All individuals are not created equal; accounting for inter-individual variation in fitting life-history responses to toxicants.

Tjalling Jager*

Dept. of Theoretical Biology, VU University Amsterdam, de Boelelaan 1085, NL-1081 HV, Amsterdam, the Netherlands

This document is the Accepted Manuscript version of a Published Work that appeared in final form in Environmental Science & Technology:

Jager T. 2013. All individuals are not created equal; accounting for interindividual variation in fitting life-history responses to toxicants. Environmental Science & Technology 47:1664-1669.

ABSTRACT

The individuals of a species are not equal. These differences frustrate experimental biologists and ecotoxicologists who wish to study the response of a species (in general) to a treatment. In the analysis of data, differences between model predictions and observations on individual animals are usually treated as random 'measurement error' around the 'true' response. These deviations, however, are mainly caused by real differences between the individuals (e.g., differences in physiology and in initial conditions). Understanding these intra-species differences, and accounting for them in the data analysis, will improve our understanding of the response to the treatment we are investigating and allow for a more powerful, less biased, statistical analysis. Here, I explore a basic scheme for statistical inference to estimate parameters governing stress that allows individuals to differ in their basic physiology. This scheme is illustrated using a simple toxicokinetic-toxicodynamic model and a data set for growth of the springtail *Folsomia candida* exposed to cadmium in food. This paper should be seen as 'proof of concept'; a first step in bringing more realism into the statistical inference for process-based models in ecotoxicology.

INTRODUCTION

The individuals of a species are not equal. These differences are the raw material for natural selection to work its magic on but frustrate experimental biologists and ecotoxicologists. These scientists wish to study the response of *a species* in general to a treatment (e.g., a toxicant); an objective that is impeded by large inter-individual differences. Differences between individuals cause 'noise' in the response, hampering the identification of the 'true' effect of the treatment. In ecotoxicology, the issue of intra-species variation has been addressed by standardisation to ensure an (as much as possible) homogeneous cohort of test animals. The animals are cultured and selected in such a way as to ensure low genetic variation; for parthenogenetic species (such as the popular test species *Daphnia magna* and *Folsomia candida*), single clones are used. Toxicity tests are initiated with animals of very similar age and size; for example, in the *D. magna* reproduction test,[1] all starting animals should be less than 24 hours old. If a perfect homogenisation could be achieved, testing a single individual for each treatment would suffice. Clearly, inter-individual differences persist, and hence the need for replication.

Toxicokinetic-toxicodynamic (TKTD) models offer important advantages in the analysis of ecotoxicological test results.[2,3] These methods make use of the toxic response on all endpoints over time in one coherent analysis to extract meaningful parameters from the data.[4] However, intra-species variation offers particular challenges when fitting these models to data. For sub-lethal endpoints, we follow a cohort of animals over time and record the response of life-history traits (e.g., body size and reproduction). In the model analysis, we depart from the (implicit) assumption that there is a single set of parameters that explains the behaviour of all individuals within a treatment over time. The deviations between model and data are treated as random 'measurement error', even though they are probably mainly caused by real differences between individuals (and thus by differences in underlying model parameters).

What consequences should we expect from ignoring inter-individual differences? Firstly, treating these deviations as measurement errors does not help to optimise statistical power. In ecotoxicology, statistical power should be an important aspect in test design and data analysis to ensure an optimal identification and quantification of toxicant impacts from the least amount of test animals. A second issue is with the evaluation and parameterisation of TKTD models: the mean response of a cohort does not equal the response of the average animal. If each individual grows according to a von Bertalanffy growth curve with different parameters, the average of the body sizes at each time point will not follow a von Bertalanffy curve anymore. Treating inter-individual differences as measurement error can thus lead to misinterpretation of the model's validity and to bias in the parameter estimates. Furthermore, when we follow different endpoints from the same individuals, these endpoints are not independent. For example, rapidly-growing individuals will experience different toxicokinetics (e.g., because of dilution by growth) compared to their slower-growing peers, and will demonstrate a different reproduction pattern over time (as body size influences reproduction). Focussing on individual responses over different endpoints may thus unlock important information about model validity and parameter values that would otherwise be buried as 'noise'. An extreme example was presented by Augustine and co-workers,[5] who concluded that the mean response of a toxicant (uranium) on growth and reproduction of zebrafish was practically useless; accounting for individual differences in initial condition, feeding, and handling of the reproduction buffer was needed to provide a clear view of the toxicant effect.

In this paper, I explore a basic scheme for statistical inference to estimate parameters of process-based models for stressor effects, which allows individuals to differ in their basic

physiology. Process-based models aim to explain the effects from underlying mechanisms, rather than providing the best possible description, and this requires a different statistical approach than is common for mixed-effects models. The scheme is illustrated using a simple TKTD model based on DEBtox[6] and a simple data set for growth of the springtail *F. candida* exposed to cadmium in food. Even though this is an ecotoxicological example, a similar framework can be used for other situations where meaningful models are fitted to repeated observations on individuals. This approach certainly requires further refinement and therefore this paper should be seen as 'proof of concept'; a first step in bringing more realism into the statistical inference for process-based models in ecotoxicology and stress ecology.

METHODS

Data set. The growth data are for the springtail *Folsomia candida* exposed to cadmium in food.[7] The mean response data have been analysed before,[4] but here I returned to the raw data. Early in the experiment, animals that died were replaced by other individuals, so the data needed to be reworked to extract the time course of body size for unique individuals only. In each treatment, 15 individuals were followed. Body size was determined as fresh weight; I derive a measure of body length (as the model in Eq. 1 is on length basis) from these data by taking the cubic root of the body volume (calculated from fresh weight assuming a density of 1 mg/mm^3). The resulting 'volumetric length' would be the length of the animal if it were shaped like a cube with the same biomass. For animals that do not change in shape over growth, volumetric length is proportional to actual body length.

In the original publication,[7] an average initial weight on 18 μg fwt was reported for 1-day old juveniles (using a different set of individuals, as the weighing procedure can easily cause injury). However, this value is questionable as Stam and co-workers [8] reported egg volumes of roughly 1-2 10^{-3} mm, which corresponds to some 1-2 μg fwt. The reported initial weight of 18 μg therefore seems much too large: I assume this was a typo and take the initial weight as 1.8 μg fwt, which translates into an initial volumetric length (L_0) of 0.12 mm.

Process model. The model for growth is the von Bertalanffy model as resulting from DEBtox.[6] The change in body length is given by:

$$\frac{d}{dt}L = r_B(fL_m - L) \quad \text{with } L(0) = L_0 \tag{1}$$

where L_m is the maximum body length, r_B the von Bertalanffy growth rate constant, and f the scaled ingestion rate (actual ingestion divided by the maximum for its size, and hence between 0 and 1). The initial size is fixed in the analysis of the data (to 0.12 mm, see previous section), and taken the same for each individual. Fixing initial size is necessary to avoid biologically unrealistic estimates, and is defensible given the synchronisation of the starting animals.

As found earlier,[4] I assume a toxic effect on assimilation, which is included as a linear effect on the scaled ingestion rate f as follows (assuming f=1 in the control):

$$f = 1 - \frac{1}{c_T}\max(0, c_V - c_0) \tag{2}$$

Where c_0 is the no-effect concentration and c_T the tolerance concentration. The scaled internal concentration c_V follows from the external concentration c using a first-order one-compartment model, extended to account for changes in size:[6]

$$\frac{d}{dt}c_V = k_e \frac{L_m}{L}(c - c_V) - c_V \frac{3}{L}\frac{d}{dt}L \quad \text{with } c_V(0) = 0 \qquad (3)$$

With this equation, toxicokinetics depends on body size and on the growth rate. This is included by adding dilution by growth (the last term in Eq. 3), and scaling the elimination rate with the surface:volume ratio of the organism. For an organism that does not change in shape during growth, we can scale the elimination rate by $1/L$. In Eq. 3, the factor L/L_m is introduced, where L_m is the maximum length in the control. Introducing L_m ensures that k_e has the usual dimension (time^{-1}), however, the absolute value of k_e should be interpreted as the elimination rate when $L=L_m$. Comparing individuals requires the reference L_m for this equation to be fixed to a common value (the mean in Table 1).

Reference fit. As a reference, the model of the previous section is fitted to the data set, ignoring differences between individuals. The assumption underlying the statistical model is that each observation (for each individual at each time point) is a sample from an independent normal distribution with a mean provided by the model. The variance of this distribution is treated as a 'nuisance parameter' for each individual.[6,9] Thus, the differences between individuals are considered to be measurement error. I fit the complete set of five parameters: r_B, L_m, k_e, c_0 and c_T. Optimisation is performed using a Nelder-Mead simplex search, and confidence intervals are generated using profile likelihoods.[9,10]

Statistical method for individual differences. In this approach, I depart from the assumption that the differences between individuals, in general, are fully reflected in differences in the model parameters and in the initial conditions. In this case study, differences in initial condition are ignored, as explained earlier. I thus want to fit a set of parameters from a non-linear model to the data for a group of individuals, where some (or all) model parameters can differ between individuals. In statistics, this is the area of nonlinear mixed-effects models, but here I propose a different type of analysis than is common, in order to stay closer to the biological and ecotoxicological strengths and weaknesses of the TKTD model. Firstly, I do not want the data for the treatments to affect the distribution of the basic animal parameters (r_B and L_m). To evaluate the adequacy of the proposed (rather crude) toxicity mechanism (in Eq. 2), I consider it unacceptable to allow a shift in the basic parameters to compensate for a lack of fit due to a poor understanding of the toxicity process. Secondly, I do not have an *a priori* idea about the joint parameter distribution, and do not want to force a distribution on them; it should be possible to judge the adequacy of a specific distributional form. And finally, I want to keep control over each step in the analysis to guard against biologically unrealistic fits. For these reasons, I here follow a non-standard approach to this regression problem.

If we would only be interested in the response in the control, it would be straightforward to fit the model to all individual responses and analyse the resulting parameters. However, because we are dealing with a treatment (toxicant stress) the problem becomes more complex. For an individual in a treatment, we do not know what it's response would have been in the control; there is no direct way to access it's basic parameters (in this case r_B and L_m). To tackle this problem requires a set of simplifying assumptions:

1) Each individual follows the specified TKTD model of Eq. 1-3, but with a different set of basic parameters (r_B and L_m).
2) The basic parameters in the population can be described by a multivariate distribution, such as the bivariate normal (specified by a vector of means $\boldsymbol{\mu}$ and a covariance matrix Σ).
3) The parameters governing toxicokinetics and toxicodynamics (k_e, c_0, and c_T) are the same for all individuals.
4) The deviations between the observations and model predictions for an individual over time follow independent normal distributions (with the variance treated as 'nuisance parameter' for each individual).

These assumptions can certainly be refined in the future, but here they serve to demonstrate the feasibility of the approach. Note that only the basic parameters are allowed to vary between individuals, but the toxicity parameters are fixed. The reason for this choice is that the data set does not provide independent information for these parameters (independent from the basic parameters). As the basic parameters already lead to different growth curves for the different individuals, there is no possibility to estimate stochastic toxicity parameters, unless there is variation in the treated individuals that cannot be explained by the basic parameter distribution.

Looking closely at the assumptions, there is a global set of model parameters that are common to all individuals. There is a vector for the means of the control parameters $\boldsymbol{\mu} = [\mu_{rB}, \mu_{Lm}]$, with Σ as the variance-covariance matrix for r_B and L_m (containing three terms; two variances and one covariance). Furthermore, a toxicity parameter set (k_e, c_0 and c_T), making a total of eight parameters (three more than for the reference fit). Even when assuming the same TK parameter (k_e) for all individuals, we still get differences in internal concentration due to differences in growth curves between individual animals (see Eq. 3).

I want to fit the data for individuals in the treatments, using the global parameters to constrain the fits. This requires an optimisation for each individual, nested in an overall optimisation of the global parameters. In words, the procedure is as follows (a more technical description is provided in the supporting information). The first step is to fit all control individuals separately, and use the resulting parameters (r_B and L_m) to construct a multivariate normal distribution. This yields the distribution parameters for the control response ($\boldsymbol{\mu}$ and Σ, best unbiased estimate, calculated using the Matlab function "cov"). In the second step, we use the data for the treatments in a meta-optimisation. This optimisation step starts by selecting values in the global parameter set for the toxicity (k_e, c_0 and c_T). For each individual, find the best fitting set of control parameters (r_B and L_m), given that they are taken from the multivariate distribution constructed in the first step. The likelihood from the fit on the data is thus multiplied by the probability to obtain that parameter combination from the bivariate distribution characterised by $\boldsymbol{\mu}$ and Σ. The highest likelihood value for each individual is collected, and an overall likelihood value is obtained by multiplying the best likelihood values for each individual. The overall likelihood is maximised to obtain the best fitting values in the global parameter set.

The parameter values from the reference fit serves as starting values for the meta-optimisation steps. Optimisation is performed using a Nelder-Mead simplex search, and confidence intervals are generated using profile likelihoods.[9,10] All calculations are coded in Matlab R2010a.

RESULTS AND DISCUSSION

Reference fit. In Figure 1, the reference fit to the data set is shown (corresponding model parameters in Table 1). A single set of model parameters is fit to the observations for all individuals, thus effectively treating inter-individual differences as measurement error. The confidence intervals reveal that the no-effect concentration (c_0) is not significantly different from zero, and that the elimination rate constant (k_e) might be infinite (instantaneous steady state). This indicates that the model is over-parameterised. However, the point of this exercise is not to provide the best description of this data set, but to fit a general (well-established) model framework, and to extract meaningful process information from the data.

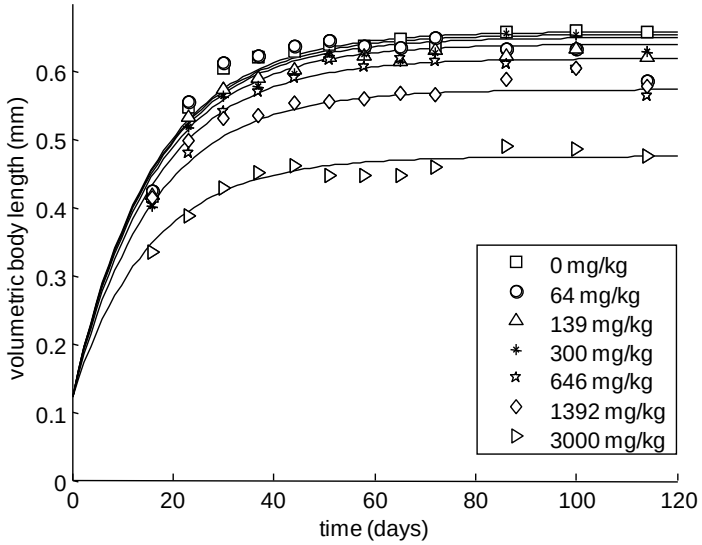

Figure 1. Reference fit for growth of *Folsomia candida* exposed to cadmium in food (concentrations on dry weight basis). Length is in volumetric length (cubic root of body volume). Fit is based on individual data points, but only means are plotted.

Table 1. Estimated model parameters with 95% likelihood-based confidence intervals. The elimination rate is referenced to an individual with L_m=0.680 mm for all fits (see Eq. 3).

Parameter	Reference fit without individual variation	Alternative fit with individual variation
(Mean) von B. growth rate constant (r_B in d^{-1})	0.0616 (0.0586-0.0645)	0.0581 (0.0536-0.0626)
(Mean) maximum volumetric body length (L_m in mm)	0.659 (0.653-0.665)	0.680 (0.659-0.700)
Variance von B. growth rate constant	n.a.	7.61 (3.76-16.0)10^{-5}
Variance maximum volumetric body length	n.a.	16 (7.89-33.6) 10^{-4}
Correlation coefficient r_B and L_m	n.a.	-0.614 (-0.842 - -0.196)
Elimination rate constant (k_e in d^{-1})	1.62 (>0.310)	0.139 (0.100-0.235)
No-effect concentration (c_0 in mg/kg food)	0 (0-84.5)	1.83 (0-46.4)
Tolerance concentration (c_T in mg/kg food)	10.8 (10.0-11.5) 10^3	8.99 (8.41-9.69) 10^3

Fit controls with individual differences. The first step in the alternative statistical framework is to scrutinise the individual growth curves in the control treatment. The left panel of Figure 2 shows all individuals and the best fitting growth curves (more detailed plots are provided in the supporting information). A slight misfit is observed where the observations suggest a somewhat

more s-shaped curve. This might relate to some unrecognised food limitation for the early juveniles.[11] For this reason, it would have been useful to have more size measurements for early juveniles.

The right panel shows the distribution of the underlying model parameters, revealing that the two basic model parameters (r_B and L_m) are negatively correlated in this data set, and supporting to some extent the choice for the bivariate normal distribution. The variation between individuals is considerable, even though we are dealing with a clone (so genetically identical individuals), with one individual per container (so no interactions between individuals), and excess food. Clearly, the response from 15 individuals is not much to specify the multivariate distribution. However, controls from other studies (using the same clone under the same conditions) might be added in the future to improve the characterisation of the distribution.

The best estimate for the model parameters r_B and L_m from standard fit (ignoring inter-individual differences) differs from the means of the fits on the control individuals. The alternative fit indicated a lower r_B and a higher L_m. This illustrates the possibility for bias when treating inter-individual differences as measurement error, although the confidence intervals of the parameters in the two approaches overlap.

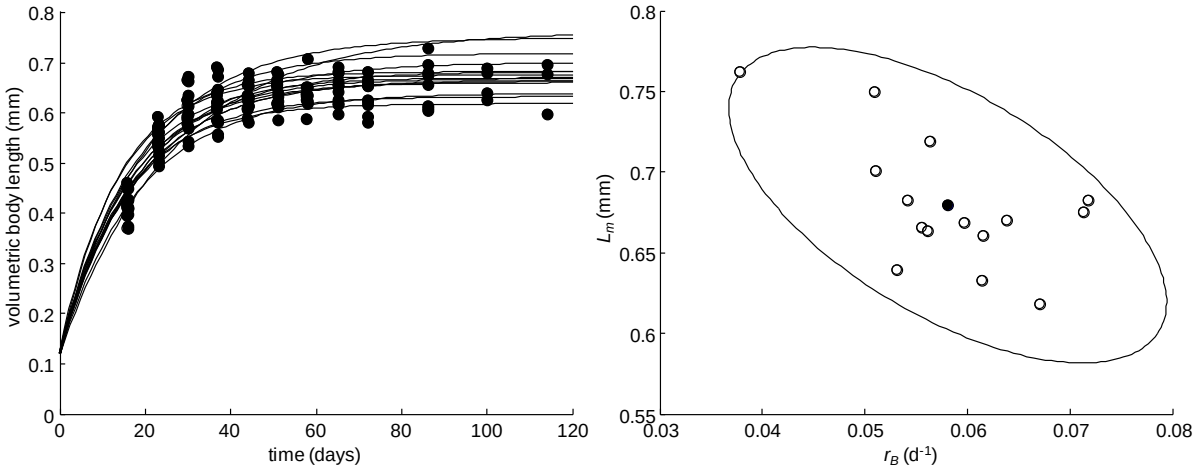

Figure 2. Fit for growth in 15 control animals (*Folsomia candida*) from the data set (lines represent fits for different individuals). Length is in volumetric length (cubic root of body volume). Right panel show the simultaneous distribution of the resulting parameter values; filled circle gives the overall mean, and the ellipse provides the 95% coverage given the best-fitting bivariate normal distribution.

Fit treatments with individual differences. The next step is to use the distribution from the previous step to constrain the model fits on the individuals in the treatments with cadmium. The individual data and fits are shown in Figure 3 (more detailed plots are provided in the supporting information). Most strikingly, this fit yields different parameter estimates for the elimination rate (k_e) and the tolerance (c_T); the confidence intervals do not even overlap (and k_e is now well defined, instead of the confidence interval extending to infinity). This is a clear indication of bias in the reference fit. The estimate for the no-effect concentration (c_0) is not very different, but the confidence interval is tighter than in the standard fit.

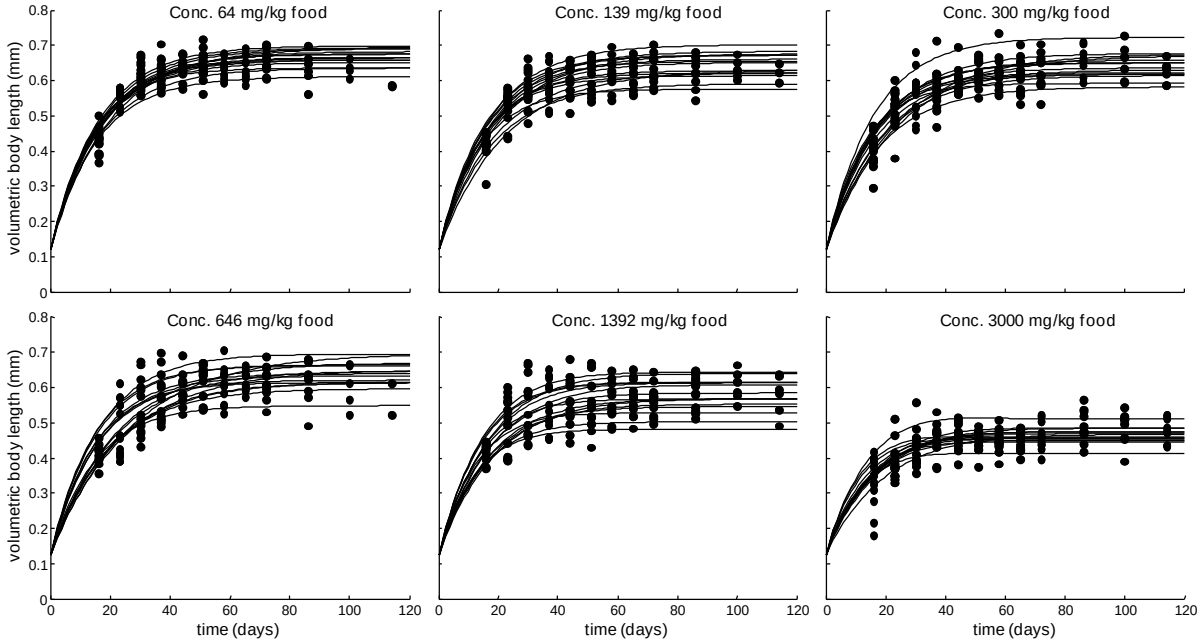

Figure 3. Fits of the model for all individuals of *Folsomia candida* in each treatment (cadmium in food).

Areas for improvement. In this analysis, I decided to allow two growth parameters to vary between individuals (r_B and L_m), while the other parameters where forced to be the same for all individuals. This was mainly done for practical reasons: there is insufficient information in the data set to deduce differences in initial size (L_0) or differences in toxicant-related parameters (k_e, c_0 and c_T).

The first assumption of this framework states that the TKTD model is 'true'. Even though the TK model (Eq. 3) includes the biological realistic impacts of growth, losses due to reproduction are ignored. This is an area for further improvement, which requires information on the transfer of the toxicant to eggs, as well as observations on reproduction on the same individuals. Reproduction might also affect the TD model (Eq. 1), as *F. candida* reproduces in clutches. This implies that a reproduction buffer is build up and emptied several times over the course of the experiment, with potential consequences for the weight measurements.

In this framework, parameters for individuals in the treatments are fitted with a constraining distribution derived from the control animals. This is equivalent to the use of informative priors in Bayesian statistics. The presented approach therefore readily lends itself to be voiced in a Bayesian framework for ecotoxicological data analysis, such as presented by Billoir an co-workers.[12]

Evaluation and outlook. The data set (Fig. 2 and 3) shows that even in a clone of a standard test species, kept individually with excess food, differences between individuals are still considerable. The framework that I present in this paper allows quantification of these differences and including them in the analysis of a stress treatment (in this case a toxicant). This paves the way for more powerful and less biased estimation of model parameters. Furthermore, quantification of intra-species differences will also benefit individual-based population models that include a mechanistic model for the individual.[13]

The analysis for the growth data of *F. candida* exposed to cadmium shows that the parameters for the toxic effect can still be extracted when allowing basic parameters (r_B and L_m) to vary between individuals. Our alternative approach does not lead to radically different conclusions regarding the overall effect of cadmium but the parameter estimates are clearly different (especially for k_e and c_T). This analysis thus shows that treating inter-individual variation like measurement error can lead to a bias in the estimation of model parameters. Distinguishing individuals probably becomes even more important when the observations include more endpoints per individual (e.g., growth and reproduction and time of death), and when initial differences in the individuals are more substantial. The work of Augustine et al [5] clearly demonstrates the need for focussing on individuals in such cases.

This scheme for statistical inference, in combination with TKTD modelling, invites a very different view on optimal test design. Test protocols have always been designed from the perspective of deriving descriptive summary statistics such as ECx and NOEC at the end of the test; focussing on the needs to parameterise TKTD models on an individual basis will lead to entirely different protocols. Clearly, this approach requires following the same individuals over time, and preferably the quantification of multiple endpoints (e.g., growth, reproduction and survival).[4] For some species, this may be a relatively simple adaptation, but for others this is a more serious issue. For a small soil-dwelling invertebrate like *F. candida*, following individuals over time is only possible under ecologically less-relevant conditions (i.e., without soil). A practical benefit of following individuals is that the starting cohort does not need to be perfectly synchronised: the initial differences between the animals are accounted for in the analysis. This claim rests on the assumption that we can account for all relevant differences in the model parameters, which requires further testing. A final consequence for testing is that if such an approach indeed proves to increases statistical power, this allows for a reduction in the number of test animals.

The framework, as presented here, should be seen as a first step, and the analysis of this data set as 'proof of concept'. Clearly, more work is needed in terms of dedicated testing and model development. However, I hope that the ideas presented in this paper will inspire further research on increasing biological and toxicological realism in the analysis of toxicity data.

ASSOCIATED CONTENT

Supporting information.
Technical details of the combined likelihood functions and additional detailed plots for the individual growth curves and parameter estimates. This material is available free of charge via the Internet at http://pubs.acs.org.

ACKNOWLEDGEMENTS

This research has been financially supported by the European Union under the 7th Framework Programme (project acronym CREAM, contract number PITN-GA-2009-238148).

REFERENCES

(1) OECD *OECD guideline for testing of chemicals 211. Daphnia magna reproduction test*; Organisation for Economic Cooperation and Development (OECD): Paris, France, **2008**.

(2) Ashauer, R.; Escher, B. I. Advantages of toxicokinetic and toxicodynamic modelling in aquatic ecotoxicology and risk assessment. *J. Environ. Monit.* **2010**, *12*, 2056-2061.

(3) Jager, T.; Heugens, E. H. W.; Kooijman, S. A. L. M. Making sense of ecotoxicological test results: towards application of process-based models. *Ecotoxicology* **2006**, *15*, 305-314.

(4) Jager, T.; Crommentuijn, T.; Van Gestel, C. A. M.; Kooijman, S. A. L. M. Simultaneous modeling of multiple endpoints in life-cycle toxicity tests. *Environ. Sci. Technol.* **2004**, *38*, 2894-2900.

(5) Augustine, S.; Gagnaire, B.; Adam-Guillermin, C.; Kooijman, S. A. L. M. Effects of uranium on the metabolism of zebrafish, *Danio rerio*. *Aquatic Toxicol.* **2012**, *118*, 9-26.

(6) Jager, T.; Zimmer, E. I. Simplified Dynamic Energy Budget model for analysing ecotoxicity data. *Ecol. Modell.* **2012**, *225*, 74-81.

(7) Crommentuijn, T.; Doodeman, C. J. A. M.; Doornekamp, A.; Van Gestel, C. A. M. In *Ecological risk assessment of contaminants in soil*; Van Straalen, N. M., Løkke, H., Eds.; Chapman & Hall: London, UK, 1997; pp 275-291.

(8) Stam, E. M.; Van de Leemkule, M. A.; Ernsting, G. Trade-offs in the life history and energy budget of the parthenogenetic collembolan *Folsomia candida* (Willem). *Oecologia* **1996**, *107*, 283-292.

(9) Pawitan, Y. *In all likelihood: statistical modelling and inference using likelihood*; Oxford University Press: Oxford, UK, **2001**.

(10) Meeker, W. Q.; Escobar, L. A. Teaching about approximate confidence regions based on maximum likelihood estimation. *Am. Stat.* **1995**, *49*, 48-53.

(11) Zimmer, E. I.; Jager, T.; Ducrot, V.; Lagadic, L.; Kooijman, S. A. L. M. Juvenile food limitation in standardized tests: a warning to ecotoxicologists. *Ecotoxicology* **2012**, 21, 2195-2204.

(12) Billoir, E.; Delignette-Muller, M. L.; Péry, A. R. R.; Charles, S. A Bayesian approach to analyzing ecotoxicological data. *Environ. Sci. Technol.* **2008**, *42*, 8978-8984.

(13) Martin, B. T.; Zimmer, E. I.; Grimm, V.; Jager, T. Dynamic Energy Budget theory meets individual-based modelling: a generic and accessible implementation. *Methods Ecol. Evol.* **2012**, *3*, 445-449.