Regulatory Toxicology and Pharmacology 69 (2014) 171–177

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Current modeling practice may lead to falsely high benchmark dose estimates

Joakim Ringblom *, Gunnar Johanson, Mattias Öberg

Karolinska Institutet, Institute of Environmental Medicine, Unit of Work Environment Toxicology, P.O. Box 210, SE-171 77 Stockholm, Sweden

ARTICLE INFO

Article history:
Received 10 December 2013
Available online 21 March 2014

Keywords:
Benchmark dose
Simulation
Risk assessment
Dose–effect
Point of departure

ABSTRACT

Benchmark dose (BMD) modeling is increasingly used as the preferred approach to define the point-of-departure for health risk assessment of chemicals. As data are inherently variable, there is always a risk to select a model that defines a lower confidence bound of the BMD (BMDL) that, contrary to expected, exceeds the true BMD. The aim of this study was to investigate how often and under what circumstances such anomalies occur under current modeling practice. Continuous data were generated from a realistic dose–effect curve by Monte Carlo simulations using four dose groups and a set of five different dose placement scenarios, group sizes between 5 and 50 animals and coefficients of variations of 5–15%. The BMD calculations were conducted using nested exponential models, as most BMD software use nested approaches. "Non-protective" BMDLs (higher than true BMD) were frequently observed, in some scenarios reaching 80%. The phenomenon was mainly related to the selection of the non-sigmoidal exponential model (Effect = $a \cdot e^{b \cdot \text{dose}}$). In conclusion, non-sigmoid models should be used with caution as it may underestimate the risk, illustrating that awareness of the model selection process and sound identification of the point-of-departure is vital for health risk assessment.

1. Introduction

Health risk assessment of chemicals is the process of characterizing and quantifying the potential adverse health effects associated with exposure to chemicals (NRC, 1994). Characterizing the hazardous properties of a chemical agent in quantitative terms includes selection of a relevant dataset and description of the dose–effect relationship for the critical effect. With the exception of direct acting carcinogenic substances it is assumed that there are some doses that do not result in adverse effects (Dybing et al., 2002; Edler et al., 2002). An important part of the risk assessment is therefore the process of finding the threshold dose, below which toxicity is not expected. Experimental data are used to describe the dose–effect relationship and to define the point of departure (POD), for development of reference or limit values such as acceptable or tolerable daily intakes, occupational exposure limits, derived no effect levels, population adjusted doses and acute guidance values. Such values are key elements in the regulatory process to define the quality standards for water, food, air, work places etc. It is therefore of vital interest that the methods used to interpret toxicological data are continuously examined and further developed to ensure the best use of data.

Traditionally, the no-observed-adverse-effect level (NOAEL) method is used as the POD in regulatory toxicity testing. The NOAEL is defined as the highest tested dose that does not give an effect which is statistically significant from that in the control group (WHO, 1999). The benchmark dose (BMD), on the other hand, is defined as the dose that results in a predefined effect level, ideally at a level that is close to, but not yet, adverse. The BMD method is suggested as a more scientifically sound alternative to the NOAEL method and has been implemented in most regulatory guidance documents as an alternative or preferred approach (ECHA, 2008; EFSA, 2009; NAC/AEGL, 2001; Solecki et al., 2005; USEPA, 1995; WHO, 2009). The BMD method involves fitting a dose–effect curve, or a set of curves, to the data of interest. A critical effect size, also referred to as benchmark response, of 5% is often used as a default for continuous data. The lower bound of the 90% confidence interval of the BMD (BMDL) is suggested to be used as the POD (Crump, 1984; Sand et al., 2008; USEPA, 2012).

Abbreviations: BMD, benchmark dose; BMDL, lower confidence bound of the benchmark dose; NOAEL, no observed adverse effect level; LOAEL, lowest observed adverse effect level; EFSA, European Food Safety Authority; OECD, Organisation for Economic Co-operation and Development; USEPA, United States Environmental Protection Agency.

* Corresponding author. Fax: +46 8 31 41 24.
E-mail addresses: joakim.ringblom@ki.se (J. Ringblom), gunnar.johanson@ki.se (G. Johanson), mattias.oberg@ki.se (M. Öberg).

http://dx.doi.org/10.1016/j.yrtph.2014.03.004
0273-2300/

The BMD approach overcomes many of the weaknesses of the NOAEL (Allen et al., 1994; Crump, 1984). BMD incorporates information on the sample size and the shape of the dose–effect curve. In addition, BMD values are not constrained to one of the experimental doses, and are less dependent on the study design. Furthermore, high variability in the data and low number of subjects both tend to result in a higher NOAEL, i.e. a less precautious POD, whereas the BMDL values tend to become lower, i.e. more precautious, as variability and sample size decreases. BMD aligned design of experiments have also been suggested as a mean to refine and reduce animal experiments (Öberg, 2010).

Several dose–effect or dose–response models may be fitted to the data. Historically, power models and polynomials were used for continuous data, but in recent years more attention has been given to the nested exponential models and nested Hill models that describe sigmoidal curves that level off at higher doses and are assumed to be more biologically relevant (Sand et al., 2008). Nested models are included in the most frequently used software for the statistical analysis of dose–effect data, i.e. BMDS (USEPA, 2013), developed by the US EPA, and PROAST (Slob, 2011), developed by the Dutch National Institute for Public Health and the Environment (RIVM). These models are also recommended by EFSA (EFSA, 2009). In the nested approaches, the software algorithm moves from simpler (few parameters) to more complex models (more parameters) according to statistically based decision rules.

In theory, the BMDL should be lower than the "true" BMD in 95%, and higher in 5% of all cases, provided that the correct model is used. The assumption of model correctness is easily ignored by risk assessors as a possible source of error. However, considering the variability of the experimental data, there is an obvious risk that the use of any approach results in an inappropriate model selection (i.e. selection of a model that deviates from the true dose–effect relationship). This might in turn result in a misleading calculation of the BMD and BMDL. Thus, there is a risk that estimated BMDL values become higher than the true BMD far more frequently than the expected 5%. A high frequency of "false" or "non-protective" BMDLs would be unfortunate, as the resulting POD would then be less protective than assumed. The aim of this study was to use simulated data to investigate how frequently and under what circumstances non-protective BMDL values occur with continuous data, and to quantify the possible impact on quantitative health risk assessment.

2. Materials and methods

The occurrence of non-protective BMDLs was investigated using one fixed dose–effect curve with added variability. The OECD standard of four dose group was tested, i.e. control, low, medium and high, with logarithmic dose spacing (1, 3, 10). Five dose-placement scenarios (A–E) were tested, in these the doses were placed on different parts of the dose–effect curve, while maintaining the 1–3–10 dose spacing. Different dose placements and variabilities were used to take into account that in a real toxicity study the potency of the tested substance is not known beforehand and that several similar endpoints with different sensitivity and variability may be evaluated from the same study.

2.1. Scenarios

Five different dose placement scenarios with the same relative logarithmic (control, 1, 3, 10) dose spacing were used, ranging from a scenario with the three dose groups at the low end of the dose–effect curve where the highest dose were corresponding to a dose that caused 50% of the maximum effect (ED_{50}) to a scenario with the dose group at the high end where the lowest dose corresponding to ED_{50} (Fig. 1). Simulations were performed with 5, 10, 20 and 50 animals per dose group. Both the number of dose groups and the number of animals per group were chosen according to standard OECD Guidelines for the testing of chemicals (OECD, 2012).

2.2. Generation of dose–effect data

1200 datasets were generated for each of the five scenarios by Monte Carlo simulation. The effect size of each individual in these datasets was calculated with the exponential model (as described by Moerbeek et al., 2004):

$$y = a \times (c - (c-1) \times e^{-bx^d}) + \varepsilon \quad (1)$$

where (y) is the effect and (x) is the dose. The background effect (a) and the shape parameter (d) were both set to 1, whereas the maximum effect (c) was set to 1.3. These values are realistic (Slob and Setzer, 2013) and the same values for (a) and (c) were also used by Slob et al. (2005). The potency parameter (b) was set to 0.231 to obtain an ED_{50} value of 3. The critical effect size was set to 5% increase over background corresponding to a "true" BMD of 0.79. A log-normally distributed residual error (ε), was added to each sample to account for inter-individual variability and measurement error. The error was defined as

$$(\varepsilon = e^{\sigma Z}) \quad (2)$$

where σ is the standard deviation of the logarithm and Z is the standard normal deviate. Coefficients of variation of 5%, 10% or 15%, were investigated, by varying σ. The choice of CVs was based on earlier studies (Slob et al., 2005), where it was noted that median CV where around 10% in historical data analyzed with the benchmark dose approach.

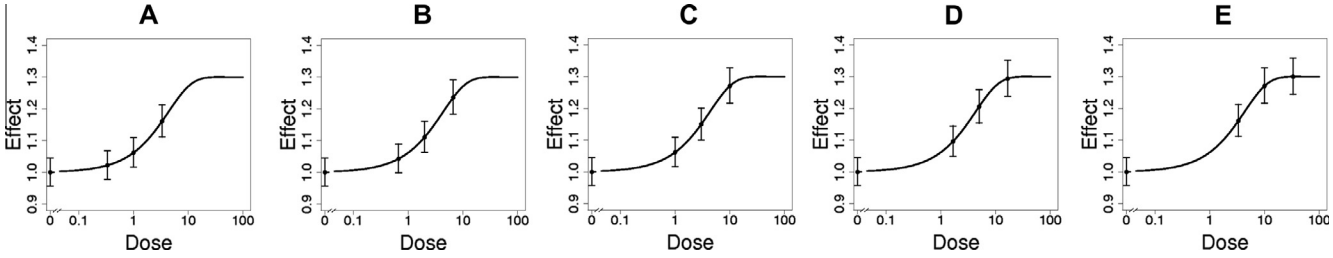

Fig. 1. Dose placement scenarios used in the Monte Carlo generation of dose–effect data. The dose placements ranged from the lowest dose placement (dose placement A) where the highest dose was equal to the EC_{50} to the highest dose placement (dose placement E) where the dose in the lowest dose group was equal to the EC_{50}. The error bars indicate the 95% confidence interval of the group means with 20 animals/group and a coefficient of variation (CV) of 10%. Data were also generated for 5, 10 and 50 animals/group and CVs of 5% and 15%.