS, Torres JPM, Marchioro CA, and Niemeyer JC. 2019. Laboratory and Field Tests for Risk Assessment of Metsulfuron-Methyl-Based Herbicides for Soil Fauna, *Chemosphere*, 222:645–655.

Demuynck S, Lebel A, Grumiaux F, Pernin C, Leprêtre A, and Lemière S. 2016. Comparative Avoidance Behaviour of the Earthworm *Eisenia fetida* towards Chloride, Nitrate and Sulphate Salts of Cd, Cu and Zn Using Filter Paper and Extruded Water Agar Gels as Exposure Media, *Ecotoxicology and Environmental Safety*, 129:66–74.

Denton D, Diamond J, and Zheng L. 2011. Test of Significant Toxicity: A Statistical Application for Assessing Whether an Effluent or Site Water is Truly Toxic, *Environmental Toxicology and Chemistry*, 30:1117–1126.

Denton D, Diamond J, and Stuber R. 2019. Comparison of False-Positive Rates of 2 Hypothesis-Test Approaches in Relation to Laboratory Toxicity Test Performance, *Environmental Toxicology and Chemistry*, 38:511–523.

Dodard SG, Sarrazin M, Hawari J, Paquet L, Ampleman G, Thiboutot S, and Sunahara GI. 2013. Ecotoxicological Assessment of a High Energetic and Insensitive Munitions Compound: 2,4-Dinitroanisole (DNAN), *Journal of Hazardous Materials*, 262:143–150.

Domínguez J, Velando A, and Ferreiro A. 2005. Are *Eisenia fetida* (Savigny, 1826) and *Eisenia andrei* Bouché (1972) (Oligochaeta, Lumbricidae) Different Biological Species?, *Pedobiologia*, 49:81–87.

Domínguez J, and Edwards CA. 2010. Biology and Ecology of Earthworm Species Used for Vermicomposting, In: *Vermiculture Technology: Earthworms, Organic Wastes, and Environmental Management*, Edwards CA, Arancon NQ, and Sherman RL (eds.), CRC Press, Boca Raton, Florida, pp. 27–40.

Drewes CD, Vining EP, and Callahan CA. 1984. Non-Invasive Electrophysiological Monitoring: A Sensitive Method for Detecting Sublethal Neurotoxicity in Earthworms, *Environmental Toxicology and Chemistry*, 3:599–607.

Dunnett CW. 1955. A Multiple Comparison Procedure for Comparing Several Treatments with a Control, *Journal of the American Statistical Association*, 50:1096–1121.

EC (Environment Canada). 1995. Guidance Document on Measurement of Toxicity Test Precision Using Control Sediments Spiked with a Reference Toxicant, Environment Canada, Ottawa, Ontario, Report EPS 1/RM/30. 69 p.

EC. 1999. Guidance Document on Application and Interpretation of Single-species Tests in Environmental Toxicology, Environment Canada, Ottawa, Ontario, Report EPS 1/RM/34. 226 p.

EC. 2004a. Inter-Laboratory Validation of Environment Canada's New Test Methods for Measuring Soil Toxicity Using Earthworms, Technical Report, Environment Canada, Ottawa, Ontario.

EC. 2004b. Biological Test Method: Tests for Toxicity of Contaminated Soil to Earthworms (*Eisenia andrei*, *Eisenia fetida*, or *Lumbricus terrestris*), Environment Canada, Ottawa, Ontario, Report EPS 1/RM/43, including June 2007 amendments. 184 p.

EC. 2004c. Proceedings to the Workshop on Toxicity Test Methodologies for Assessing the Impacts of Contaminant Mixtures in Soil Using Terrestrial Species of Ecological Relevance to Canadian Soil Systems, Workshop Held February 19–21, 2003 at the Pacific Environmental Science Centre, North Vancouver, British Columbia, Report published by Environment Canada, Ottawa, Ontario. 69 p.

EC. 2005a. Guidance Document on Statistical Methods for Environmental Toxicity Tests, Environment Canada, Ottawa, Ontario, Report EPS 1/RM/46. 283 p.

EC. 2005b. Biological Test Method: Test for Measuring Emergence and Growth of Terrestrial Plants Exposed to Contaminants in Soil, Environment Canada, Ottawa, Ontario, Report EPS 1/RM/45, including June 2007 amendments. 159 p.

EC. 2007a. Biological Test Method: Test of Reproduction and Survival Using the Cladoceran *Ceriodaphnia dubia*, 2nd edition, Environment Canada, Ottawa, Ontario, Report EPS 1/RM/21. 100 p.

EC. 2007b. Biological Test Method: Test for Measuring the Inhibition of Growth Using the Freshwater Macrophyte, *Lemna minor*, 2nd edition, Environment Canada, Ottawa, Ontario, Report EPS 1/RM/37. 141 p.

EC. 2010. Development and Standardization of New Toxicity Test Methodologies and Guidance for Assessment and Remediation of Impacts from Hydrocarbon and Brine Contamination on Boreal Forest, Taiga and Northern Soil using Organisms Representative of the Eco-zone Regions, Five Year Technical Progress Report, Report prepared for Program for Energy Research and Development, Ottawa, Ontario, 167 p.

EC. 2011a. Biological Test Method: Test of Larval Growth and Survival Using Fathead Minnows, 2nd edition, Environment Canada, Ottawa, Ontario, Report EPS 1/RM/22. 100 p.

EC. 2011b. Biological Test Method: Fertilization Assay Using Echinoids (Sea Urchins and Sand Dollars), 2nd edition, Environment Canada, Ottawa, Ontario, Report EPS 1/RM/27. 140 p.

EC. 2012. Guidance Document on the Sampling and Preparation of Contaminated Soil for Use in Biological Testing, Environment Canada, Ottawa, Ontario, Report EPS 1/RM/53. 256 p.

EC. 2013a. Development and Standardization of Toxicity Test Methodologies for Assessment of Impacts from Contamination using Species and Microbial Communities Representative of Canadian Boreal and Taiga Eco-Zones, Technical Report: 2010–2013, Report prepared for Program for Energy Research and Development, Ottawa, Ontario, 245 p.

EC. 2013b. Biological Test Method: Test for Growth in Contaminated Soil Using Terrestrial Plants Native to the Boreal Region, Environment Canada, Ottawa, Ontario, Report EPS 1/RM/56. 137 p.

EC. 2014a. Biological Test Method: Test for Measuring Survival and Reproduction of Springtails Exposed to Contaminants in Soil, 2nd edition, Environment Canada, Ottawa, Ontario, Report EPS 1/RM/47. 184 p.

EC. 2014b. Estimation of Chemical Bioavailability to Soil Organisms with CMP Substances from the Medium Priority List: Comparison of the Toxicity of Zinc Moieties, Technical Report, Environment Canada, Ottawa, Ontario, 148 p.

ECASG (Ecological Criteria Advisory Sub Group). 2006. Five-Year Review of the Canada-Wide Standards for Petroleum Hydrocarbons (PHC CWS); Ecological, Direct Soil Contact Guidance, Report to the Canadian Council of Ministers of Environment Soil Quality Guidelines Task Group, 5 p.

ECCC (Environment and Climate Change Canada). 2020a. Biological Test Method: Test for Measuring Reproduction of Oribatid Mites Exposed to Contaminants in Soil, Environment and Climate Change Canada, Ottawa, Ontario, Report STB 1/RM/61. 135 p.

ECCC. 2020b. An Evaluation of the Earthworm Reproduction Test using *Eisenia andrei* and *Dendrodrilus rubidus*, Technical Report, Environment and Climate Change Canada, Ottawa, Ontario. 85 p.

EcoDynamics Consulting Inc. 2007. Sampling in the Swan Hills and Peace River Areas of Alberta and the Torch River Area of Saskatchewan, Report prepared for Environment Canada, Ottawa, Ontario. 23 p.

EcoDynamics Consulting Inc. 2008. Sampling of a Reference Soil in the Bathurst Region of New Brunswick – Ecological Site Characterization, Report prepared for Environment Canada, Ottawa, Ontario. 23 p.

EcoDynamics Consulting Inc. 2011a. Ecological Description of the Ontario Reference Site, Report prepared for Environment Canada, Ottawa, Ontario.

EcoDynamics Consulting Inc. 2011b. Soil Sampling and Ecological Description at the Newfoundland Reference Site, Report prepared for Environment Canada, Ottawa, Ontario. 8 p.

EcoDynamics Consulting Inc. 2011c. Site SK09 Robins Lake, Saskatchewan – Soil Sampling and Ecological Description, Report prepared for Environment Canada, Ottawa, Ontario. 10 p.

Edwards CA, and Bohlen PJ. 1992. The Effects of Toxic Chemicals on Earthworms, *Reviews of Environmental Contamination and Toxicology*, 125:23–99.

Edwards CA, and Bohlen PJ. 1996. *Biology and Ecology of Earthworms*, 3rd edition, Chapman & Hall, London, United Kingdom. 426 p.

Edwards CA, and Lofty JR. 1977. *Biology of Earthworms*, Chapman & Hall, London, United Kingdom. 334 p.

Elvira C, Domínguez J, and Mato S. 1997. The Growth and Reproduction of *Lumbricus rubellus* and *Dendrobaena rubida* in Cow Manure Mixed Cultures with *Eisenia andrei*, *Applied Soil Ecology*, 5:97–103.

ESG (Ecological Services Group International Inc.). 2000. Final Report on the Acute Screening and Definitive, Chronic Toxicity Tests with Motor Gasoline, Report prepared for Petroleum Technology Alliance Canada, Calgary, Alberta.

ESG. 2001. Toxicity of Petroleum Hydrocarbons to Soil Organisms and the Effects on Soil Quality: Phase 1 Fraction-Specific Toxicity of Crude Oil, Report prepared for Petroleum Technology Alliance Canada, Calgary, Alberta.

ESG. 2002. Impacts of Metal-Contaminated Forest Soils from the Canadian Shield on Terrestrial Organisms, Report prepared for Metals in the Environment Research Network Canadian Network of Toxicology Centres, Guelph, Ontario and Environment Canada, Ottawa, Ontario.

ESG and Aquaterra Environmental. 2002. Assessment of the Biological Test Methods for Terrestrial Plants and Soil Invertebrates: Pesticides, Report prepared for Environment Canada, Ottawa, Ontario.

Fender WM. 1985. Earthworms of the Western United States, Part 1. Lumbricidae, *Megadriologica*, 4:93–129.

Ferreiro A, Domínguez J, and Velado A. 2002. The Taxonomy and Reproductive Behaviour of *Eisenia fetida* and *Eisenia andrei* (Lumbricidae), In: *Proceedings from the 7th International Symposium on Earthworm Ecotoxicology*, Cardiff, Wales, p. 245.

Fischer E, and Molnár L. 1997. Growth and Reproduction of *Eisenia fetida* (Oligochaeta, Lumbricidae) in Semi-Natural Soil Containing Various Metal Chlorides, *Soil Biology and Biochemistry*, 29:667–670.

Fitzpatrick LC, Sassani R, Venables BJ, and Goven AJ. 1992. Comparative Toxicity of Polychlorinated Biphenyls to Earthworms *Eisenia foetida* and *Lumbricus terrestris*, *Environmental Pollution*, 77:65–69.

Frenot Y. 1992. Introduced Populations of *Dendrodrilus rubidus* ssp. (Oligochaeta:Lumbricidae) at Crozet, Kerguelen and Amsterdam Islands: Effects of Temperature on Growth Patterns During the Juvenile Stages, *Soil Biology and Biochemistry*, 24:1433–1439.

Gibbs MH, Wicker LF, and Stewart AJ. 1996. A Method for Assessing Sublethal Effects of Contaminants in Soils to the Earthworm, *Eisenia foetida*, *Environmental Toxicology and Chemistry*, 15:360–368.

Giggleman MA, Fitzpatrick LC, Goven AJ, and Venables BJ. 1998. Effects of Pentachlorophenol on Survival of Earthworms (*Lumbricus terrestris*) and Phagocytosis by their Immmunoactive Coelomocytes, *Environmental Toxicology and Chemistry*, 17:2391–2394.

GISD (Global Invasive Species Database). 2020. Species profile: *Dendrodrilus rubidus*. URL: http://www.iucngisd.org/gisd/speciesname/Dendrodrilus+rubidus. Last accessed: June 23, 2021.

Goven AJ, Fitzpatrick LC, Eyambe GS, Venables BJ, and Cooper EL. 1993. Cellular Biomarkers for Measuring Toxicity of Xenobiotics: Effects of Polychlorinated Biphenyls on Earthworm *Lumbricus terrestris* Coelomocytes, *Environmental Toxicology and Chemistry*, 12:863–870.

Goven AJ, Chen SC, Fitzpatrick LC, and Venables BJ. 1994. Lysozyme Activity in Earthworm (*Lumbricus terrestris*) Coelomic Fluid and Coelomocytes: Enzyme Assay for Immunotoxicity of Xenobiotics, *Environmental Toxicology and Chemistry*, 13:607–613.

Green JW. 2014. Power and Control Choice in Aquatic Experiments with Solvents, *Ecotoxicology and Environmental Safety*, 102:142–146.

Haimi J, and Paavola S. 1998. Responses of Two Earthworm Populations with Different Exposure Histories to Chlorophenol Contamination, *Environmental Toxicology and Chemistry*, 17:1114–1117.

Hartenstein R. 1982. Effect of Aromatic Compounds, Humic Acids and Lignins on Growth of the Earthworm (*Eisenia foetida*), *Soil Biology and Biochemistry*, 14:595–599.

Hausenbuiller RL. 1985. *Soil Science–Principles & Practices, 3rd edition*, W.C. Brown Publisher, Dubuque, Iowa. 610 p.

Heimbach F. 1993. Use of Laboratory Toxicity Tests for the Hazard Assessment of Chemicals of Earthworms Representing the Soil Fauna, In: *Integrated Soil and Sediment Research: A Basis for Proper Protection*, Eijsackers HJP and Hamers T (eds.), Kluwer Academic Publishers, The Netherlands, pp. 299–302.

Heimbach F. 1997. Comparison of the Sensitivities of an Earthworm (*Eisenia fetida*) Reproduction Test and a Standardized Field Test on Grassland, In: *Advances in Earthworm Ecotoxicology*, Sheppard SC, Bembridge JD, Holmstrup M, and Posthuma L (eds.), SETAC Press, Pensacola, Florida, pp. 235–245.

Heimbach F, and Edwards PJ. 1983. The Toxicity of 2-Chloroacetamide and Benomyl to Earthworms Under Various Test Conditions in an Artificial Soil Test, *Pesticide Science*, 14:635–636.

Hendershot WH, Lalonde H, and Duquette M. 1993. Soil Reaction and Exchangeable Acidity, In: *Soil Sampling and Methods of Analysis*, Carter MR (ed.), Lewis Publishers, Boca Raton, Florida, pp. 141–145.

Hirano T, and Tamae K. 2011. Earthworms and Soil Pollutants, *Sensors*, 11:11157-11167.

Hoenig JM, and Heisey DM. 2001. The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis, *The American Statistician*, 55:19–24.

Hund K. 1998. Earthworm Avoidance Test for Soil Assessment: Alternative for Acute and Reproduction Test, In: *Contaminated Soil '98*, Vol. 2, Proceedings of the Sixth International FZK/TNO Conference on Contaminated Soil, Thomas Telford Publishing Ltd., London, United Kingdom, pp. 1039–1040.

Hund-Rinke K, and Wiechering H. 2001. Earthworm Avoidance Test for Soil Assessments: An Alternative for Acute and Reproduction Tests, *Journal of Soils and Sediments*, 1:15–20.

Hund-Rinke K, Achazi R, Römbke J, and Warnecke D. 2003. Avoidance Test with *Eisenia fetida* as Indicator for the Habitat Function of Soils: Results of a Laboratory Comparison Test, *Journal of Soils and Sediments*, 3:7–12.

Hund-Rinke K, Lindemann M, and Simon M. 2005. Experiences with Novel Approaches in Earthworm Testing Alternatives, *Journal of Soils and Sediments*, 5:233–239.

Hutchinson TH, Shillabeer N, Winter MJ, and Pickford DB. 2006. Acute and Chronic Effects of Carrier Solvents in Aquatic Organisms: A Critical Review, *Aquatic Toxicology*, 76:69–92.

Ingraldi S, Princz J, and Scroggins R. 2004. Comparison of *E. andrei* and *E. fetida* and their Response to Contamination, Technical Report prepared for Environment Canada, Ottawa, Ontario.

ISO (International Organization for Standardization). 1991. Soil – Determination of the Effect of Chemical Substances on the Reproduction of Earthworms, Proposal from The Netherlands, Geneva, Switzerland. 9 p.

ISO. 1993. Soil Quality – Effects of Pollutants on Earthworms (*Eisenia fetida*) – Part 1: Determination of Acute Toxicity Using Artificial Soil Substrate, Geneva, Switzerland, ISO 11268-1:1993. 6 p.

ISO. 1995. Soil Quality – Determination of Organic and Total Carbon After Dry Combustion (Elementary Analysis), Geneva, Switzerland, ISO 10694:1995. 7 p.

ISO. 1998. Soil Quality – Effects of Pollutants on Earthworms (*Eisenia fetida*) – Part 2: Determination of Effects on Reproduction, Geneva, Switzerland, ISO 11268-2:1998. 16 p.

ISO. 1999. Soil Quality – Inhibition of Reproduction of Collembola (*Folsomia candida*) by Soil Pollutants, Geneva, Switzerland, ISO 11267:1999. 16 p.

ISO. 2003. Avoidance Test for Evaluating the Quality of Soils and the Toxicity of Chemicals – Test with Earthworms (*Eisenia fetida/ andrei*), Draft Document, Geneva, Switzerland, Doc ISO/TC/190/SC 4/WG 2 N1238. 19 p.

ISO. 2005. Soil quality – Vocabulary. Geneva, Switzerland, ISO 11074:2005. 113 p.

ISO. 2008. Soil quality – Avoidance test for determining the quality of soils and effects of chemicals on behavior – Part 1: Test with Earthworms (*Eisenia fetida* and *Eisenia andrei*), Geneva, Switzerland, ISO 17512-1:2008. 25 p.

ISO. 2012. Soil Quality – Effects of Pollutants on Earthworms – Part 2: Determination of Effects on Reproduction of *Eisenia fetida*/*Eisenia andrei*, Geneva, Switzerland, ISO 11268-2:2012. 21 p.

ISO. 2019. Soil Quality – Identification of Ecotoxicological Test Species by DNA Barcoding, Geneva, Switzerland, ISO 21286:2019. 20 p.

Jesmer AH, Velicogna JR, Schwertfeger DM, Scroggins RP, and Princz JI. 2017. The Toxicity of Silver to Soil Organisms Exposed to Silver Nanoparticles and Silver Nitrate in Biosolids-Amended Field Soil, *Environmental Toxicology and Chemistry*, 36:2756–2765.

Jaenike J. 1982. *Eisenia foetida* is Two Biological Species, *Megadrilogica*, 4:6–8.

Karnak RE, and Hamelink JL. 1982. A Standardized Method for Determining the Acute Toxicity of Chemicals to Earthworms, *Ecotoxicology and Environmental Safety*, 6:216–222.

Keddy CJ, Greene JC, and Bonnell MA. 1995. Review of Whole-Organism Bioassays: Soil, Freshwater Sediment, and Freshwater Assessment in Canada, *Ecotoxicology and Environmental Safety*, 30:221–251.

Kilpi-Koski J, Penttinen O, Väisänen AO, and van Gestel CAM. 2020. Toxicity of Binary Mixtures of Cu, Cr and As to the Earthworm *Eisenia andrei*, *Ecotoxicology*, 29:900–911.

Komex International. 1995. Sampling and Shipping of Reference Soil for the Terrestrial Soil Toxicity Method Development Project, Report prepared for Environment Canada, Ottawa, Ontario. 17 p.

Kula H. 1995. Comparison of Laboratory and Field Testing for the Assessment of Pesticide Side Effects on Earthworms, *Acta Zoologica Fennica*, 196:338–341.

Kula H, and Kokta C. 1992. Side Effects of Selected Pesticides on Earthworms Under Laboratory and Field Conditions, *Soil Biology and Biochemistry*, 24:1711–1714.

Kula H, and Larink O. 1997. Tests on the Earthworms *Eisenia fetida* and *Aporrectodea caliginosa*, In: *Handbook of Soil Invertebrate Toxicity Tests*, Løkke H and van Gestel CAM (eds.), John Wiley & Sons, Chichester, United Kingdom, pp. 95–112.

Lackmann C, Velki M, Seiler TB, and Hollert H. 2018. Herbicides Diuron and Fluazifop-p-Butyl Affect Avoidance Response and Multixenobiotic Resistance Activity in Earthworm *Eisenia andrei*, *Chemosphere*, 210:110–119.

Lanno RP, Oorts K, Smolders E, Albanese K, and Chowdhury MJ. 2019. Effects of Soil Properties on the Toxicity and Bioaccumulation of Lead in Soil Invertebrates, *Environmental Toxicology and Chemistry*, 38:1486–1494.

Lee KE. 1985. *Earthworms: Their Ecology and Relationship with Soils and Land Use*, Academic Press, London, United Kingdom. 411 p.

Lenth RV. 2007. Statistical Power Calculations, *Journal of Animal Science*, 85(E. Suppl.):E24–E29.

Leon CD, and van Gestel CAM. 1994. Selection of a Set of Laboratory Ecotoxicity Tests for the Effects Assessment of Chemicals in Terrestrial Ecosystems – Discussion Paper, Department of Ecology and Ecotoxicology, Vrije Universiteit, Amsterdam, The Netherlands, Report No. D94004. 134 p.

Lofs-Holmin A. 1980. Measuring Growth of Earthworms as a Method of Testing Sublethal Toxicity of Pesticides – Experiments with Benomyl and Trichloroacetic Acid (TCA), *Swedish Journal of Agricultural Research*, 10:25–33.

Lofs-Holmin A, and Bostrom U. 1988. The Use of Earthworms and Other Soil Animals in Pesticide Testing, In: *Earthworms in Waste and Environmental Management*, Edwards CA and Neuhauser EF (eds.), Academic Publishing, The Hague, The Netherlands, pp. 303–313.

Macdonald DW. 1983. Predation on Earthworms by Terrestrial Vertebrates, In: *Earthworm Ecology: From Darwin to Vermiculture*, Stachell JE (ed.), Chapman & Hall, New York, New York, pp. 393–414.

McAlpine DF, and Reynolds JW. 1977. Terrestrial Oligochaeta of Some New Brunswick Caves with Remarks on Their Ecology, *Canadian Field Naturalist*, 91:360-366.

McCann J. 2004. Report on the Identification of Earthworm Test Cultures (*Eisenia andrei* and *E. fetida*) and Potential Commercial Sources of Earthworms, Report prepared for Environment Canada, Ottawa, Ontario.

McElroy TC, and Diehl WJ. 2001. Heterosis in Two Closely Related Species of Earthworm (*Eisenia fetida* and *E. andrei*), *Heredity*, 87:598–608.

McGuirk B, Theron P, and Maboeta M. 2020. The Effects of Different Gold Mine Tailings on Growth, Reproduction and Avoidance-Behaviour of Earthworms, *African Zoology*, 55:35–42.

Mather JG, and Christensen OM. 1992. Surface Migration of Earthworms in Grassland, *Pedobiologia*, 36:51–57.

Meharg AA, Shore RF, and Broadgate K. 1998. Edaphic Factors Affecting the Toxicity and Accumulation of Arsenate in the Earthworm *Lumbricus terrestris*, *Environmental Toxicology and Chemistry*, 17:1124–1131.

Meier JR, Chang LW, Jacobs S, Torsella J, Meckes MC, and Smith MK. 1997. Use of Plant and Earthworm Bioassays to Evaluate Remediation of Soil from a Site Contaminated with Polychlorinated Biphenyls, *Environmental Toxicology and Chemistry*, 16:928–938.

Menzie CA, Burmaster DE, Freshman JS, and Callahan CA. 1992. Assessment of Methods for Estimating Ecological Risk in the Terrestrial Component: A Case Study at the Baird & McGuire Superfund Site in Holbrook, Massachusetts, *Environmental Toxicology and Chemistry*, 11:245–260.

MESI (Miller Environmental Sciences Inc.). 2014. Review of Technical Issues and Test Performance of Environment Canada's Test Method: Tests for Toxicity of Contaminated Soil to Earthworms (*Eisenia andrei*, *Eisenia fetida*, or *Lumbricus terrestris*), Report prepared for Environment Canada, Ottawa, Ontario. 60 p.

MESI. 2020. Revision and Amendment of Environment and Climate Change Canada's Earthworm Test Method (EPS 1/RM/43): Review and Recommendations, Report prepared for Environment and Climate Change Canada, Ottawa, Ontario. 24 p.

Natal-da-Luz T, Römbke J, and Sousa JP. 2008. Avoidance Tests in Site-Specific Risk Assessment – Influence of Soil Properties on the Avoidance Response of Collembola and Earthworms, *Environmental Toxicology and Chemistry*, 27:1112–1117.

NERI (National Environmental Research Institute). 1993. Manual of SECOFASE: Development, Improvement and Standardization of Test Systems for Assessing Sublethal Effects on Chemicals on Fauna in the Soil Ecosystem, Løkke H and van Gestel CAM (eds.), Report from a Workshop held in Silkeborg January 18–19, 1993, Silkeborg, Denmark. 44 p.

Neuhauser EF, and Callahan CA. 1990. Growth and Reproduction of the Earthworm *Eisenia fetida fetida* Exposed to Sublethal Concentrations of Organic Chemicals, *Soil Biology and Biochemistry*, 22:175–179.

Neuhauser EF, Malecki MR, and Loehr RC. 1984. Growth and Reproduction of the Earthworm *E. fetida* after Exposure to Sublethal Concentrations of Metals, *Pedobiologia*, 27:89–97.

Newman MC. 2008. "What Exactly are You Inferring?" A Closer Look at Hypothesis Testing, *Environmental Toxicology and Chemistry*, 27:1013–1019.

Niemeyer JC, de Santo FB, Guerra N, Ricardo Filho AM, and Pech TM. 2018. Do Recommended Doses of Glyphosate-Based Herbicides Affect Soil Invertebrates? Field and Laboratory Screening Tests to Risk Assessment, *Chemosphere*, 198:154–160.

Norberg-King TJ. 1993. A Linear Interpolation Method for Sublethal Toxicity: the Inhibition Concentration (ICp) Approach (Version 2.0), U.S. Environmental Protection Agency, Environmental Research Laboratory, Duluth, Minnesota, Technical Report 03-93 of National Effluent Toxicity Assessment Center. 25 p.

OECD (Organisation for Economic Co-operation and Development). 1984. Test No. 207: Earthworm, Acute Toxicity Tests, OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris, France. 9 p.

OECD. 2000. OECD Guideline for the Testing of Chemicals: Proposal for a New Guideline – Earthworm Reproduction Test (*Eisenia fetida/andrei*), Draft document, January 2000. 17 p.

OECD. 2006. Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application, OECD Environment Health and Safety Publications Series on Testing and Assessment, No. 54, Paris, France. 147 p.

OECD. 2016. Test No. 222: Earthworm Reproduction Test (*Eisenia fetida/Eisenia andrei*), OECD Guideline for the Testing of Chemicals, Section 2, OECD Publishing, Paris, France. 21 p.

Øien N, and Stenersen J. 1984. Esterases of Earthworms – III. Electrophoresis Reveals that *Eisenia fetida* (Sav.) is Two Species, *Comparative Biochemistry and Physiology*, 78:277–282.

OMEE (Ontario Ministry of Environment and Energy). 1996. Guidance on Sampling and Analytical Methods for Use at Contaminated Sites in Ontario, Version 1.1, Queen's Printer, Toronto, Ontario. 162 p.

Pereira PCG, Soares LOS, Júnior SFS, Saggioro EM, and Correia FV. 2020. Sub-Lethal Effects of the Pesticide Imazalil on the Earthworm *Eisenia andrei*: Reproduction, Cytotoxicity, and Oxidative Stress, *Environmental Science and Pollution Research*, 27:33474–33485.

Plytycz B, Lis-Molenda U, Cygal M, Kielbasa E, Grebosz A, Duchnowski M, Andre J, and Morgan AJ. 2009. Riboflavin Content of Coelomocytes in Earthworm (*Dendrodrilus rubidus*) Field Populations as a Molecular Biomarker of Soil Metal Pollution, *Environmental Pollution*, 157:3042–3050.

Princz JI, Moody M, Fraser C, Van der Vliet L, Lemieux H, Scroggins R, and Siciliano SD. 2012. Evaluation of a New Battery of Toxicity Tests for Boreal Forest Soils: Assessment of the Impact of Hydrocarbons and Salts, *Environmental Toxicology and Chemistry*, 31:766–777.

Princz J, Becker L, Scheffczyk A, Stephenson G, Scroggins R, Moser T, and Römbke J. 2017. Ecotoxicity of Boric Acid in Standard Laboratory Tests with Plants and Soil Organisms, *Ecotoxicology*, 26:471–481.

Prodana M, Silva C, Gravato C, Verheijen FGA, Keizer JJ, Soares AMVM, Loureiro S, and Bastos AC. 2019. Influence of Biochar Particle Size on Biota Responses, *Ecotoxicology and Environmental Safety*, 174:120–128.

Puurtinen HM, and Martikainen EAT. 1997. Effect of Soil Moisture on Pesticide Toxicity to an Enchytraeid Worm, *Enchytraeus* sp., *Archives of Environmental Contamination and Toxicology*, 33:34–41.

Ramires MF, de Souza EL, de Castro Vasconcelos M, Clasen BE, Fontanive DE, Bianchetto R, Grasel Cezimbra JC, and Antoniolli ZI. 2020. Enzyme Assays and Toxicity of Pig Abattoir Waste in *Eisenia andrei*, *Environmental Pollution*, 260:113928.

Reinecke AJ, and Viljoen AJ. 1991. A Comparison of the Biology of *Eisenia fetida* and *Eisenia andrei* (Oligochaeta), *Biology and Fertility of Soils*, 11:295–300.

Reinecke AJ, Viljoen AJ, and Saayman RJ. 1992. The Suitability of *Eudrilus eugeniae, Perionyx excavatus* and *Eisenia fetida* (Oligochaeta) for Vermicomposting in Southern Africa in Terms of their Temperature Requirements, *Soil Biology and Biochemistry*, 24:1295–1307.

Reinecke AJ, and Reinecke SA. 1996. The Influence of Heavy Metals on the Growth and Reproduction of the Compost Worm *Eisenia fetida* (Oligocaheta), *Pedobiologia*, 40:439–448.

Renaud M, Akeju T, Natal-da-Luz T, Leston S, Rosa J, Ramos F, Sousa JP, and Azevedo-Pereira HMVS. 2018. Effects of the Neonicotinoids Acetamiprid and Thiacloprid in Their Commercial Formulations on Soil Fauna, *Chemosphere*, 194:85–93.

Renoux AY, Zajdlik B, Stephenson GL, and Moulins J. 2013. Risk-Based Management of Site Soils Contaminated with a Mixture of Hazardous Substances: Methodological Approach and Case Study, *Human and Ecological Risk Assessment: An International Journal*, 19:1127–1146.

Reynolds JW. 1977. The Earthworms (Lumbricidae and Sparganophilidae) of Ontario, Life Sciences Miscellaneous Publications, Royal Ontario Museum, Toronto, Ontario. 141 p.

Reynolds JW. 1994. The Distribution of the Earthworms (Oligochaeta) of Indiana: A Case for the Post Quaternary Introduction Theory for Megadrile Migration in North America, *Megadrilogica*, 5:13–32.

Reynolds JW. 1995. Status of Exotic Earthworm Systematics and Biogeography in North America, In: *Earthworm Ecology and Biogeography in North America*, Hendrix PF (ed.), Lewis Publishers, CRC Press, Boca Raton, Florida, pp. 1–28.

Reynolds JW. 2015. A Checklist of Earthworms (Oligochaeta: Lumbrididae and Megascolecidae) in Western and Northern Canada, *Megadrilogica*, 17:141–156.

Rico A, Sabater C, and Castillo MA. 2016. Lethal and Sub-Lethal Effects of Five Pesticides Used in Rice Farming on the Earthworm *Eisenia fetida*, *Ecotoxicology and Environmental Safety*, 127:222–229.

Ritchie EE, Princz JI, Robidoux PY, and Scroggins RP. 2013. Ecotoxicity of Xanthene Dyes and a Non-Chlorinated Bisphenol in Soil, *Chemosphere*, 90:2129–2135.

Ritchie E, Boyd P, Lawson-Halasz A, Hawari J, Saucier S, Scroggins R, and Princz J. 2017. The Ecotoxicity of Zinc and Zinc-Containing Substances in Soil with Consideration of Metal-Moiety Approaches and Organometal Complexes, *Environmental Toxicology and Chemistry*, 36:3324–3332.

Ritchie EE, Maisonneuve F, Scroggins RP, and Princz, JI. 2019. Lethal and Sublethal Toxicity of Thiamethoxam and Clothianidin Commercial Formulations to Soil Invertebrates in a Natural Soil, *Environmental Toxicology and Chemistry*, 38:2111–2120.

Riepert F, and Kula C. 1996. Development of Laboratory Methods for Testing Effects of Chemicals and Pesticides on Collembola and Earthworms, Mitteilungen aus der Biologischen Bundesantalt für Land- und Forstwirtschaft Berlin-Dahlem. 82 p.

Robidoux PY, Svendsen C, Caumartin J, Hawari J, Ampleman G, Thiboutot S, Weeks JM, and Sunahara GI. 2000. Chronic Toxicity of Energetic Compounds in Soil Determined Using the Earthworm (*Eisenia andrei*) Reproduction Test, *Environmental Toxicology and Chemistry*, 19:1764–1773.

Robidoux PY, Hawari J, Thiboutot S, Ampleman G, and Sunahara GI. 2001. Chronic Toxicity of Octahydro-1,3,5,7-Tetranitro-1,3,5,7-Tetrazocine (HMX) in Soil Determined Using the Earthworm (*Eisenia andrei*) Reproduction Test, *Environmental Pollution*, 111:283–292.

Robidoux PY, Sunahara GI, Savard K, Berthelot Y, Dodard S, Martel M, Gong P, and Hawari J. 2004. Acute and Chronic Toxicity of the New Explosive CL-20 to the Earthworm (*Eisenia andrei*) Exposed to Amended Natural Soils, *Environmental Toxicology and Chemistry*, 23:1026–1034.

Rocchini RJ, Clark MJR, Jordan AJ, Horvath S, McLeay DJ, Servizi JA, Sholund A, Singleton HJ, Watts RG, and Young RH. 1982. Provincial Guidelines and Laboratory Procedures for Measuring Acute Lethal Toxicity of Liquid Effluents to Fish, British Columbia Ministry of Environment, Victoria, British Columbia. 18 p.

Römbke J, Knacker T, Förster B, and Marcinkowski A. 1994. Comparison of Effects of Two Pesticides on Soil Organisms in Laboratory Tests, Microcosms, and in the Field, In: *Ecotoxicology of Soil Organisms*, Donker MH, Eijsackers H, and Heimbach F (eds.), Lewis Publishers, Chelsea, Michigan, pp. 229–240.

Römbke J, Jänsch S, and Scroggins R. 2006. Identification of Potential Organisms of Relevance to Canadian Boreal Forest and Northern Lands for Testing of Contaminated Soils, *Environmental Reviews*, 14:137–167.

Römbke J, Aira M, Backeljau T, Breugelmans K, Domínguez J, Funke E, Graf N, Hajibabaei M, Pérez-Losada M, Porto PG, Schmelz RM, Vierna J, Vizcaíno A, and Pfenniger M. 2016. DNA Barcoding of Earthworms (*Eisenia fetida*/*andrei* Complex) from 28 Ecotoxicological Test Laboratories, *Applied Soil Ecology*, 104:3–11.

Rundgren S, and Nilsson P. 1997. Sublethal Effects of Aluminium on Earthworms in Acid Soil: The Usefulness of *Dendrodrilus rubidus* (Sav.) in a Laboratory Test System, *Pedobiologia*, 41:417–436.

Sager JC and McFarlane C. 1997. Radiation, In: *Plant Growth Chamber Handbook*, Langhans RW and Tibbits TW (eds.), North Central Regional Research Publication No. 340, Iowa Agriculture and Home Economics Experiment Station Special Report No. 99, Iowa State University of Science and Technology, Ames, Iowa, pp. 1–30.

Saggioro EM, do Espírito Santo DG, Sales Júnior SF, Hauser-Davis RA, and Correia FV. 2019. Lethal and Sublethal Effects of Acetamiprid on *Eisenia andrei*: Behavior, Reproduction, Cytotoxicity and Oxidative Stress, *Ecotoxicology and Environmental Safety*, 183:109572.

Saterbak A, Toy RJ, Wong DCL, McMain BJ, Williams MP, Dorn PB, Brzuzy LP, Chai EY, and Salanitro JP. 1999. Ecotoxicological and Analytical Assessment of Hydrocarbon-Contaminated Soils and Application to Ecological Risk Assessment, *Environmental Toxicology and Chemistry*, 18:1591–1607.

Sauvé S, Cook N, Hendershot WH, and McBride MB 1996. Linking Plant Tissue Concentrations and Soil Copper Pools in Urban Contaminated Soils, *Environmental Pollution*, 94:153–157.

Sauvé S, Dumestre A, McBride MB, and Hendershot WH. 1998. Derivation of Soil Quality Criteria Using Predicted Chemical Speciation of Pb^{2+} and Cu^{2+}, *Environmental Toxicology and Chemistry*, 17:1481–1489.

Sauvé S, Norvell WA, McBride MB, and Hendershot WH. 2000. Speciation and Complexation of Cadmium in Extracted Soil Solutions, *Environmental Science & Technology*, 34:291–296.

Schaefer M. 2003. Behavioural Endpoints in Earthworm Ecotoxicology: Evaluation of Different Test Systems in Soil Toxicity Assessment, *Journal of Soils and Sediments*, 3:79–84.

Scheffczyk A, Frankenbach S, Jänsch S, and Römbke J. 2014. Comparison of the Effects of Zinc Nitrate-Tetrahydrate and Tributyltin-Oxide on the Reproduction and Avoidance Behavior of the Earthworm *Eisenia andrei* in Laboratory Tests Using Nine Soils, *Applied Soil Ecology*, 83:253–257.

Scheu S, and Falca M. 2000. The Soil Food Web of Two Beech Forests (*Fagus sylvatica*) of Contrasting Humus Type: Stable Isotope Analysis of a Macro- and a Mesofauna-Dominated Community, *Oecologia*, 123:285–296.

Scott-Fordsmand JJ, Weeks JM, and Hopkin SP. 2000. Importance of Contamination History for Understanding Toxicity of Copper to Earthworm *Eisenia fetida* (Oligochaeta: Annelida), Using Neutral-Red Retention Assay, *Environmental Toxicology and Chemistry*, 19:1774–1780.

Sheppard PS. 1988. Specific Differences in Cocoon and Hatchling Production in *Eisenia fetida* and *E. andrei*, In: *Earthworms in Waste and Environmental Management*, Edwards CA and Neuhauser EF (eds.), SPB Academic Publishing, The Hague, The Netherlands, pp. 83–92.

Sheppard SC, and Evenden WG. 1998. An Approach to Defining a Control or Diluent Soil for Ecotoxicity Assays, In: *Environmental Toxicology and Risk Assessment*, Little EE, DeLonay AJ, and Greenberg BM (eds.), 7th Vol, ASTM STP 1333, American Society for Testing and Materials, Philadelphia, Pennsylvania, pp. 215–226.

Sims RW, and Gerard BM. 1985. Earthworms: Keys and Notes for the Identification and Study of the Species – Synopses of the British Fauna (New Series), Kermack DM and Barnes RSK (eds.), Report No. 31, The Linnaean Society of London and The Estuarine and Brackish-Water Sciences Association, Brill EJ and Backhuys W, London, United Kingdom. 171 p.

Sims RW, and Gerard BM. 1999. Earthworms: Keys and Notes for the Identification and Study of the Species – Synopses of the British Fauna (New Series), Barnes RSK and Crothers JH (eds.), Report No. 31 (Revised), The Linnaean Society of London and the Estuarine and The Estuarine and Coastal Sciences Association, Field Studies Council, Shrewsbury, United Kingdom. 169 p.

Sivakumar S. 2015. Effects of Metals on Earthworm Life Cycles: A Review, *Environmental Monitoring and Assessment*, 187:530.

Slimak KM. 1997. Avoidance Response as a Sublethal Effect of Pesticides on *Lumbricus terrestris* (Oligochaeta), *Soil Biology and Biochemistry*, 29:713–715.

SPAC (Soil and Plant Analysis Council Inc.). 1992. *Soil Analysis Handbook of Reference Methods*, Soil and Plant Analysis Council Inc., Georgia University Station, Athens, Georgia, 202 p.

Spurgeon DJ, and Hopkin SP. 1995. Extrapolation of the Laboratory-Based OECD Earthworm Toxicity Test to Metal-Contaminated Field Sites, *Ecotoxicology*, 4:190–205.

Spurgeon DJ, and Hopkin SP. 1996a. Effects of Metal-Contaminated Soils on the Growth, Sexual Development, and Early Cocoon Production of the Earthworm *Eisenia fetida*, with Particular Reference to Zinc, *Ecotoxicology and Environmental Safety*, 35:86–95.

Spurgeon DJ, and Hopkin SP. 1996b. Effects of Variations of the Organic Matter Content and pH of Soils on the Availability and Toxicity of Zinc to the Earthworm *Eisenia fetida*, *Pedobiologia*, 40:80–96.

Spurgeon DJ, Hopkin SP, and Jones DT. 1994. Effects of Cadmium, Copper, Lead and Zinc on Growth, Reproduction and Survival of the Earthworm *Eisenia fetida* (Savigny): Assessing the Environmental Impact of Point-Source Metal Contamination in Terrestrial Ecosystems, *Environmental Pollution*, 84:123–130.

Stantec, and Aquaterra Environmental. 2004. Developmental Studies in Support of Environment Canada's Biological Test Methods for Measuring Soil Toxicity Using Earthworms, Report prepared for Environment Canada, Ottawa, Ontario.

Stephenson GL. 2002. Unpublished Data. Ecological Services Group International Inc., Guelph, Ontario.

Stephenson GL. 2003a. Terrestrial Test Methods for Plants and Soil Invertebrates, Ph.D. Thesis, Department of Environmental Biology, University of Guelph, Guelph, Ontario. 230 p.

Stephenson GL. 2003b. Unpublished Data. Stantec Consulting, Ltd., Guelph, Ontario.

Stephenson GL, Wren CD, Middelraad ICJ, and Warner JE. 1997. Exposure of the Earthworm, *Lumbricus terrestris*, to Diazinon, and the Relative Risk to Passerine Birds, *Soil Biology and Biochemistry*, 29:717–720.

Stephenson GL, Kaushik A, Kaushik NK, Solomon KR, Steele T, and Scroggins RP. 1998. Use of an Avoidance-Response Test to Assess the Toxicity of Contaminated Soils to Earthworms, In: *Advances in Earthworm Ecotoxicology*, Sheppard S, Bembridge J, Holmstrup M, and Posthuma L (eds.), Society of Environmental Toxicology and Chemistry, SETAC Press, Pensacola, Florida, pp. 67–81.

Stephenson GL, Koper N, McCann J, and Wang Z. 1999a. Draft Report on the Acute Screening and Definitive, Chronic Toxicity Tests with Whole Federated Crude Oil, Unpublished (Draft) Report prepared for Petroleum Technology Alliance Canada (Calgary, Alberta) and the Canadian Association of Petroleum Producers (Calgary, Alberta). 52 p. + app.

Stephenson GL, Koper N, McCann J, and Wang Z. 1999b. Draft Report on the Acute Screening and Definitive, Chronic Toxicity Tests with Fraction 3 Derived from Federated Crude Oil, Unpublished (Draft) Report prepared for Petroleum Technology Alliance Canada (Calgary, Alberta) and the Canadian Association of Petroleum Producers (Calgary, Alberta) 48 p. + app.

Stephenson GL, McCann J, Jokuty P, and Wang Z, 2000a. Draft Report on the Acute Screening and Definitive, Chronic Toxicity Tests with Fraction 2 Derived from Federated Crude Oil, Unpublished (Draft) Report prepared for Petroleum Technology Alliance Canada (Calgary, Alberta) and the Canadian Association of Petroleum Producers (Calgary, Alberta). 32 p. + app.

Stephenson GL, Koper N, Atkinson GF, Solomon KR, and Scroggins RP. 2000b. Use of Nonlinear Regression Techniques for Describing Concentration-Response Relationships of Plant Species Exposed to Contaminated Site Soils, *Environmental Toxicology and Chemistry*, 19:2968–2981.

Stephenson GL, Princz JI, Koper N, and Miasek PG. 2001. Terrestrial Toxicity Testing with Volatile Substances, In: *Environmental Toxicology and Risk Assessment: Science, Policy, and Standardization – Implications for Environmental Decisions*, 10th Volume, Greenberg BM, Hull RN, Roberts MH Jr., and Gensemer RW (eds.), American Society for Testing and Materials, ASTM STP 1403, West Conshohocken, Pennsylvania, pp. 236–252.

Stephenson GL, Kuperman RG, Linder GL, and Visser S. 2002. Toxicity Tests for Assessing Contaminated Soils and Ground Water, In: *Environmental Analysis of Contaminated Sites*, Sunahara GI, Renoux AY, Thellen C, Gaudet CL, and Pilon A (eds.), Ecological and Environmental Toxicology Series, John Wiley & Sons Ltd., Chichester, United Kingdom, pp. 25–43.

Stephenson G, Feisthauer N, Bessie K, and Scroggins R. 2008. Terrestrial Toxicity Testing in Support of Site-specific Risk Assessments: An Integrative Tool for Contaminated Site Management, Proceedings of the Federal Contaminated Sites National Workshop, Vancouver, BC, April 28–May 2, 2008, hosted by the Real Property Institute of Canada.

Suzuki MM, Cooper EL, Eyambe GS, Goven AJ, Fitzpatrick LC, and Venables BJ. 1995. Polychlorinated Biphenyls (PCBs) Depress Allogeneic Natural Cytotoxicity by Earthworm Coelomocytes, *Environmental Toxicology and Chemistry*, 14:1697–1700.

Tatsi K, Shaw BJ, Hutchinson TH, and Handy RD. 2018. Copper Accumulation and Toxicity in Earthworms Exposed to CuO Nanomaterials: Effects of Particle Coating and Soil Ageing, *Ecotoxicology and Environmental Safety* 166:462–473.

Thursby GB, Heltshe J, and Scott KJ. 1997. Revised Approach to Toxicity Test Acceptability Criteria Using a Statistical Performance Assessment, *Environmental Toxicology and Chemistry*, 16:1322–1329.

Tomlin AD. 1995. Behaviour as a Source of Earthworm Susceptibility to Ecotoxicants, In: *Ecotoxicology of Earthworms*, Greig-Smith PW, Becker H, Edwards PJ, and Heimbach F (eds.), Intercept Ltd., Andover, United Kingdom, pp. 116–125.

TOXSTAT. 1996. "Version 3.5", Lincoln Research Associates, Inc., PO Box 4276, Bisbee, Arizona 85603, USA, phone 520 432-4092, email danlra@msn.com. [Programs on disk, with printed user's manual.]

Tsukamoto J, and Watanabe H. 1977. Influence of Temperature on Hatching and Growth of *Eisenia foetida* (Oligochaeta, Lumbricidae), *Pedobiologia*, 17:338–342.

USEPA (United States Environmental Protection Agency). 1989. Protocols for Short-term Toxicity Screening of Hazardous Waste Sites, Corvallis Environ. Research Lab., Corvallis, Oregon, Report EPA/600/3-88/029. 102 p.

USEPA. 1995. Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to West Coast Marine and Estuarine Organisms, Office of Research and Development, USEPA, Washington, DC, Report EPA/600/R-95/136. 661 p.

USEPA. 2002. Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Marine and Estuarine Organisms, 4th edition, Office of Water, USEPA, Washington, DC, Report EPA 821-R-02-014. 464 p.

USEPA. 2008. Guidance for the Use of Dilution-Water (Negative) and Solvent Controls in Statistical Data Analysis for Guideline Aquatic Toxicology Studies, Memo from Statistics Workgroup and Aquatic Biology Technical Team to Donald Brady, Director, Environmental Fate and Effects Division, Office of Pesticides Programs, Office of Chemical Safety and Pollution Prevention (OCSPP), Washington, DC.

USEPA. 2010. National Pollutant Discharge Elimination System Test of Significant Toxicity Implementation Document, Office of Wastewater Management, Washington, DC, Report EPA 833-R-10-003.

USEPA. 2012. Ecological Effects Test Guidelines, OCSPP 850.3100: Earthworm Subchronic Toxicity Test, Office of Pesticides Programs, Office of Chemical Safety and Pollution Prevention (OCSPP), Washington, DC, Report EPA 712-C-016. 19 p.

Uwizeyimana H, Wang M, Chen W, and Khan K. 2017. The Eco-Toxic Effects of Pesticide and Heavy Metal Mixtures towards Earthworms in Soil, *Environmental Toxicology and Pharmacology*, 55:20–29.

van Dam RA, Harford AJ and Warne M St J. 2012. Time to Get Off the Fence: The Need for Definitive International Guidance on Statistical Analysis of Ecotoxicity Data, Debate and Commentary, *Integrated Environmental Assessment and Management*, 8:242–245.

Van der Hoeven N. 1998. Power Analysis for the NOEC: What is the Probability of Detecting Small Toxic Effects on Three Different Species using the Appropriate Standardized Test Protocols?, *Ecotoxicology*, 7:355–361.

Van der Vliet L, Taylor LN, and Scroggins R. 2012. NOEC: Notable Oversight of Enlightened Canadians: A Response to van Dam *et al.* (2012), Letter to the Editor, *Integrated Environmental Assessment and Management*, 8:397–398.

van Gestel CAM. 1991. Earthworms in Ecotoxicology, Ph.D. Thesis, Utrecht University, Utrecht, The Netherlands. 197 p.

van Gestel CAM. 1992. Validation of Earthworm Toxicity Tests by Comparison with Field Studies: A Review on Benomyl, Carbendazim, Carbofuran and Carbaryl, *Ecotoxicology and Environmental Safety*, 23:221–236.

van Gestel CAM. 1997. Scientific Basis for Extrapolating Results from Soil Ecotoxicity Tests to Field Conditions and the Use of Bioassays, In: *Ecological Risk Assessment of Contaminants in Soil*, van Straalen NM and Løkke H (eds.), Chapman & Hall, London, United Kingdom, pp. 25–50.

van Gestel CAM. 2012. Soil Ecotoxicology: State of the Art and Future Directions, *Zookeys*, 176:275–296.

van Gestel CAM, and van Dis WA. 1988. The Influence of Soil Characteristics on the Toxicity of Four Chemicals to the Earthworm *Eisenia fetida andrei* (Oligochaeta), *Biology and Fertility of Soils*, 6:262–265.

van Gestel CAM, van Dis WA, van Breemen EM, and Sparenburg PM. 1988. Comparison of Two Methods for Determining the Viability of Cocoons Produced in Earthworm Toxicity Experiments, *Pedobiologia*, 32:367–371.

van Gestel CAM, van Dis WA, van Breemen EM, and Sparenburg PM. 1989. Development of a Standardized Reproduction Toxicity Test with the Earthworm Species *Eisenia fetida andrei* Using Copper, Pentachlorophenol, and 2,4-Dichloroaniline, *Ecotoxicology and Environmental Safety*, 18:305–312.

van Gestel CAM, Dirven-van Breemen EM, and Baerselman R. 1992a. Influence of Environmental Conditions on the Growth and Reproduction of the Earthworm *Eisenia andrei* in an Artificial Soil Substrate, *Pedobiologia*, 36:109–120.

van Gestel CAM, Dirven-van Breemen EM, Baerselman R, Emans HJB, Janssen JAM, Postuma R, and van Vliet PJM. 1992b. Comparison of Sublethal and Lethal Criteria for Nine Different Chemicals in Standardized Toxicity Tests Using the Earthworm *Eisenia andrei*, *Ecotoxicology and Environmental Safety*, 23:206–220.

Vasseur P, and Bonnard M. 2014. Ecogenotoxicology in Earthworms: A Review, *Current Zoology*, 60:255–272.

Velicogna J, Ritchie E, Princz J, Lessard ME, and Scroggins R. 2012. Ecotoxicity of Siloxane D5 in Soil, *Chemosphere*, 87:77–83.

Velicogna JR, Ritchie EE, Scroggins RP, and Princz JI. 2016. A Comparison of the Effects of Silver Nanoparticles and Silver Nitrate on a Suite of Soil Dwelling Organisms in Two Field Soils, *Nanotoxicology*, 10:1144–1151.

Velicogna JR, Schwertfeger DM, Beer C, Jesmer AH, Kuo J, Chen H, Scroggins RP, and Princz JI. 2020. Phytotoxicity of Copper Oxide Nanoparticles in Soil with and without Biosolid Amendment, *NanoImpact*, 17:100196.

Venter JM, and Reinecke AJ. 1988. Sublethal Ecotoxicological Effects of Dieldrin on the Earthworm *Eisenia foetida* (Oligochaeta), In: *Earthworms in Waste and Environmental Management*, Edwards CA and Neuhauser EF (eds.), SPB Academic Publishing, The Hague, The Netherlands, pp. 337–353.

Wallwork JA. 1983. *Earthworm Biology*, Studies in Biology No. 161, Camelot Press Ltd., Southampton, United Kingdom. 58 p.

Wells JB, and Lanno RP. 2001. Passive Sampling Devices (PSDs) as Biological Surrogates for Estimating the Bioavailability of Organic Chemicals in Soil, In: *Environmental Toxicology and Risk Assessment: Science, Policy, and Standardization – Implications for Environmental Decisions*, 10th Volume, Greenberg BM, Hull RN, Roberts Jr MH, and Gensemer RW (eds.), American Society for Testing and Materials, ASTM STP 1403, West Conshohocken, Pennsylvania, pp. 253–270.

Wentsel RS, and Guelta MA. 1988. Avoidance of Brass Powder-contaminated Soil by the Earthworm, *Lumbricus terrestris*, *Environmental Toxicology and Chemistry*, 7:241–243.

Williams DA. 1972. The Comparison of Several Dose Levels with a Zero Dose Control, *Biometrics*, 28:519–532.

Yeardley RB Jr., Lazorchak JM, and Gast LC. 1996. The Potential of an Earthworm Avoidance Test for Evaluation of Hazardous Waste Sites, *Environmental Toxicology and Chemistry*, 15:1532–1537.

Zoran MJ, Heppner TJ, and Drewes CD. 1986. Teratogenic Effects of the Fungicide Benomyl on Posterior Segmental Regeneration in the Earthworm, *Eisenia fetida*, *Pesticide Science*, 17:641–652.

Appendix A

Biological Test Methods and Supporting Guidance Documents Published by Environment and Climate Change Canada's Method Development and Applications Unit[a]

Title of Biological Test Method or Guidance Document	Report Number	Publication Date	Applicable Amendments
A. Generic (Universal) Biological Test Methods			
Acute Lethality Test Using Rainbow Trout	EPS 1/RM/9	July 1990	May 1996 and May 2007
Acute Lethality Test Using *Daphnia* spp.	EPS 1/RM/11	July 1990	May 1996
Test of Reproduction and Survival Using the Cladoceran *Ceriodaphnia dubia*	EPS 1/RM/21 2nd Edition	February 2007	–
Test of Larval Growth and Survival Using Fathead Minnows	EPS 1/RM/22 2nd Edition	February 2011	–
Toxicity Test Using Luminescent Bacteria	EPS 1/RM/24	November 1992	–
Growth Inhibition Test Using a Freshwater Alga	EPS 1/RM/25 2nd Edition	March 2007	–
Acute Test for Sediment Toxicity Using Marine or Estuarine Amphipods	EPS 1/RM/26	December 1992	October 1998
Fertilization Assay Using Echinoids (Sea Urchins and Sand Dollars)	EPS 1/RM/27 2nd Edition	February 2011	–
Toxicity Tests Using Early Life Stages of Salmonid Fish (Rainbow Trout)	EPS 1/RM/28 2nd Edition	July 1998	–
Test for Survival and Growth in Sediment Using Larvae of Freshwater Midges (*Chironomus tentans* or *Chironomus riparius*)	EPS 1/RM/32	December 1997	–
Test for Survival, Growth and Reproduction in Sediment and Water Using the Freshwater Amphipod *Hyalella azteca*	EPS 1/RM/33 3rd Edition	September 2017	–

[a] These documents are available for purchase from the Publication Catalogue, Environment and Climate Change Canada, Ottawa ON K1A 0H3, Canada. Printed copies can also be requested by email from methods@ec.gc.ca. These documents are available free of charge in electronic format at the following website: https://www.canada.ca/en/environment-climate-change/services/wildlife-research-landscape-science/biological-test-method-publications.html. For further information or comments, contact the Manager, Method Development and Applications Unit, Environment and Climate Change Canada, Ottawa ON K1A 0H3.

Title of Biological Test Method or Guidance Document	Report Number	Publication Date	Applicable Amendments
A. Generic (Universal) Biological Test Methods (continued)			
Test for Measuring the Inhibition of Growth Using the Freshwater Macrophyte, *Lemna minor*	EPS 1/RM/37 2nd Edition	January 2007	–
Test for Survival and Growth in Sediment Using Spionid Polychaete Worms (*Polydora cornuta*)	EPS 1/RM/41	December 2001	–
Tests for Measuring Avoidance Behaviour or Reproduction of Earthworms (*Eisenia andrei* or *Dendrodrilus rubidus*) Exposed to Contaminants in Soil	STB 1/RM/43 2nd Edition	August 2022	–
Test for Measuring Emergence and Growth of Terrestrial Plants Exposed to Contaminants in Soil	EPS 1/RM/45	February 2005	June 2007
Test for Measuring Survival and Reproduction of Springtails Exposed to Contaminants in Soil	EPS 1/RM/47 2nd Edition	February 2014	–
Test for Growth in Contaminated Soil Using Terrestrial Plants Native to the Boreal Region	EPS 1/RM/56	August 2013	–
Test for Measuring Reproduction of Oribatid Mites Exposed to Contaminants in Soil	STB 1/RM/61	September 2020	–
B. Reference Methods[b]			
Reference Method for Determining Acute Lethality Using Threespine Stickleback	EPS 1/RM/10 2nd Edition	December 2017	–
Reference Method for Determining Acute Lethality of Effluents to Rainbow Trout	EPS 1/RM/13 2nd Edition	December 2000	May 2007 and February 2016
Reference Method for Determining Acute Lethality of Effluents to *Daphnia magna*	EPS 1/RM/14 2nd Edition	December 2000	February 2016
Reference Method for Determining Acute Lethality of Sediment to Marine or Estuarine Amphipods	EPS 1/RM/35	December 1998	–
Reference Method for Determining the Toxicity of Sediment Using Luminescent Bacteria in a Solid-Phase Test	EPS 1/RM/42	April 2002	–
Reference Method for Measuring the Toxicity of Contaminated Sediment to Embryos and Larvae of Echinoids (Sea Urchins or Sand Dollars)	EPS 1/RM/58	July 2014	–
Reference Method for Determining Acute Lethality Using *Acartia tonsa*	STB 1/RM/60	June 2019	–

[b] For this series of documents, a reference method is defined as a specific biological test method for performing a toxicity test, i.e., a toxicity test method with an explicit set of test instructions and conditions that is described precisely in a written document. Unlike other generic (multipurpose or "universal") biological test methods published by Environment and Climate Change Canada, the use of a reference method is frequently restricted to testing requirements associated with specific regulations.

Title of Biological Test Method or Guidance Document	Report Number	Publication Date	Applicable Amendments
C. Supporting Guidance Documents			
Guidance Document on Control of Toxicity Test Precision Using Reference Toxicants	EPS 1/RM/12	August 1990	–
Guidance Document on Collection and Preparation of Sediments for Physicochemical Characterization and Biological Testing	EPS 1/RM/29	December 1994	–
Guidance Document on Measurement of Toxicity Test Precision Using Control Sediments Spiked with a Reference Toxicant	EPS 1/RM/30	September 1995	–
Guidance Document on Application and Interpretation of Single-species Tests in Environmental Toxicology	EPS 1/RM/34	December 1999	–
Guidance Document for Testing the Pathogenicity and Toxicity of New Microbial Substances to Aquatic and Terrestrial Organisms	EPS 1/RM/44 2nd Edition	December 2016	–
Guidance Document on Statistical Methods for Environmental Toxicity Tests	EPS 1/RM/46	March 2005	June 2007
Procedure for pH Stabilization During the Testing of Acute Lethality of Wastewater Effluent to Rainbow Trout	EPS 1/RM/50	March 2008	–
Guidance Document on the Sampling and Preparation of Contaminated Soil for Use in Biological Testing	EPS 1/RM/53	February 2012	–
Procedure for pH Stabilization During the Testing of Acute Lethality of Pulp and Paper Effluent to Rainbow Trout	STB 1/RM/59	March 2018	–
Recommended Procedure for the Importation of Test Organisms for Sublethal Toxicity Testing	–	September 1999	–
Revised Procedures for Adjusting Salinity of Effluent Samples for Marine Sublethal Toxicity Testing Conducted under Environmental Effects Monitoring (EEM) Programs	–	December 2001	–
Supplementary Background and Guidance for Investigating Acute Lethality of Wastewater Effluent to Rainbow Trout	–	March 2008	–
Supplementary Guidance for Investigating Acute Lethality of Pulp and Paper Mill Effluents due to Ammonia	–	March 2018	–

Appendix B

Members of the Inter-Governmental Ecotoxicological Testing Group (as of September 2021)

Federal, Environment and Climate Change Canada

Suzanne Agius
Marine Protection Programs Section
Gatineau, Quebec

Adrienne Bartlett
Aquatic Contaminants Research Division
Burlington, Ontario

Lee Beaudette
Wildlife Toxicology Research
Ottawa, Ontario

Rene Beaulieu
Prairie & Northern Laboratory for Environmental Testing
Edmonton, Alberta

Christian Blaise (Emeritus)
Centre St. Laurent
Montréal, Quebec

Patrick Boyd
Biological Assessment & Standardization Section
Ottawa, Ontario

Lorraine Brown
Pacific & Yukon Laboratory for Environmental Testing
North Vancouver, British Columbia

Joy Bruno
Pacific & Yukon Laboratory for Environmental Testing
North Vancouver, British Columbia

Julia Brydon
Marine Protection Programs Section
Gatineau, Quebec

Craig Buday
Pacific & Yukon Laboratory for Environmental Testing
North Vancouver, British Columbia

Melanie Camplin
Prairie & Northern Laboratory for Environmental Testing
Edmonton, Alberta

Marshneil Chandra
Prairie & Northern Laboratory for Environmental Testing
Edmonton, Alberta

Ajith Dias Samarajeewa
Biological Assessment & Standardization Section
Ottawa, Ontario

Heather Dillon
Prairie & Northern Laboratory for Environmental Testing
Edmonton, Alberta

Ken Doe (Emeritus)
Atlantic Laboratory for Environmental Testing
Moncton, New Brunswick

Tamzin El-Fityani
National Guidelines and Standards Office
Ottawa, Ontario

Richard Frank
Aquatic Contaminants Research Division
Burlington, Ontario

François Gagné
Fluvial Ecosystem Research
Montréal, Quebec

Patricia Gillis
Aquatic Ecosystem Protection Research Division
Burlington, Ontario

Christina Heise
Prairie & Northern Laboratory for Environmental Testing
Edmonton, Alberta

Natasha Hostal
Prairie & Northern Laboratory for Environmental Testing
Edmonton, Alberta

Paula Jackman
Atlantic Laboratory for Environmental Testing
Moncton, New Brunswick

Stephanie Kvas
Biological Assessment & Standardization Section
Ottawa, Ontario

Christopher Le
Pacific & Yukon Laboratory for Environmental Testing
North Vancouver, British Columbia

Heather Lemieux
Biological Assessment & Standardization Section
Ottawa, Ontario

Michelle Linssen-Sauvé
Pacific & Yukon Laboratory for Environmental Testing
North Vancouver, British Columbia

Carolyn Martinko
Biological Assessment & Standardization Section
Ottawa, Ontario

Danielle Milani
Aquatic Ecosystem Impacts Research Division
Burlington, Ontario

Rachel Miliano
Pacific & Yukon Laboratory for Environmental Testing
North Vancouver, British Columbia

Joanne Parrott
Aquatic Ecosystem Protection Research Division
Burlington, Ontario

Linda Porebski
Marine Protection Programs Section
Gatineau, Quebec

Juliska Princz
Biological Assessment & Standardization Section
Ottawa, Ontario
Rick Scroggins
Biological Assessment & Standardization Section
Ottawa, Ontario

David Taillefer
Marine Environmental Protection
Gatineau, Quebec

Sylvain Trottier
Quebec Laboratory for Environmental Testing
Montréal, Quebec

Graham van Aggelen
Pacific & Yukon Laboratory for Environmental Testing
North Vancouver, British Columbia

Leana Van der Vliet
Biological Assessment & Standardization Section
Ottawa, Ontario

Jessica Velicogna
Biological Assessment & Standardization Section
Ottawa, Ontario

Brian Walker
Quebec Laboratory for Environmental Testing
Montréal, Quebec

Peter Wells (Emeritus)
Environmental Conservation Service
Dartmouth, Nova Scotia

Federal, Natural Resources Canada

Philippa Huntsman-Mapila
Ecosystem Risk Management Program
Mining & Mineral Sciences Laboratory
CANMET, NRCan
Ottawa, Ontario

Morgan King
Ecosystem Risk Management Program
Mining & Mineral Sciences Laboratory
CANMET, NRCan
Ottawa, Ontario

Carrie Rickwood
Ecosystem Risk Management Program
Mining & Mineral Sciences Laboratory
CANMET, NRCan
Ottawa, Ontario

Provincial

Lisa Kennedy (co-Chair)
Ontario Ministry of Environment, Conservation and Parks
Etobicoke, Ontario

Jennifer Koene-Fenton
Ontario Ministry of Environment, Conservation and Parks
Etobicoke, Ontario

Jasen Nelson
British Columbia Ministry of the Environment and Climate Change Strategy
Victoria, British Columbia

Heather Osachoff
British Columbia Ministry of the Environment and Climate Change Strategy
Victoria, British Columbia

David Poirier (Emeritus)
Ontario Ministry of Environment, Conservation and Parks
Etobicoke, Ontario

Éloïse Veilleux
Centre d'expertise en analyse environnementale du Québec
Ste. Foy, Quebec

Trudy Watson-Leung (co-Chair)
Ontario Ministry of Environment, Conservation and Parks
Etobicoke, Ontario

Appendix C

Environment and Climate Change Canada, National Capital Region (NCR) and Regional Environmental Testing Laboratories

Soil Toxicology Laboratory
River Road S & T Branch Laboratories
335 River Road
Ottawa, Ontario
K1A 0H3

Atlantic Laboratory for Environmental Testing
Environmental Science Building
443 Université Avenue, Université de Moncton
Moncton, New Brunswick
E1A 3E9

Pacific and Yukon Laboratory for Environmental Testing
Pacific Environmental Science Centre
2645 Dollarton Hwy
North Vancouver, British Columbia
V7H 1B1

Québec Laboratory for Environmental Testing
105 McGill Street
Montréal, Quebec
H2Y 2E7

Prairie and Northern Laboratory for Environmental Testing
Northern Forestry Building
5320 122 St NW
Edmonton, Alberta
T6H 3S5

For current regional laboratory contact information please contact:

Method Development and Applications Unit
Science and Technology Branch
Environment and Climate Change Canada
335 River Road
Ottawa, Ontario
K1A 0H3
Email: methods@ec.gc.ca

Appendix D

Members of the Scientific Advisory Group for the First Edition Test Method Document

SAG Members

Mr. Christian Bastien
Centre d'expertise en analyse
environnementale du Québec
Ministère de l'Environnement
2700 Einstein
Saint-Foy, Quebec G1P 3W8
Phone: (418) 643-8225
Fax: (418) 643-9023
Email: christian.bastien@menv.gouv.qc.ca

Dr. Clive Edwards
Ohio State University
Department of Entomology
1735 Neil Avenue
Columbus, Ohio
USA 43210
Phone: (614) 292-3786
Fax: (614) 688-4222
Email: edwards.9@osu.edu

Dr. Roman G. Kuperman
U.S. Army Edgewood Chemical Biological Center
AMSSB-RRT-TE E5641 DR KUPERMAN
5183 Blackhawk Road
Aberdeen Proving Ground, Maryland
USA 21010-5424
Phone: (410) 436-4697
Fax: (410) 436-4846
Email: roman.kuperman@us.army.mil

Dr. Roman P. Lanno
Ohio State University
Department of Entomology
1735 Neil Avenue
Columbus, Ohio
USA 43210
Phone: (614) 292-4943
Fax: (614) 292-2180
Email: lanno.1@osu.edu

Dr. Frank Riepert
Biologische Bundesantalt fur Land- und
Forstwirtschaft (BBA)
Konigin-Luise-Str. 19
D-14195
Berlin, Germany
Phone: 0049 30 8304 2406
Fax: 0049 30 8304 2403
Email: f.riepert@bba.de

Dr. Jörg Römbke
ECT Oekotoxikologie GmbH
Boettgerstrasse 2-14
65439 Flörsheim am Main
Germany
Phone: 49 6145 95640
Fax: 49 6145 95649 9
Email: j-roembke@ect.de

Dr. Geoffrey Sunahara
National Research Council
Biotechnology Research Institute
6100 Royalmount Avenue
Montreal, Quebec H4P 2R2
Phone: (514) 496-8030
Fax: (514) 496-6265
Email: geoffrey.sunahara@nrc.ca

Mr. Graham van Aggelen
Environment Canada
Pacific Environmental Science Centre
2645 Dollarton Highway
North Vancouver, BC V7H 1B1
Phone: (604) 924-2513
Fax: (604) 924-2555
Email: graham.vanaggelen@ec.gc.ca

Dr. Kees van Gestel
Institute of Ecological Science
Vrije Universiteit Amsterdam
De Boelelaan 1087
1081 HV Amsterdam
The Netherlands
Phone: 31 20 444-7079/7004
Fax: 31 20 444-7123
Email: kees.van.gestel@vu.nl

Dr. Suzanne Visser
Department of Biological Sciences
University of Calgary
2500 University Drive NW
Calgary, Alberta T2N 1N4
Phone: (403) 220-6375
Fax: (403) 289-9311
Email: svisser@acs.ucalgary.ca

Scientific Authority

Mr. Rick Scroggins
Environment Canada
Biological Methods Division
Environmental Technology Centre
335 River Road
Ottawa, ON K1A 0H3
Phone: (613) 990-8569
Fax: (613) 990-0173
Email: rick.scroggins@ec.gc.ca

Consultants

Dr. Don McLeay
McLeay Environmental Ltd.
2999 Spring Bay Road
Victoria, BC V8N 5S4
Phone: (250) 472-2608
Fax: (250) 472-2609
Email: dmcleay@telus.net

Dr. Gladys Stephenson
Aquaterra Environmental Consulting Inc.
RR1, Site 5936
Orton, ON L0N 1N0
Phone: (519) 836-6050
Fax: (519) 836-2493
Email: gstephenson@stantec.com

Appendix E

Procedural Variations for Tests of Effects of Contaminated Soil on the Survival and Reproduction of Earthworms (*Eisenia andrei* and *Dendrodrilus rubidus*), as Described in International Methodology Documents

The following source documents are listed chronologically, by originating agency rather than by author(s).

EC, 2004b—is the first edition of Environment Canada's biological test method for measuring soil toxicity using a test for effects on reproduction and growth of *Eisenia andrei* or *E. fetida*, published in 2004, EPS 1/RM/43 (Ottawa, Canada).

ECCC, 2022—is the test method described herein for measuring soil toxicity using a test for effects on reproduction of *Eisenia andrei* or *Dendrodrilus rubidus*, EPS 1/RM/43, second edition (Ottawa, Canada).

ISO, 2012—is a standard guideline for assessing the effects of chemicals on the reproduction of the earthworm *Eisenia fetida* or *E. fetida andrei*, published in 2012 by the International Standardization Organisation (Geneva, Switzerland).

OECD, 2016—is a standard guideline for assessing the effects of chemicals on the reproduction of the earthworm *Eisenia fetida* or *E. fetida andrei*, published in 2016 by the Organisation for Economic Co-operation and Development (Paris, France).

Parameter	EC, 2004b	ECCC, 2022	ISO, 2012	OECD, 2016
Test type	• Whole soil static (i.e., non-renewal) toxicity test	• Whole soil static (i.e., non-renewal) toxicity test	• Whole soil static (i.e., non-renewal) toxicity test	• Whole soil static (i.e., non-renewal) toxicity test
Soil type	• Field-collected or chemical-spiked soil	• Field-collected or chemical-spiked soil	• Field-collected or chemical-spiked soil	• Chemical-spiked soil only
Test duration	• 56 or 63 days	• 56 days (63 days, in some cases)	• 56 days	• 56 days
Test organisms	• Laboratory-cultured *E. andrei* or *E. fetida*; sexually mature adults with clitellum and with wet wt of 250–600 mg • 2 adult earthworms per test vessel • Acclimated to negative control soil and test conditions for ≥ 7 days	• Laboratory-cultured *E. andrei* or *D. rubidus*; sexually mature adults with clitellum and with wet wt of 250–600 mg for *E. andrei*, and 50–200 mg for *D. rubidus* • 4 adult earthworms per test vessel • Acclimated to negative control soil and test conditions for ≥ 7 days	• Laboratory-cultured *E. andrei* or *E. fetida*; 2–12 months old; age-synchronized from cocoons; with clitellum and with wet wt of 250–600 mg for *E. andrei* • 10 adult earthworms per test vessel; each group of 10 adult worms is weighed prior to placement in vessel; the range of mean biomass between vessels should be ≤ 100 mg • Acclimated to negative control soil and food for 1–7 days	• Laboratory-cultured *E. andrei* or *E. fetida*; 2–12 months old; homogeneous age-structure, age-synchronized from cocoons; with clitellum and with wet wt of 250–600 mg for *E. andrei*; individuals do not differ in age by more than 4 weeks • 10 adult earthworms per test vessel; each group of 10 adult worms is weighed prior to placement in the vessel • Acclimated to negative control soil and food for ≥ 1 days
Negative control soil	• Inclusion as a treatment with each toxicity test • Natural clean field-collected or formulated artificial soil for which prior tests with the test species demonstrated that test validity could be regularly met • All field soil horizons are tested separately (no mixing or layering)	• Inclusion as a treatment with each toxicity test • Natural clean field-collected or formulated artificial soil for which prior tests with the test species demonstrated that test validity could be regularly met • All field soil horizons are tested separately (no mixing or layering)	• Inclusion as a treatment with each toxicity test • Natural clean field-collected or formulated artificial soil for which prior tests with the test species demonstrated that test validity could be regularly met	• Inclusion as a treatment with each toxicity test • Artificial soil • Natural soil may be used in additional (i.e., for higher tier testing) testing

Number of replicates	• 10 replicates per control and treatment		• 5 replicates per control and treatment for multi-concentration test • For single-concentration tests, depending on the target effect size (see Section 5.6.2): o a minimum of 13–21 replicates per control and treatment for *E. andrei*; and o a minimum of 5–13 replicates per control and treatment for *D. rubidus*		• 4 replicates per control and treatment for NOEC/LOEC and mixed (NOEC/LOEC and ECx) • 2 replicates for treatments and 6 replicates for controls for ECx		• 4 replicates per control and treatment for NOEC/LOEC and mixed (NOEC/LOEC and ECx) • 2 replicates for treatments and 6 replicates for controls for ECx • 8 replicates for limit test	
Number of treatments	Single-concentration • ≥ 1, plus negative control soil	Multi-concentration • ≥ 7, plus negative control soil	Single-concentration • ≥ 1, plus negative control soil	Multi-concentration • ≥ 7, plus negative control soil	Single-concentration • ≥ 1, plus negative control soil	Multi-concentration • ≥ 5, plus negative control soil for NOEC and mixed; ≥ 12 for ECx	Limit test • 1000 mg/kg, plus negative control soil	Multi-concentration • ≥ 5, plus negative control soil for NOEC; enough to produce ≥ 4 statistically significant different mean responses for ECx; 8, plus negative control for mixed
Test vessel	• 500-mL glass wide-mouth Mason jar covered with perforated translucent or transparent cover		• For *E. andrei*: 500-mL glass wide-mouth Mason jar covered with perforated cover • For *D. rubidus*: 250-mL glass wide-mouth Mason jar covered with 50-µm Nitex mesh held on with metal screw ring; a metal lid is placed on top of the mesh.		• 1–2 L with cross-sectional area of ~200 cm^2; covered to permit gas exchange and light		• 1–2 L with cross-sectional area of ~200 cm^2; inert material; covered to permit gas exchange and light	

Amount of soil per test unit	• Identical wet wt equivalent to a volume of 350 mL soil at optimal moisture content	• Identical wet wt equivalent to a volume of 350 mL soil for *E. andrei*, and 200 mL of soil for *D. rubidus*, at optimal moisture content	• 500 g dry mass • 5–6 cm depth	• 500 g dry mass • 5–6 cm depth
Moisture content	• Hydrate to the optimal moisture content • Maintain moisture by observation/squeeze test	• Hydrate to the optimal moisture content • Maintain moisture by weighing, observation, squeeze test	• Hydrate to 40–60% or higher of WHC • Maintain moisture by weighing containers periodically (<10% difference by test end)	• Hydrate to 40–60% of WHC • Maintain moisture by weighing containers periodically (<10% difference by test end)
Temperature	• Daily average, 20 ± 2 °C; instantaneous, 20 ± 3 °C	• Daily average, 20 ± 2 °C; instantaneous, 20 ± 3 °C	• Daily average, 20 ± 2 °C	• Daily average, 20 ± 2 °C
Lighting	• Incandescent or fluorescent with intensity of 400–800 lux; fixed photoperiod of 16 h light and 8 h dark, or 12 h light and 12 h dark	• Incandescent, fluorescent, or LED with intensity of 6.96–11.92 µmol/(m^2 · s); fixed photoperiod of 16 h light and 8 h dark, or 12 h light and 12 h dark	• Incandescent or fluorescent with intensity of 400–800 lux; fixed photoperiod of 16 h light and 8 h dark, or 12 h light and 12 h dark	• Incandescent or fluorescent with intensity of 400–800 lux; fixed photoperiod of 16 h light and 8 h dark
Feeding	• ~½ tsp cooled cooked oatmeal on Days 0, 14, 28, and 42 • Food is placed in a depression beneath the surface of the test soil	• For *E. andrei*: ~2 g or 1 g (if large amount of uneaten food remains) of oatmeal, MWF or organic mixed grains per test vessel, on Days 0, 14, 28, and 42 • For *D. rubidus*: ~1 g of MWF or organic mixed grains per test vessel, on Days 0, 14, 28, and 42 • Food is placed in a depression beneath the surface of the test soil	• 5 g air dried finely ground food (e.g., oatmeal, mashed potato powder, cow or horse manure) per test vessel, on Day 1, and once per week thereafter, and Day 28 after adults are removed • Food shown to not affect growth and cocoon production • Food is placed on surface and moistened with potable water (5–6 mL per container); reduce feeding if consumption is low • Juveniles are fed once (Day 28) with 5 g of food gently mixed into substrate	• 5 g air dried finely ground food (e.g., oatmeal, mashed potato powder, cow or horse manure) per test vessel, on Day 1, and once per week, and Day 28 after adults are removed • Food shown to not affect growth and cocoon production • Food is placed on surface and moistened with potable water (5–6 mL per container); reduce feeding if consumption is low • Juveniles are fed once (Day 28) with 5 g of food gently mixed into substrate

Measurements and observations during test	• Daily temperature in environmentally-controlled incubator • pH and moisture content at beginning and end of test • Presence of uneaten food and qualitative moisture level at each feeding • Light intensity at least once during the test	• Daily temperature in environmentally-controlled incubator • pH and moisture content at beginning and end of test • Weekly test vessel weights • Presence of uneaten food and qualitative moisture level at each feeding • Light intensity at least once during the test	• pH and moisture content at beginning and end of test • Presence of uneaten food	• pH and moisture content at beginning and end of test • Presence of uneaten food
Counting juveniles	• Heat extraction (40–45 °C water bath for 15 minutes) • Hand sorting	• Hand sorting • Heat extraction (40–45 °C water bath for 15 minutes) offered as alternative	• Heat extraction (50–60 °C water bath for 20 minutes); efficiency of the method is checked • Sieving (two 0.5-mm sieves) • Hand sorted twice	• Heat extraction (50–60 °C water bath for 20 minutes); efficiency of the method is checked • Sieving (two 0.5-mm sieves) • Hand sorted twice
Test validity in controls	• Survival must be ≥ 90% at Day 28 or Day 35 • Reproduction ≥ 3 live juveniles per adult at test end • Mean dry wt of individual live juveniles is ≥ 0.2 mg	• Survival must be ≥ 90% at Day 28 • Reproduction ≥ 3 live juveniles per adult at test end (56 days)	• Survival must be ≥ 90% • Reproduction ≥ 30 live juveniles per replicate • CV of reproduction ≤ 30%	• Survival must be ≥ 90% • Reproduction ≥ 30 live juveniles per replicate • CV of reproduction ≤ 30%
Biological endpoints	• Number of live adults and cocoons (optional) in each test vessel on Day 28 or 35, with observations of adult health • Wet wt of selected organisms (subset of 20 worms) on day of organism addition and on day of adult removal (Day 28 or 35)	• Number of live adults and cocoons (optional) in each test vessel on Day 28 with observations of adult health • Wet wt of selected organisms (subset of 20 worms) on day of organism addition and on day of adult removal (Day 28) • Number of live juveniles and cocoons (optional) in each vessel on Day 56	• Number and mass of live adults on Day 28 • Number of live juveniles and cocoons on Day 56	• Number and mass of live adults on Day 28 • Number of live juveniles and cocoons on Day 56

Biological endpoints (continued)	• Number of live juveniles and cocoons (optional) in each vessel on Day 56 or Day 63 • Individual juvenile dry mass after Day 56 or 63			
Statistical endpoints	Single-Concentration: • Mean percent survival (± SD) of adults in each treatment on Day 28 or 35 • Mean (± SD) number of live juveniles in each treatment on Day 56 or 63 • Mean (± SD) dry wt of individual live juveniles in each treatment on Day 56 or 63 Multi-Concentration: • Mean percent survival (± SD) of adults in each treatment on Day 28 or 35 • Mean (± SD) number of live juveniles in each treatment on Day 56 or 63 • 28-d or 35-d LC50 and 95% confidence limits for adult survival, data permitting • 56-d or 63-d ICps and 95% confidence limits for number of live juveniles and mean dry wt of juveniles	Single-Concentration: • Mean percent survival (± SD) of adults in each treatment on Day 28 or 35 • Mean (± SD) number of live juveniles in each treatment on Day 56 or 63 Multi-Concentration: • Mean percent survival (± SD) of adults in each treatment on Day 28 or 35 • Mean (± SD) number of live juveniles in each treatment on Day 56 or 63 • 28-d or 35-d LC50 and 95% confidence limits for adult survival, data permitting • 56-d or 63-d ICps and 95% confidence limits for number of live juveniles	Single-Concentration: • Mean percent survival (± SD) of adults in each treatment on Day 28 • Mean percent loss/increase in biomass[a] (± SD) of adults in each treatment on Day 28 • Mean (± SD) number of live juveniles in each treatment on Day 56 • ANOVA Multi-Concentration: • Mean percent survival (± SD) of adults in each treatment on Day 28 • Mean percent loss/increase in biomass[1] (± SD) of adults in each treatment on Day 28 • Mean (± SD) number of live juveniles in each treatment on Day 56 • ECx for reduction of number of live juveniles • NOEC/LOEC for lethality or mass alteration of adults, or reduction of reproduction	Limit test: • NOEC for reproduction Multi-Concentration: • Mean (± SD) number of live juveniles in each treatment on Day 56 • 28-d LC50 and 95% confidence limits for adult survival • ECx or NOEC/LOEC for reduction of number of live juveniles

[a] Initial wet mass of each group of 10 worms introduced into every test vessel is measured and used for determination of change in mass of adults at Day 28.

Appendix F

Natural and Artificial Negative Control Soils Used for Method Development and the Establishment of Test Validity Criteria

Negative control soil must be included as one of the experimental treatments in each soil toxicity test. This treatment requires a soil that is essentially free of any contaminants that could adversely affect the performance of test organisms during the test (see Section 3.3). Before applying either of the test methods described in this document as a standardized test to be conducted according to Environment and Climate Change Canada, it was necessary to first assess the performance of test organisms in different types of negative control soil representative of an array of clean soils found within Canada. Five types of negative control soils were used to develop the biological test methods described in the first edition of this test method document and to further assess the robustness of each test method with samples of soil that varied considerably in their physical and chemical characteristics. These soils were also used to establish reasonable criteria for valid test results, based on control performance. The five soils tested include an artificial soil (see Section 3.3.2) and four natural soils (see Section 3.3.1) (Aquaterra Environmental, 1998; Stephenson *et al.*, 1999a, 1999b, 2000a; Aquaterra Environmental and ESG, 2000; ESG, 2001, 2002; ESG and Aquaterra Environmental, 2002; Stantec and Aquaterra Environmental, 2004). The artificial soil was formulated in the laboratory from natural ingredients. The four natural soils included two agricultural soils from southern Ontario, a prairie soil from Alberta, and a forest soil from northern Ontario. The physicochemical characteristics of all five soils are summarized in Table F-1.

The artificial control soil (AS) used in this series of performance evaluation studies with diverse soil types was the same as that recommended for use herein (see Section 3.3.2). It consists of 70% silica sand, 20% kaolin clay, 10% *Sphagnum* sp. peat, and calcium carbonate (10–30 g per 1 kg peat). The soil was formulated by mixing the ingredients in their dry form thoroughly, then gradually hydrating with deionized water, and mixing further until the soil was visibly uniform in colour, texture, and degree of wetness. This artificial soil is much the same as that described by OECD (2016) and ISO (2012).

The four natural soils used as negative control soil while developing these biological test methods and establishing the test validity criteria herein (see Sections 4.2.3 and 4.3.3) do not represent all Canadian soil types. However, they do vary greatly in their physicochemical characteristics and include agricultural soils with diverse textures as well as a forest soil (see Table F-1). The soils originated from areas that had not been subjected to any direct application of pesticides in recent years. They were collected with either a shovel or a backhoe, depending on the location and the amount of soil collected. Sampling depth depended on the nature of the soil and the site itself.

The sample of clay loam soil, classified as a Delacour Orthic Black Chernozem, was collected in May 1995 from an undeveloped road allowance east of Calgary, Alberta. The soil beneath the sod was air dried to about 10–20% moisture content, sieved (4 or 9 mm), placed into 20-L plastic pails, and shipped to the University of Guelph (Guelph, ON) where it was kept in cold storage (4 °C) until needed. The soil was determined to be virtually free of any contaminants (Komex International, 1995). The physicochemical characteristics of the soil show that it is a moderate-to-fine clay loam, with a relatively high organic content and cation exchange capacity compared to the other clean soils used during the development of the first edition of these biological test methods and the establishment of test validity criteria (see Table F-1).

Table F-1 Physicochemical Characteristics of Candidate Artificial and Natural Negative Control Soils [a]

Parameter	Artificial Soil	Clay Loam	Sandy Loam	Silt Loam	Forest Soil	Analytical Method
Source	formulated from constituents	field-collected from Alberta	field-collected from Ontario	field-collected from Ontario	field-collected from Ontario	—
Soil Texture	Fine Sandy Loam	Clay Loam	Fine Sandy Loam	Silt Loam	Loam	as per Hausenbuiller (1985); based on grain size distribution
Sand (%)	77.3	26.6	60.8	36.6	48.6	gravimetric grain size distribution
Silt (%)	7.8	43.3	27.8	50.1	36.9	gravimetric grain size distribution
Clay (%)	14.9	30.1	11.4	13.3	14.5	gravimetric grain size distribution
Gravel (%)	—[b]	—	0	0	0	gravimetric grain size distribution
Very Coarse Sand (%)	—	—	1.5	1.2	0.6	gravimetric grain size distribution
Coarse Sand (%)	—	—	3.2	2.3	2.2	gravimetric grain size distribution
Medium Sand (%)	—	—	10.1	5.4	9	gravimetric grain size distribution
Fine Sand (%)	—	—	25.9	13.4	20.4	gravimetric grain size distribution
Very Fine Sand (%)	—	—	20.2	14.3	16.4	gravimetric grain size distribution
Water-holding capacity (%)	71.5	80.3	44	56.5	75.6	gravimetric analysis[c]
pH (units)	6	5.9	7.3	7.4	4.2	0.01 *M* $CaCl_2$ method[d]
Electrical conductivity (mS/cm)	0.3	1.52	0.092	0.373	0.39	saturated paste method
Bulk Density (g/cm^3)	0.98	0.83	—	—	0.51	clod method
Total Carbon (%)	4.46	6.83	1.88	2.57	11.9	Leco furnace method
Inorganic Carbon (%)	—	—	0.18	0.58	< 0.05	Leco furnace method
Organic Carbon (%)	—	—	1.7	1.99	11.9	Leco furnace method

Parameter	Artificial Soil	Clay Loam	Sandy Loam	Silt Loam	Forest Soil	Analytical Method
Organic Matter (%)	9	12.8	2.9	3.5	19.9	dichromate oxidation
Cation Exchange Capacity ($Cmol^+/kg$)	18.5	34.5	16.1	21.9	20	barium chloride method
Total Nitrogen (%)	0.05	0.59	0.115	0.166	0.74	Kjeldahl method
NH_4-N (mg/kg)	—	—	0.53	10.25	260	Kjeldahl method
NO_3-N (mg/kg)	—	—	6.94	5.44	2.26	Kjeldahl method
NO_2-N (mg/kg)	—	—	0.94	< 0.1	< 0.1	Kjeldahl method
Phosphorus (mg/kg)	23	12	6	10	35	nitric/perchloric acid digestion
Potassium (mg/kg)	22	748	61	75	250	NH_4 acetate extraction, colourimetric analysis
Magnesium (mg/kg)	149	553	261	256	192	NH_4 acetate extraction, colourimetric analysis
Calcium (mg/kg)	1848	5127	1846	4380	963	NH_4 acetate extraction, colourimetric analysis
Chloride (mg/kg)	—	—	69	42	113	H_2O extraction, colourimetric analysis
Sodium (mg/kg)	67	57	33	19	38	NH_4 acetate extraction, colourimetric analysis

[a] Characteristics of the artificial and various negative control soils that have been used to develop the definitive biological test methods and associated criteria for test validity described in this test method document (Aquaterra Environmental, 1998; Stephenson *et al.*, 1999a, 1999b, 2000a; Aquaterra Environmental and ESG, 2000; ESG, 2001, 2002; ESG and Aquaterra Environmental, 2002; Stantec and Aquaterra Environmental, 2004).

[b] Not determined.

[c] Determined according to USEPA (1989) using a Fisherbrand P8 creped filter paper (see Section 5.3).

[d] Determined by Becker-van Slooten *et al.* (2004) according to Hendershot *et al.* (1993) (see Section 4.2.5).

A large (~3000 L) sample of sandy loam soil was collected in June 1999 from Beauchamp Farms, Eramosa, Ontario, from a site that had been cultivated regularly for crop production but not subjected to pesticide application. The soil was air-dried and sieved (2 or 5 mm), placed into 20-L plastic buckets, and kept in cold storage (4 °C) until needed. This soil was analyzed for common organic and inorganic contaminants, and its physicochemical characteristics established to determine if any unusual soil characteristics (e.g., high electrical conductivity or anomalous nutrient levels) were present. The sample was found to be virtually free of both contaminants and anomalies. This soil is a fine sandy loam with a moderate organic content and a moderate cation exchange capacity compared to the other clean soils included in these studies (see Table F-1).

The sample of silt loam soil was collected in June 1999 from the University of Guelph Elora Research Station, in Nichol Township, Ontario. The topsoil had been removed several years ago when the research facility was built, and had been stockpiled beside a field. Soil collected for these method development studies was removed from the interior of the pile to avoid collecting soil that might have been inadvertently contaminated with pesticide or fertilizer spray drift from the adjacent field. The soil was air-dried and sieved (2 or 5 mm), placed into 20-L plastic buckets, and kept in cold storage (4 °C) until needed. The soil was also analyzed and found to be free of both organic and inorganic contaminants and anomalies. The measured physicochemical characteristics of this silt loam soil showed that it had a moderate organic content and a moderate cation exchange capacity, compared to the other four soils included in the method development studies for the first edition of this biological test method document (see Table F-1).

A 400-L sample of forest soil, classified as Orthic Humo-Ferric Podzols, was collected in June 2001 from a forested area located on the Canadian Shield, in Sudbury, Ontario. The leaf litter was gently raked away, and a hand trowel was used to remove soil to a depth ranging from 5 to 10 cm. The soil was placed without sieving into 20-L plastic-lined buckets, and transported to ESG International Inc. in Guelph, Ontario. It was air-dried for 48 hours to no less than ~10% moisture content, homogenized, and then sieved through 6-mm mesh. Once the sample was sieved, it was thoroughly homogenized and stored in the same 20-L plastic buckets until used. This soil was stored at room temperature (20 °C) until use. The physicochemical characteristics of the forest soil show that it is a loam with a moderate cation exchange capacity, and the highest total organic carbon content (11.9%) and highest percentage of organic matter (19.9%) of the five soils used in the method development studies for the first edition of this biological test method document (see Table F-1).

For this second edition test method document, the performance of *D. rubidus* was assessed in different types of negative control soil representative of an array of clean soils collected from the boreal and taiga ecozones within Canada. Nine negative control soils were used to develop the biological test methods described herein for use with *D. rubidus* and to further assess the robustness of the test methods with samples of soil that varied considerably in their physical and chemical characteristics. These soils were also used to establish reasonable test validity criteria based on control performance in the 56-day reproduction test for *D. rubidus*. The nine soils tested included an artificial soil (see Section 3.3.2) and 8 natural soils with various soil horizons (see Section 3.3.1) (EC, 2010; ECCC, 2020b). The natural soils consisted of 1 agronomic soil and 7 natural soils from the boreal and taiga ecozones. The physicochemical characteristics of these soils are summarized in Table F-2.

The artificial control soil used in this series of performance evaluation studies with diverse soil types was the same formulated soil as that recommended for use herein (see Section 3.3.2) and the same as that used for *E. andrei* in the first edition of this test method document, described earlier in this section. It consists of 70% silica sand, 20% kaolin clay, 10% *Sphagnum* sp. peat and calcium carbonate (10–30 g $CaCO_3$/kg peat). The soil was formulated by mixing the ingredients in their dry form thoroughly, then gradually hydrating with deionized water, and mixing further until the soil was visibly uniform in colour, texture, and degree of wetness. This artificial soil is much the same as that described by ISO (2012) and OECD (2016).

Table F-2 Physicochemical characteristics of candidate artificial and natural negative control boreal soils and soil horizons[a]

			Soil type:	Artificial soil	LUFA	NFLD01 podzol	NB podzol	
			Source:	In-house	Standard soil from Europe	Newfoundland	New Brunswick	
			Soil classification:	n/a	n/a	Gleyed humo-ferric podzol	Gleyed humo-ferric podzol	
			Horizon:	n/a	n/a	Bf	A	B
Parameter	**Units**	**Analytical method**						
Soil texture[b]		n/a[c]		–	LS	–	SCL	SL
Sand	%	Particle size distribution (filter candle system)		72	77	72	79	62
Silt	%			20	17	20	1	28
Clay	%			8	6.5	8	20	10
Water-holding capacity	%	EC (2005b)		41.9	47.9	41.9	67.6	80.6
Optimal moisture content	%			55.0	57.5	55.0	65	65
pH	units	1:1 water method		4.2	5.6	4.2	4.7	4.6
Electrical conductivity	mS/cm	Saturated paste method		–	1.6	–	0.23	0.06
Organic carbon	%	Leco furnace method		–	–	–	41.1	3.7
Organic matter	%	Loss on ignition		4.6	3	4.6	77.1	10.9
Cation exchange capacity	$Cmol^+/kg$	Barium chloride method		–	< 10	–	–	–
Total nitrogen	%	Kjeldahl method		–	1640	–	1.72	0.23
NH_3	mg/kg	2N KCl extractable		15	< 20	15	783	19
NO_3-N	mg/kg			<10	36	<10	3	9
NO_2-N	mg/kg			<1	< 1	< 1	–	–
Phosphorous (total)	%			0.04	0.03	0.04	–	–
Phosphorous	mg/kg	$NaHCO_3$ extractable		4	230[d]	4	99	18
Potassium	mg/kg	NH_4 acetate extraction, colourimetric analysis		20	360	20	917	1030
Magnesium	mg/kg			20	590	20	784	6560
Calcium	mg/kg			< 100	1400	< 100	4190	608
Sodium	mg/kg			10	< 50	10	128	< 100
C/N				–	–	–	23.9	16
Sodium adsorption ratio		Saturated paste method		–	0.25	–	1.8	1.2

		Soil type:	ON Podzol			AB02 Chernozem	
		Source:	Ontario			Alberta	
		Soil classification:	Gleyed Humo-ferric Podzol			Rego dark grey chernozem	
		Horizon:	Ahe	Of/Oh	Of/Oh	Ah	Ck
Parameter	**Units**	**Analytical method**					
Soil texture[b]		n/a[c]	LS	LS	LS	SL	SL
Sand	%	Particle size distribution (filter candle system)	82	88	86	51	71
Silt	%		12	6	6	43	24
Clay	%		6	6	8	6	6
Water-holding capacity	%	EC (2005b)	41.0	181.9	40.9	68.3	51.4
Optimal moisture content	%		65.0	52.5	47.5	55.0	47.5
pH	units	1:1 water method	4.6	4.6	5.8	7.1	7.7
Electrical conductivity	mS/cm	Saturated paste method	–	–	–	0.34	0.2
Organic carbon	%	Leco furnace method	32.1	1.6	1.0	6.3	1.5
Organic matter	%	Loss on ignition	58.1	2.1	2.2	9.5	2.6
Cation exchange capacity	$Cmol^+/kg$	Barium chloride method	26	9	12	25	16
Total nitrogen	%	Kjeldahl method	0.96	0.06	0.05	0.43	0.09
NH_3	mg/kg	2N KCl extractable	128	4	2	2	1
NO_3-N	mg/kg		< 1	< 1	< 1	15	1
NO_2-N	mg/kg		< 1	< 1	< 1	< 1	< 1
Phosphorous (total)	%		–	–	–	–	–
Phosphorous	mg/kg	$NaHCO_3$ extractable	16	2	< 2	17	8
Potassium	mg/kg	NH_4 acetate extraction, colourimetric analysis	143	23	16	430	203
Magnesium	mg/kg		151	31	40	431	235
Calcium	mg/kg		765	184	191	3380	2400
Sodium	mg/kg		57	35	21	–	12
C/N			33.4	26	20.6	14.6	16.2
Sodium adsorption ratio		Saturated paste method	2.0	2.8	2.4	1.2	1.2

		Soil type:	**SK01 Luvisol**	**SK02 Brunisol**		**SK09**
		Source:	Saskatchewan	Saskatchewan		Saskatchewan
		Soil classification:	Dark grey luvisol	Orthic eutric brunisol		Eluviated dystric brunisol
		Horizon:	Bt	FH	AB	LFH / Ae
Parameter	**Units**	**Analytical method**				
Soil texture[b]		n/a[c]	L	SL	LS	LS
Sand	%	Particle size distribution (filter candle system)	35	89	82	77
Silt	%		55	7	12	17
Clay	%		10	6	4	6
Water-holding capacity	%	EC (2005b)	42.1	174.1	39.5	–
Optimal moisture content	%		42.5	55.0	45.0	–
pH	units	1:1 water method	6.6	6.9	6.8	4.2
Electrical conductivity	mS/cm	Saturated paste method	–	–	–	–
Organic carbon	%	Leco furnace method	1.0	11.4	1.0	8.1
Organic matter	%	Loss on ignition	2.0	15.8	1.8	9.6
Cation exchange capacity	$Cmol^+/kg$	Barium chloride method	11	22	6	< 1
Total nitrogen	%	Kjeldahl method	0.07	0.65	0.05	0.14
NH_3	mg/kg	2N KCl extractable	5	23	6	55
NO_3-N	mg/kg		3	86	< 1	< 10
NO_2-N	mg/kg		< 1	< 1	< 1	< 1
Phosphorous (total)	%		0.06	0.05	0.02	0.04
Phosphorous	mg/kg	$NaHCO_3$ extractable	9	24	16	20
Potassium	mg/kg	NH_4 acetate extraction, colourimetric analysis	170	200	83	70
Magnesium	mg/kg		198	785	196	20
Calcium	mg/kg		1780	2860	795	< 100
Sodium	mg/kg		67	64	50	20
C/N			0.3	4	0.6	–
Sodium adsorption ratio		Saturated paste method	0.2	0.4	0.1	–

[a] Characteristics of the artificial and various negative control soils that have been used to develop the definitive biological test method and associated criteria for test validity described herein (EC, 2010; ECCC, 2020b).

[b] SL = sandy loam; LS = loam sand; L = loam; SCL = sandy clay loam.

[c] Not applicable.

[d] Strong acid leachable metal analysis.

The agronomic soil was a loam sand (LUFA 2.2) from Germany Rheinland P-falz Hanhofen, sampled at GroBer Stret, Nr.585 in 2017. The soil is free from pesticides, biocidal fertilizers, or organic manure for at least 5 years prior to collection. The soil was sampled from a 0–20-cm depth, and sieved with a 2-mm mesh screen. The specific soil lot was purchased in 2017, and shipped to Environment and Climate Change Canada (Ottawa, Ontario), where it was stored at ~23 °C until needed. The physicochemical characteristics of the soil are presented in Table F-2.

The Newfoundland soil (NL Podzol) was classified as a Gleyed Humo-ferric Podzol, developed on a stony, loamy-to-sandy, non-calcareous glacial till (EcoDynamics Consulting Inc., 2011a). The main canopy within the site was dominated by balsam fir and scattered black spruce. The understory consisted of sheep laurel (*Kalmia angustifolia*) and creeping snowberry (*Gaulteria hispidula*), regenerating trees, bunchberry (*Cornus canadense*), with lesser amounts of spinulose woodfern (*Dryopteris spinulosa*), cinnamon fern (*Osmunda cinnamomea*), two-leaved solomonseal (*Maianthemum canadense*), and blue bead lily (*Clintonia borealis*). The ground surface was dominated by feathermosses (e.g., Shreber's moss [*Pleurozium schreberi*], stair-step moss [*Hylocomium splendens*], and knight's plume [*Ptilium crista-castrensis*]). Prior to sampling, woody debris and leaf litter were removed, and the underlying organic F and H horizons were collected together, followed by the separate collection of the Ahe (to a depth of 3 cm), Ae (to a depth of 25 cm), and Bf horizons. Only the Bf horizon was used in the establishment of test validity criteria for *D. rubidus* (see Table F-2).

The New Brunswick soil (NB Podzol) was classified as an imperfectly drained Gleyed Humo-ferric Podzol, developed in non-calcareous, medium to moderately fine-textured basal or lodgement till (EcoDynamics Consulting Inc., 2008). The main canopy consisted of a mixed-wood forest, consisting of beech (*Fagus grandifolia*), red maple (*Acer rubrum*), yellow birch (*Betula alleghaniensis*), and sugar maple (*Acer saccharum*), underlain by balsam fir (*Abies balsamea*), with an understory of hazel (*Corylus cornuta*), and regenerating maple and balsam fir (EcoDynamics Consulting Inc., 2008). The forest litter (L horizon) was removed, and the underlying FH and Ahe-Aegj horizons were collected separately and placed into 25-L pails. The underlying Bf horizon was then collected; however, given the variation and wavy nature of the soil horizon boundaries, the collection of some BCgj material was unavoidable. Both the A and B horizons were used in the establishment of test validity criteria for *D. rubidus* (see Table F-2).

The Ontario soil (ON Podzol) was classified as a Gleyed Humo-ferric Podzol developed within a non-calcareous fluvial-lacustrine deposit (EcoDynamics Consulting Inc., 2011b). The site was a coniferous-dominant mixed-wood forest, with a mixture of both coniferous and deciduous species. The upper canopy consisted mainly of red pine (*Pinus resinosa*) and eastern white pine (*Pinus strobus*), with scattered sugar maple (*Acer saccharum*), and with a lower canopy consisting of a mixture of white birch (*Betula papyrifera*), eastern white cedar (*Thuja occidentalis*), black spruce (*Picea mariana*), white spruce (*Picea glauca*), red maple (*Acer rubra*) and eastern hemlock (*Tsuga canadensis*). The understory was dominated by regenerating tree species, with lesser amounts of speckled alder (*Alnus incana*), beaked hazelnut (*Corylus cornuta*), eastern leatherwood (*Dirca palustris*), wild raisin (*Viburnum nudum*), velvet blueberry (*Vaccinium myrtilloides*), and twinflower (*Linnaea borealis*). The ground surface was dominated by bunch berry (*Cornus canadensis*) and goldthread (*Coptis trifolia*). Three horizons were collected following the removal of the forest litter: the Ahe (to a depth of 2 cm), Ae (to a depth of 7 cm), and Bf horizons (to a depth of 20 cm). Only the Ahe horizon was used in the establishment of test validity criteria for *D. rubidus* (see Table F-2).

The Alberta soil (AB02 Chernozem) was collected on a river floodplain terrace, and was characterized as a well-to-moderately well-drained Rego Dark Gray Chernozem (EcoDynamics Consulting Inc., 2007). The texture of the organic-rich Ah horizon was classified as a silt loam, with a very fine sand/loamy to very fine sand-to-very-sandy loam texture occurring with depth. The dominant vegetation consisted of smooth brome (*Bromus inermis* Leyss.), interspersed with small amounts of rose (*Rosa* sp.), northern bedstraw (*Galium boreale* L.), and fireweed *(Epilobium angustifolium* L.). Forested areas close to the river valley slopes contained an aspen overstory, with scattered white spruce. Two horizons were collected: the Ah horizon to a depth of 11cm, and the Ckgj horizon to a

depth of approximately 25–30 cm; there was no defined B horizon. Both horizons were used in the establishment of test validity criteria for *D. rubidus* (see Table F-2).

Three soils were collected from Saskatchewan. The first soil (SK01 Luvisol) was classified as a well- to moderately well-drained Dark Grey Luvisol, developed on stone-free, loamy-to-clayey glaciolacustrine materials (EcoDynamics Consulting Inc., 2007). The forest cover was a mixture of white spruce (*Picea glauca*) and trembling aspen (*Populus tremuloides*), with an understory of aspen suckers, rose (*Rosa* sp.), willow (*Salix* spp.), bunchberry (*Cornus canadensis*), and twinflower (*Linnaea borealis*). Three horizons were collected: LFH (10 cm depth), Ahe (10 cm depth), and Bt (to a depth of 19 cm), but only the Bt horizon was used in the establishment of test validity criteria for *D. rubidus* (see Table F-2).

The second soil collected from Saskatchewan (SK02 Brunisol) was classified as a rapidly-drained, Orthic Eutric Brunisol, developed on stone-free, sandy glaciofluvial materials (EcoDynamics Consulting Inc., 2007). The forest cover consisted of pure jack pine (*Pinus banksiana*), with an understory dominated by aspen (*Populus tremuloides*), green alder (*Alnus crispa*), bearberry (*Arctostaphylos uva-ursi*), and reindeer lichens (*Cladina* spp.). The leaf litter was removed, and the FH was collected to a depth of approximately 6 cm; the Ah and Bm horizons were collected together to a depth of approximately 25–30 cm, as the Ah was discontinuous and thin (2 cm). Both horizons were used in the establishment of test validity criteria for *D. rubidus* (see Table F-2).

The third soil collected from Saskatchewan (SK09 Brunisol) was classified as a well-drained Eluviated Dystric Brunisol profile, underlain by stratified, sandy glaciofluvial, which in turn is underlain by eroded sandy glacial till (EcoDynamics Consulting Inc., 2011c). The vegetation overstory contained mainly a mixture of black spruce (Picea mariana), jack pine (Pinus banksiana), and white birch (Betula papyrifera), with an understory dominated by reindeer lichens (mostly Cladina mitis) and feather mosses (mostly Pleurozium schreberi), along with a mixture of shrubs including Labrador tea (*Ledum* groenlandicum), bog cranberry (Vaccinium vitis-idaeus), blueberry (Vaccinium myrtilloides), bog bilberry (Vaccinium uliginosum), and crowberry (Empetrum nigrum). Surface woody debris and leaf litter were removed prior to sampling, and the underlying organic F and H horizons were sampled; thereafter, the underlying A and B mineral horizons were collected up to a depth of approximately 10 cm.

Appendix G

Logarithmic Series of Concentrations Suitable for Toxicity Tests[a]

Column (Number of concentrations between 10.0 and 1.00, or between 1.00 and 0.10)[b]

1	2	3	4	5	6	7
10.0	10.0	10.0	10.0	10.0	10.0	10.0
3.2	4.6	5.6	6.3	6.8	7.2	7.5
1.00	2.2	3.2	4.0	4.6	5.2	5.6
0.32	1.00	1.8	2.5	3.2	3.7	4.2
0.10	0.46	1.00	1.6	2.2	2.7	3.2
	0.22	0.56	1.00	1.5	1.9	2.4
	0.10	0.32	0.63	1.00	1.4	1.8
		0.18	0.40	0.68	1.00	1.3
		0.10	0.25	0.46	0.72	1.00
			0.16	0.32	0.52	0.75
			0.10	0.22	0.37	0.56
				0.15	0.27	0.42
				0.10	0.19	0.32
					0.14	0.24
					0.10	0.18
						0.13
						0.10

[a] Modified from Rocchini *et al.* (1982).

[b] A series of successive concentrations may be chosen from a column. Midpoints between concentrations in column (x) are found in column (2x + 1). The values listed can represent concentrations expressed as a percentage by weight (e.g., mg/kg) or weight-to-volume (e.g., mg/L) basis. As necessary, values can be multiplied or divided by any power of 10. Column 2, which spans two orders of magnitude in concentration, might be used if there was considerable uncertainty about the degree of toxicity. More widely spaced concentrations should not be used, since such usage gives poor resolution of the confidence limits surrounding any threshold-effect value calculated. The finer gradations of columns 4 to 7 might occasionally be useful for testing chemicals that have an abrupt threshold of effect.

Appendix H

Determining a Positive Control Concentration and Defining Warning Limits – Worked Example

1. Use a minimum of five valid (i.e., test validity criteria must be met) multi-concentration tests in one soil type (e.g., negative control soil or artificial soil), with the same reference toxicant. In this example, 56-day reproduction tests were conducted in a clean field soil (i.e., negative control soil) using boric acid as the reference toxicant.
2. For each test, tabulate the mean total number of juveniles produced per treatment (Table H-1).
3. For each test, calculate and tabulate the percent reduction of juvenile production relative to the control response (Table H-2) using the following formula:

$$\%\ Reduction = \left(\frac{(Control\ Response\ -\ Treatment\ Response)}{Control\ Response}\right) \times 100$$

4. Calculate the mean percent reduction of juvenile production for each treatment (Table H-2). *Optional*: plot the data.
5. Select a concentration where the data tend to be less variable (i.e., range of the data spans ~20%), but still show a partial effect (i.e., 30–70% reduction; see shaded cells in Table H-2).
6. Calculate the standard deviation (SD) and two standard deviations (2 SD) of the mean percent reduction for the selected test concentration (Table H-2).
7. Calculate the mean percent reduction ± 2 SD for the selected test concentration (Table H-2) and compare these values to the minimum and maximum percent reduction observed within that treatment, to ensure that the proposed warning limits (i.e., mean percent reduction ± 2 SD) capture the response data. Use the mean percent reduction at that treatment to define the target effect size.
8. **In this example, 245 mg H_3BO_3/kg dry soil produced a 72% mean percent reduction of progeny production (i.e., target effect size) with proposed warning limits of ≥ 60% and ≤ 84%. Based on these results, this is the test concentration of boric acid that a laboratory might choose and then run concurrently with each definitive test for the positive control treatment.**
9. For tests where the positive control is included as part of the definitive reproduction test, the percent reduction of juvenile production (i.e., effect) is compared to the established warning limits. This is carried out and documented following the same procedures as those used for comparing multi-concentration reference toxicity tests in reference toxicant warning charts (Section 4.4). If the percent reduction of juveniles in a positive control run with a definitive test is within the established warning limits (i.e., mean % reduction ± 2 SD), the positive control is acceptable. If the response is outside of those limits, an investigation into the test conduct and sensitivity of the test population (i.e., in-house cultures) or group of test organisms used in the test (i.e., batch of test organisms obtained from an outside source) must be launched (see Section 4.4). This investigation might include, for example: determining if the positive control concentration was prepared properly, checking test calculations, confirming the positive control concentration analytically, investigating the negative control data, examining culture health data, investigating technician proficiency, or investigating soil age quality (e.g., stored too long in buckets). In addition to maintaining warning charts of positive control data, a laboratory should monitor the variability of the positive control response over time by calculating the coefficient of variation (CV) of the response and evaluating it relative to a predefined acceptability limit (e.g., lab defines ≤ 30% CV as acceptable). In this example the CV is 8.2 % for six data points (Table H-2).

Table H-1 Mean number of *D. rubidus* juveniles produced upon exposure to boric acid in clean field soil

	Boric acid concentration (mg/kg)					
Test No.	**0**	**84**	**120**	**171**	**245**	**350**
1	29.0	13.4	34.4	20.2	6.6	1.5
2	41.0	42.5	49.7	29.7	10.6	0.2
3	56.0	58.2	56.1	35.4	17.1	3.2
4	62.0	63.7	58.8	31.0	18.2	1.2
5	34.0	31.6	28.6	21.4	7.1	1.5
6	25.0	23.5	17.2	20.0	9.3	0.0

Table H-2 Percent reduction in *D. rubidus* juvenile production, relative to the control response, upon exposure to boric acid in clean field soil

	Boric acid concentration (mg/kg)					
Test No.	**0**	**84**	**120**	**171**	**245**	**350**
1	0	53.8	-18.6	30.3	77.2	94.8
2	0	-3.7	-21.2	27.6	74.2	99.5
3	0	-3.9	-0.2	36.8	69.5	94.3
4	0	-2.7	5.2	50.0	70.6	98.1
5	0	7.1	15.9	37.1	79.1	95.6
6	0	6.0	31.2	20.0	62.8	100.0
Mean	**-**	**9.4**	**2.1**	**33.6**	**72.2**	**97.1**
SD[a]					5.9	
2 SD					11.8	
Mean + 2 SD					84.0	
Mean − 2 SD					60.4	
%CV					8.2	

[a] Standard deviation