Talanta 85 (2011) 1966–1973

Contents lists available at ScienceDirect

Talanta

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

# CALUX measurements: Statistical inferences for the dose–response curve

M. Elskens[a,*], D.S. Baston[a], C. Stumpf[b], J. Haedrich[b], I. Keupers[a], K. Croes[a], M.S. Denison[c], W. Baeyens[a], L. Goeyens[a]

[a] Laboratorium voor Analytische en Milieu Chemie, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium
[b] European Union Reference Laboratory for Dioxins and PCBs in Feed and Food, Bioassay Research Laboratory, State Institute for Chemical and Veterinary Analysis of Food, Bissierstrasse 5, 79114, Freiburg, Germany
[c] Department of Environmental Toxicology, 4241 Meyer Hall, One Shields Avenue, University of California, Davis, CA 95616, USA

ARTICLE INFO

*Article history:*
Received 20 December 2010
Received in revised form 22 June 2011
Accepted 7 July 2011
Available online 18 July 2011

*Keywords:*
Bioassay CALUX
Statistical inference
Dose–response curve
Box–Cox transformation

ABSTRACT

Chemical Activated LUciferase gene eXpression [CALUX] is a reporter gene mammalian cell bioassay used for detection and semi-quantitative analyses of dioxin-like compounds. CALUX dose–response curves for 2,3,7,8-tetrachlorodibenzo-p-dioxin [TCDD] are typically smooth and sigmoidal when the dose is portrayed on a logarithmic scale. Non-linear regression models are used to calibrate the CALUX response *versus* TCDD standards and to convert the sample response into Bioanalytical EQuivalents (BEQs). Several complications may arise in terms of statistical inference, specifically and most important is the uncertainty assessment of the predicted BEQ. This paper presents the use of linear calibration functions based on Box–Cox transformations to overcome the issue of uncertainty assessment. Main issues being addressed are (i) confidence and prediction intervals for the CALUX response, (ii) confidence and prediction intervals for the predicted BEQ-value, and (iii) detection/estimation capabilities for the sigmoid and linearized models. Statistical comparisons between different calculation methods involving inverse prediction, effective concentration ratios ($ECR_{20-50-80}$) and slope ratio were achieved with example datasets in order to provide guidance for optimizing BEQ determinations and expand assay performance with the recombinant mouse hepatoma CALUX cell line H1L6.1c3.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

Contaminating chemicals are known to enter the environment by many different pathways, e.g. industrial discharges and leakage, municipal waste production, agricultural run-off, forestry applications as well as accidents [1,2]. The US EPA has identified a diverse array of anthropogenic environmental contaminants [3] amongst them are halogenated aromatic hydrocarbons [HAHs], such as the polychlorinated dibenzo-p-dioxins [PCDDs], furans [PCDFs] and biphenyls [PCBs]. Man-made xenobiotics like HAHs are a significant global burden, in that they are highly toxic and their pronounced lipophilicity and resistance to metabolic degradation cause them to bioaccumulate in biota, and biomagnify in the food chain. Accordingly, the World Health Organization [WHO] is leading an initiative to provide reliable and accurate estimates of the global burden of all food-borne diseases caused by chemicals, parasites and enteric infections by 2012 [4]. Decreases in dioxin body burden have been repeatedly reported [5,6], however recent data seem to indicate that the decline has slowed down to a standstill [7]. Vigilance and an expanded focus on hitherto unidentified contaminants and their sources explain the need for efficient control strategies and analysis techniques for both environmental and food matrices [2,8]. Moreover, an integrated approach to identify and characterize hazards, to assess human exposure and to evaluate risks preferably relies on analytical techniques to quantify both the chemical contaminants' concentrations as well as their resulting biological (toxicological) effects [9]. While chemo-analytical methods, including rapid chromatographic screening techniques are currently used for dioxin analysis [8], there is a growing demand for the combined use of chemo- and bio-analytical approaches in order to link chemical contamination to measurable toxic and biological effects [2,10].

The Chemical Activated LUciferase gene eXpression [CALUX] is a recombinant mammalian cell bioassay based on the Ah receptor that is used for the detection and semi-quantitative analysis of dioxins and dioxin-like compounds [11]. The recombinant cells contain a stably transfected AhR-responsive firefly luciferase reporter gene, which responds to dioxin-like HAHs and AhR agonists with the induction of luciferase production [11,12]. The measured luminescence is converted into a bioanalytical potency estimate (BEQ) by comparison of the sample induction responses to a dose–response curve obtained with 2,3,7,8-tetrachlorodibenzo-p-dioxin [TCDD] standards. BEQ-values can be generated in several ways as reviewed in [13]. The underlying assumption of all of these techniques is that the sample induces a CALUX response through

* Corresponding author.
*E-mail address:* melskens@vub.ac.be (M. Elskens).

0039-9140/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2011.07.014

the same mechanism as the TCDD standard. These techniques were applied to a wide variety of matrices (reviewed in [11]), including, food and feed matrices [14], human fluids [15], environmental samples [16,17], and marine biological matrices [18].

The CALUX dose–response follows an S-shaped curve in a two-dimensional space, defined by an *X*-exposure axis [concentration or dose on a logarithmic scale] and a *Y*-response axis [luminescence signal on a linear scale]. This relationship may be equally well described by the Gaussian cumulative distribution or by logistic type models [19]. Yet fitting sigmoid dose–response curves with log transformed data as usual can generate inference problems [20,21], and hence make the BEQ interpretation difficult. To minimize these shortcomings, alternative linear and nonlinear regressions were examined and compared using Box–Cox transformations [22].

The objective of this paper is to determine key performance parameters (precision, accuracy and working range) for the different fitting approaches in order to optimize the BEQ determinations and to expand the assay performance. The mathematical and statistical bases for the regressions are described and illustrated by dose–response curves generated with the recombinant mouse hepatoma cell lines H1L6.1c3.

## 2. Materials and methods

### 2.1. Reagents and chemicals

Dimethylsulfoxide (DMSO) was purchased from Merck (Pro analysi ACS, Germany) and used to prepare stock and standard solutions of 2,3,7,8-tetrachlorodibenzo-p-dioxin (2,3,7,8-TCDD, 99% purity) as supplied by Campro Scientific (Cambridge Isotope Laboratories CIL-ED-901). Tissue culture media was purchased from Invitrogen (Life Science Technologies, UK), including α-MEM, FBS (South American Origin) as well as trypsin (0.25%; stock solution) and 10X PBS (Ambion).

### 2.2. Cell line and procedure

The stably transfected mouse hepatoma cell line H1L6.1c3 was used to evaluate TCDD standards ranging from 0.3 to 3000 pM to define a dose–response curve. Saturation of the Ah receptors initially present at the time of exposure, and formed during exposure of the cells, occurs at a concentration of ~1000 pM TCDD. The recombinant cell line was developed by stably transfecting a mouse hepatoma (hepa1c1c7) cell line with the plasmid pGudLuc6.1, which contains the Mouse Mammary Tumour Virus (MMTV) viral promoter and the firefly luciferase gene under AhR-dependent control of 4 native Dioxin Responsive Elements (DREs) [23]. Induction of the AhR signalling pathway is measured by quantifying the increase in luciferase activity and is expressed as Relative Light Units (RLUs). This activity is proportional to the concentration of AhR agonist(s) extracted from a sample.

Cells were grown in α-MEM (Gibco® Invitrogen, UK) supplemented with 10% FBS (South American; Gibco, UK) to near confluence (roughly 85–90% coverage of the tissue culture plate) and harvested by adding a 0.05% trypsin–EDTA solution. After centrifugation and suspension in culture media, cells were seeded into white clear bottom 96-well plates (Perkin Elmer, USA) by adding 100 μl of the suspension at a cell density of $7.5 \times 10^5$ cells/ml. Subsequently, cells were incubated for 24 h at 37 °C in 5% $CO_2$ and 80% relative humidity. Media was then removed and cells were dosed with a 100-fold dilution of TCDD standard solutions in DMSO. Dosing of the cells was accomplished by adding 100 μl to the wells in triplicate. All TCDD standard curve experiments consist of 8 to 10 points, a media blank and associated DMSO blanks. Environmental samples were analysed with 8–10 dilutions, allowing for identification and delineation of upper (or maximum) and lower sample extract dose curve plateaus. The procedure for sample extraction and clean-up is described in [17].

Cells were incubated for another 24 h at 33 °C, the wells were rinsed with 50 μl of 1X PBS (Gibco, UK), checked for viability, followed by addition of 50 μl of 1X lysis buffer (Promega, USA). Afterwards, wells were read using a Glomax 96-well plate luminometer (Promega, USA), or a Lucy 2 luminometer (Anthos, Germany), for 15 s, with a 5.6 s incubation (lag) time, after injection of 50 μl of the reconstituted luciferase assay buffer.

Method-performance studies were carried out on two independent datasets produced by the Bioassay Research Laboratory at the European Union Reference Laboratory for Dioxins and PCBs in Feed and Food (EURL, Freiburg) and the Laboratory for Analytical and Environmental Chemistry (ANCH, Brussels).

### 2.3. Fitting dose–response curves

In pharmacology, the 4-parameter Hill equation, which is mathematically analogous to the logistic equation [19], is often used to describe the effects of different dose levels of a xenobiotic. This model represents a significant improvement over the 3-parameter version because the background level ($y_0$) is included amongst the parameters instead of being considered a perfectly known constant:

$$y_i = y_0 + \frac{m \cdot x_i^h}{(EC_{50})^h + x_i^h} + \varepsilon_i \quad (1)$$

where $x_i$ is the TCDD concentration or sample dilution and $y_i$ the responsiveness normalized to the maximum possible responsiveness (efficacy). The parameter $y_0$ denotes the intercept or experimental background response, m the asymptotic maximum or efficacy, $h$ the Hill coefficient describing the steepness of the dose–response curve, $EC_{50}$ the concentration that produces a half-maximal response and $\varepsilon_i$ the stochastic or residual term for the $i$th observation. When plotted on a linear scale, Eq. (1) appears similar to a Langmuir adsorption isotherm (Fig. 1A) but plotted on a log scale, it is equivalent to a logistic function (Fig. 1B). Accordingly, the model parameters ($y_0$, $m$, $h$ and $EC_{50}$) can be estimated with the state variable ($x$) entered into a linear or a logarithmic form. The main difference between these options regards the statistical inference on the parameter estimates (see Fig. 1). For example, the outcome from the logarithmic regression $Log(EC_{50})$ must be converted to the corresponding $EC_{50}$ for further BEQ assessments (see Section 2.5). This switch from a log to a linear scale results in variance inflation and asymmetric confidence bounds on the fitted $EC_{50}$-value [20,21].

Alternatively, simple transformations of the response variable ($y$) are helpful to linearize non-linear regression models. A general rule proposed by [22] is to appropriately re-express curvilinear relationships (Fig. 2A) as linear ones by using transformations of the form:

$$z_i = \frac{y_i^\lambda - 1}{\lambda \cdot \left(\sqrt[n]{\prod_{i=1} y_i}\right)^{\lambda-1}} = \alpha \cdot x_i + \beta + \varepsilon_i \quad (2)$$

where $z_i$ represents the transformed variable, λ is a parameter determined from the data (Fig. 2B), $\alpha$ is the slope of the line and $\beta$ its intercepts on the $z$-axis. A Box–Cox procedure based on the method of maximum likelihood [20] was used to estimate λ and the other parameters for the linear regression model (Eq. (2)). A limitation of the Box–Cox transformation is its difficulty of fitting the entire dose–response curve into one unique model. The maximum likelihood estimation was, therefore, implemented for induction responses ranging from 6 to 75% of the maximum responsiveness

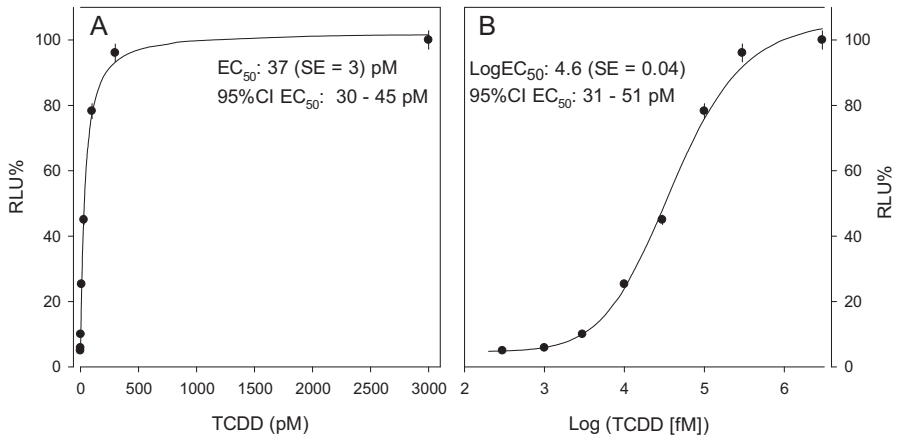


**Fig. 1.** Dose–response curves generated with the cell line H1L6.1c3. (A) Hill regressions with TCDD on a linear scale and (B) with TCDD on a on a log scale.

(Fig. 2). In all cases, scatter and residual plots were utilized to examine the appropriateness of the Box–Cox transformation. The relevance of the linear and nonlinear estimation procedures was investigated for the uncertainty assessment of the predicted BEQ (see Section 2.5).

### 2.4. Confidence intervals for the response variable

Accurate model prediction estimations are improved when the variances in the measurements are known. They are taken into account to compute the variance-covariance matrix of parameter estimates according to:

$$Var\text{-}Cov = (J^T \cdot W \cdot J)^{-1} \quad (3)$$

with $J$ the Jacobian and $W$ the weighting matrix, respectively. $J$ is the matrix of all first-order partial derivatives of the model with respect to the parameters, while $W$ is a square diagonal matrix in which the entries on the main diagonal are the predicted value of the variance at $y_i$ or $z_i$ (see Eqs. (1) and (2)). The variance on any point on the regression line is then given by a Taylor series expansion:

$$Var(\hat{y}_i \text{ or } \hat{z}_i) = \sum_{i=1}^{q} \left( \frac{\partial f(x_i)}{\partial p_j} \right)^2 \cdot Var(p_j) + 2 \cdot \sum_{\substack{j,l=1 \\ j \neq l}}^{q} \frac{\partial f(x_i)}{\partial p_j} \cdot \frac{\partial f(x_i)}{\partial p_l} \cdot Cov(p_j, p_l) \quad (4)$$

where $Var(p_j)$ represents the variance on the parameter estimates and $Cov(p_j,p_l)$, the covariance between parameters (see Eq. (3)).

For direct comparison of a sample response with a TCDD dose–response curve, the uncertainty on the sample value $u$(sample) depends on the variances of the regression line (Eq. (4)) and of the sample. The variances being independent are additive:

$$u(\text{sample}) = \sqrt{Var(\text{sample}) + Var(\hat{y}_i \text{ or } \hat{z}_i)} \quad (5)$$

### 2.5. Assessment of the sample potency (BEQ)

Depending of the sample source (either environmental or biological matrices) BEQ-values were assessed using one of the following methods:

(i) *Inverse prediction* or conversion of the sample response using the TCDD dose–response curve. The method, frequently used for the screening of biological matrices [15,24], is based on a single point estimate, and assumes that the sample extract acts merely as a dilution of TCDD. Ideally, the sample extract must be diluted to provide a response that lies between 20 and 70% of the maximal induced TCDD response. The extract potency is calculated by inverting the regression models described by Eqs. (1) and (2). The uncertainty range on the potency estimate is calculated by substituting $y_i$ for $y_{\text{sample}} \pm u$(sample) in the inverse models.

(ii) *Effective concentration ratios*. This method is often used for assessing the potency of environmental samples (reviewed in [13]). The $EC_{50}$-values are directly generated from the Hill regression of the TCDD standards and sample extract

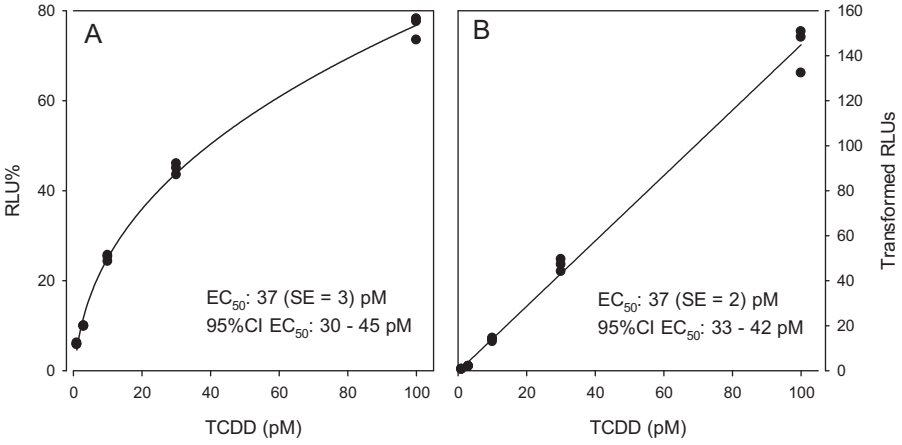


**Fig. 2.** Dose–response curves generated with the cell line H1L6.1c3. (A) Hill regression with TCDD on a linear scale and (B) linear regression after Box–Cox transformation.

dose–response curves. The potency is then assessed as the ratio $[EC_{50}]_{STANDARDS}$ over $[EC_{50}]_{SAMPLES}$. If the confidence intervals on the $EC_{50}$-estimates are symmetric (Hill regression on a linear scale), the prediction interval on the relative potency may be determined analytically. For asymmetric confidence limits (Hill regression on a log scale), the error propagation becomes too complex to derive an analytic solution and the prediction interval was assessed with Monte-Carlo simulations. When the dose–response curves for TCDD and the sample extract are not parallel, a relative potency range can be determined using various EC-TCDD to sample ratios (e.g. $EC_{20}$ and $EC_{80}$). The spread of the potency range defined as the minimum to the maximum potency can then be used as a measure of the confidence for the result [25].

(iii) *Slope ratio*. Another comparison for calculating BEQs is based on a modification of the slope ratio technique described by [26]. By comparing slopes of the Box–Cox transformations for the TCDD and the sample dose–response curves, the potency can be assessed as the ratio of the sample to TCDD slopes. Since the confidence intervals on the slope estimates are symmetric, the prediction interval on the relative potency may be determined analytically.

## 3. Results and discussion

### 3.1. Background noise

The background noise is defined as the RLU (relative light unit) level below which it is difficult to distinguish an actual response signal from a random noise. It was assessed from the mean response signal and standard deviation obtained with the solvent (DMSO) according to IUPAC definitions [27]. The sampling distribution of the solvent response is normal (Shapiro–Wilk test) resulting in a critical signal level of ~6.0 and a minimum detectable signal of ~7.3% of the TCDD efficacy (Fig. 3). These limits are independent from the model applied to fit the dose–response curves.

### 3.2. Expected precision on response signals for TCDD standards

The Horwitz curve is a useful generalization of the reproducibility standard deviations expected in collaborative trials [28,29]. The form of this curve is a power relationship between the standard deviations amongst laboratories to concentration expressed in mass per mass units. Although in CALUX, the measurand is expressed in relative light units (RLUs), the power law was used to establish a functional relationship between precision values and the response signal. The results are illustrated in Fig. 4. The best fit appears as a straight line in a log–log plot. In the interlaboratory assessment, neither slopes nor intercepts significantly differ on a 95% confidence level (two-sided *t*-test). Slopes are significantly different from 0, which are typical of heteroscedastic data, and R-squared suggests that >60% of the variance in the standard deviation can be explained by the regressions. Hence the following expression of the reproducibility standard deviation ($s_R$) can be derived for the entire dataset:

$$s_R = 0.05 \cdot (\text{RLU\%})^{0.87} \quad (6)$$

Eq. (6) is needed to fix the weighting matrix (Eq. (3)) that is subsequently used to fit the dose–response models, and is applicable as a functional performance criterion for others applying the CALUX methodology. For example, the coefficient of variation or expected precision on TCDD replicates ($CV = 100 S_R/\text{RLU\%}$) should be less than 5% on an average basis, and the 95% prediction bands indicate that the probability of observing a CV greater than 20% is less than 0.05 (see Fig. 4).

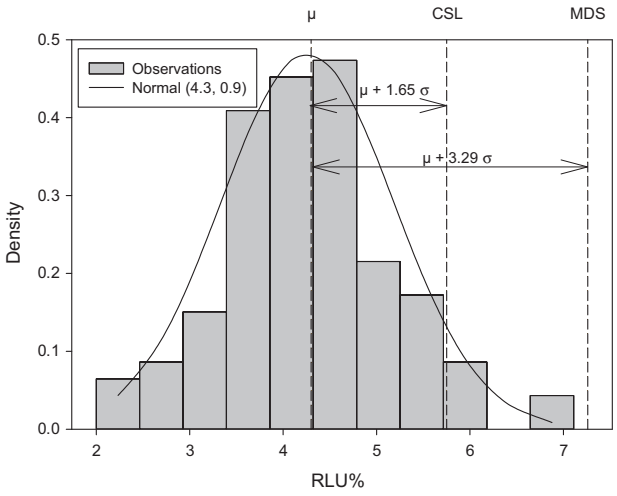


**Fig. 3.** Probability distribution function for the solvent DMSO signal (Shapiro–Wilk test; $p = 0.313$; $n = 100$). The critical signal level (CSL) and the minimum detectable signal (MDS) were determined according to IUPAC definition with $\alpha = \beta = 0.05$.

### 3.3. Estimation properties of the dose–response models

EC-results obtained when fitting dose–response curves are moderate-tailed distributions with 95% confidence intervals

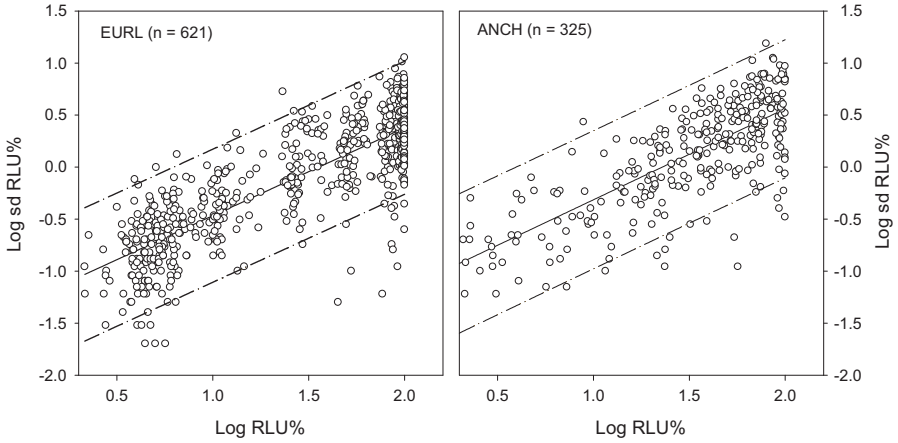


**Fig. 4.** Relationships between the standard deviation and the response signal. The equation corresponding to the fitted line is given by Log sd RLU = −1.31 (SE = 0.035) + 0.85 (SE = 0.024) for the EURL data set and by: Log sd RLU = −1.19 (SE = 0.065) + 0.87 (SE = 0.041) for the ANCH data set. Dashed dotted lines represent the 95% prediction intervals.

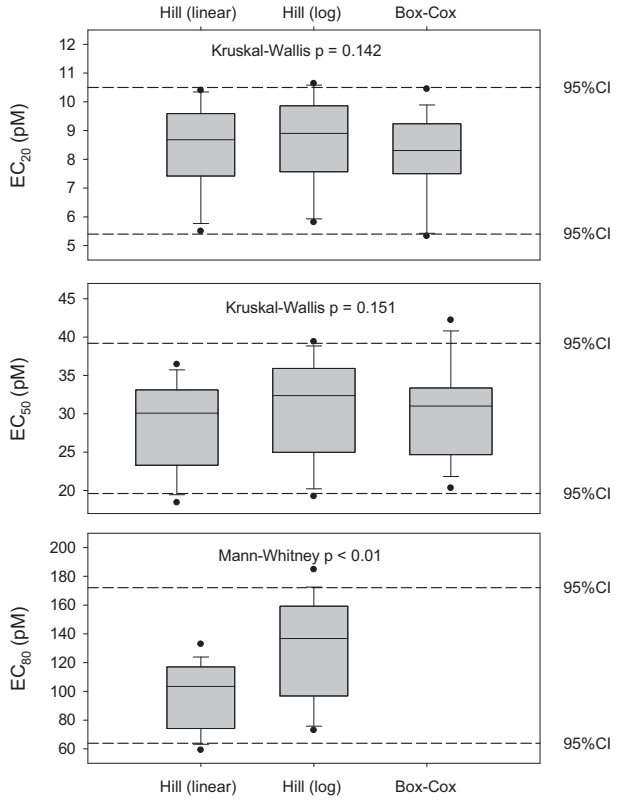


**Fig. 5.** EC-values estimated from TCDD dose–response curves using linear and non-linear regression models.

extending from 5 to 10, 20 to 40 and 64 to 172 pM TCDD for $EC_{20}$, $EC_{50}$ and $EC_{80}$, respectively (Fig. 5). This unpredictability is due to variability in the biological response, an inherent property of the studied system, and uncertainty, the lack of knowledge about the studied system (randomness and model parameterization). For the linear model, the effective prediction range ($EC_{20-50}$) is limited by the Box–Cox transformations (see Section 2.3). Statistically, there was no significant difference for the $EC_{20-50}$ values between the various models (Kruskall–Wallis test) but the $EC_{80}$ results appeared systematically higher for the Hill regression on a log scale (Mann–Whitney test). Uncertainty on EC determinations is calculated by substituting $y_i$ for $y_i \pm u(y_i)$ in Eqs. (1) and (2), and solving for $x_i$. Accordingly, model results may be compared on the basis of their respective CV, a dimensionless measure of uncertainty. Confidence bounds for the CVs are calculated according to [30] and reported in Table 1. On the whole the CVs for the Hill function appear significantly greater than the ones for the Box–Cox transformation indicating that the linear model is more precise for EC assessment.

**Table 1**
Model estimation properties.

| | Hill (linear scale) | Hill (log scale) | Box–Cox |
|---|---|---|---|
| Goodness of fit | | | |
| $SSR/\chi^2_{95\%}$ | 0.5 (0.1–1.4)[a] | 0.4 (0.2–1.2)[a] | 0.7 (0.1–1.3)[a] |
| CV% on | | | |
| $EC_{20}$ | 7.1 (5.8–9.2)[b] | 6.6 (5.4–8.6)[b] | 4.1 (3.3–5.3)[b] |
| $EC_{50}$ | 7.2 (5.9–9.4)[b] | 8.7 (7.1–11)[b] | 4.3 (3.5–5.6)[b] |
| $EC_{80}$ | 13 (11–17)[b] | 19 (15–25)[b] | NA |
| Fitted parameters | | | |
| Hill coefficient | 1.1 (1.0–1.2)[a] | 10.5 (10.2–11.2)[a] | NA |
| Lambda | NA | NA | 1.7 (1.4–2.2)[a] |

SSR = sum of squared residuals; NA = not applicable; CV = coefficient of variation.
[a] Mean (min–max).
[b] Mean (exact 95% confidence intervals).

Additionally, it is an essential aspect of the analysis to test the appropriateness of the overall models. The statistical evaluation of the goodness of fit was based on a comparison between the Sum of Squared Residuals (SSR) and the theoretical $\chi^2$ distribution from which SSR should be derived [31]. SSR is a measure of the discrepancy between the data and the dose–response model. A small SSR indicates a tight fit of the model to the data. For a valuable fit with assumptions satisfied, SSR must be lower than the critical $\chi^2$ value (upper one-tail 95%, $p > 0.05$), while SSR $> \chi^2_{99\%}$ indicates substantial lack of fit or some failures of assumptions ($p < 0.01$). In Table 1, it can be seen that the SSR to $\chi^2_{95\%}$ ratios were less than 1 in most cases, indicating that model regressions are statistically significant, i.e. around 9 and 1% of the SSR-values exceed the 95 and 99th percentiles of the $\chi^2$ distribution, respectively. This is well acceptable since the calibration curves do not represent realizations of the same experiment, but corresponds to calibration curves for runs carried out during a one-year period. The typical shape of the TCDD dose–response curve appears hyperbolic rather than sigmoidal, the Hill coefficient being very close to 1 (Table 1). Parameterized in terms of log(X), the Hill function follows a sigmoidally shaped curve (Hill coefficient close to 10). A general use of this transformation is mathematically not required, and even not recommended here since it does not improve the goodness of fit and deteriorates the estimation properties of the r egression model, especially for EC-prediction in the range of $EC_{80}$ (Table 1). On the other hand, the Box–Cox transformation helps to linearize the curvilinear dose–response function without impairing the assumptions of independence, identicalness and normality of the error distribution [32]. For the cell line H1L6.1c3, the transformation shows linear responses with TCDD doses up to 100pM (see Fig. 2), beyond this value, departures from the error distribution assumptions were detected. Despite this limitation, the linearization improves the estimation properties enabling a rigorous statistical analysis (Table 1).

The detection capability (DC) is calculated by substituting $y_i$ for the minimum detectable signal in Eqs. (1) and (2), and solving for $x_i$. While DC differs slightly for different calibrations and model approaches, its distribution was unimodal and the experimental design (including the number of replicates) was identical for each of the calibrations and models. Accordingly the median of the distribution equal to ~1.6 pM TCDD was used as the detection capability for the cell line H1L6.1c3. This limit, valid for the standard curves, is independent from the model applied to fit the dose–response curves.

### 3.4. Example data sets

Five example data sets from environmental extracts were used to compare the fitting approaches in estimating BEQ-values. The dose–response curves for those test samples were compared to a TCDD standard curve to calculate a potency factor (pg BEQ/g) as described in Section 2.5. In Fig. 6, it can be seen that: S1 is almost parallel to TCDD with comparable efficacy (maximal response, $m$) and steepness (Hill coefficient, $h$); S2 lacks a finite upper asymptote, and is characterized by a putative supra maximal induction ($m_{sample} \geq 20\%\ m_{TCDD}$); S3 seems typical of antagonism where the maximal response is not observed ($m_{sample} \leq m_{TCDD}$); S4 exhibits a shallower response than TCDD ($h_{sample} \leq 20\%$ of $h_{TCDD}$) but with comparable efficacy; S5 is characterized by a steeper response ($h_{sample} \geq 20\%$ of $h_{TCDD}$) with a well-define maximum but for a higher sediment mass the sample response drastically decreases. The estimated standard deviation on those sample measurements

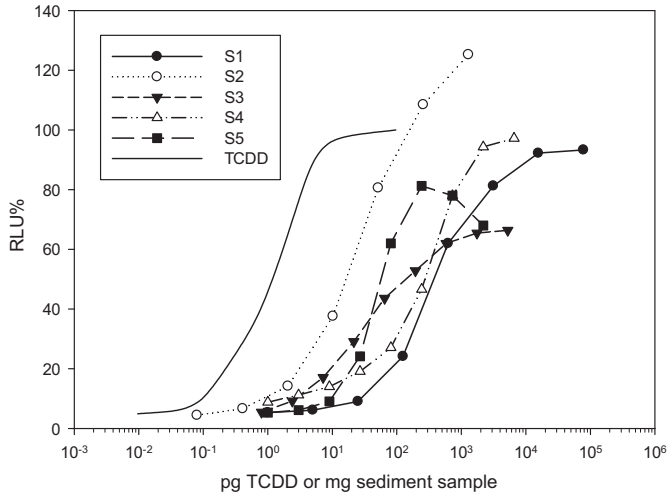


**Fig. 6.** TCDD and sample dose–response curves generated with the cell line H1L6.1c3.

($s_{sample}$) was estimated with a power law as previously performed for TCDD standards:

$$s_{sample} = 0.23 \cdot (RLU\%)^{0.71} \quad (7)$$

Eq. (7) shows that the reproducibility standard deviation estimated for the environmental sample extracts are higher than for the TCDD standards. The 95% CV prediction bands for the regression lies between 5 and 25%. BEQ estimates are shown in Fig. 7 and Table 2. In cases where m and h of both samples and TCDD do not markedly differ, all models and calculation methods provide BEQs that are not statistically different (e.g. S1). In the other situations (S2–S5) substantial differences are observed amongst calculation methods but not between model results (Table 2). So far the hypothesis that the BEQ distributions for sample S1–S5 are derived from normal populations cannot be rejected (Kolmogorov–Smirnov test). The potency range, calculated using the effective concentration ratios (ECR), increases as the difference between the steepness of the TCDD and sample dose–response curves increases with $ECR_{20} > ECR_{80}$ for $h_{sample} < h_{TCDD}$ while it is not influenced markedly by differences between the sample and TCDD efficacy. Two extreme $ECR_{80}$-values (|*z*-score| > 2) were obtained using the Hill regression on a log scale (Table 2). This corroborates our previous findings with TCDD curves regarding the relevance of some $EC_{80}$-values forecasted with a Hill fit on a log scale (see Table 1 and Fig. 5). Thus as long as the Hill fit remains satisfactory, the ECR approach provides robust BEQ estimates. Under these conditions, with only few exceptions, the expected uncertainty on a BEQ estimate should at least be smaller than 25% (Table 2). It is noted that the $ECR_{50}$ is close to the overall BEQ mean calculated for each sample (Table 2), and that the BEQ based on the slope ratio after Box–Cox transformation always lies in the $ECR_{20-80}$ range (Fig. 7). Inverse prediction is based on a single point estimate and is influenced both by differences in the steepness and efficacy of the TCDD and sample dose–response curves. Compared to the ECR approach, the BEQ will be overestimated in situations where $m_{sample} > m_{TCDD}$ and $h_{sample} > h_{TCDD}$ and reversely (Fig. 7). Practically this bias remains unnoticed since $m_{sample}$ and $h_{sample}$ are unknown when assessing a BEQ from a TCDD calibration curve, but it may contribute significantly to the overall unpredictability of the method. For safety performance it was usually recommended to take a sample dilution providing a response signal that lies between 20 and 70% of the maximal induced TCDD response [15]. Apparently that was not enough to warrant the BEQ result. In Fig. 8, it can be seen that the estimated BEQ for S3 (1–25 pg/g) highly depends on the dilution that is used through the inverse prediction method. This unpredictability in BEQ appears substantially larger than the propagated uncertainty on the model results (Eq. (5)). Here again the results are influenced by the calculation method and not by the dose–response model (Box–Cox *versus* Hill). The interpretation of BEQ via single point estimate has therefore to be considered carefully, and requires testing the fundamental assumption that sample extracts behaves similarly to TCDD dilutions. Rigorous statistical analyses for comparing dose–response curves in terms of basal and maximal responses, curve shape and steepness are outlined in [19].

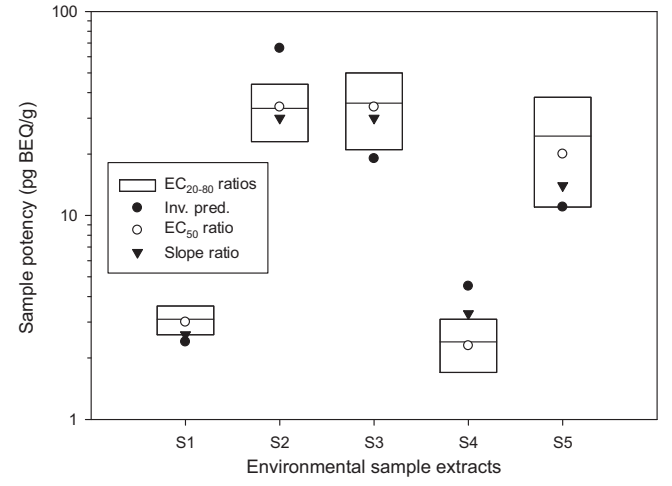


**Fig. 7.** Potency range for the five example datasets from environmental extracts.

**Table 2**
Potency estimates for environmental sample extracts (pg TCDD/g sediment).

| Sample | Hill (Linear scale) | Hill (log scale) | Box–Cox |
|---|---|---|---|
| S1 (overall mean ± sd) | | 2.8 ± 0.5 | |
| Inverse prediction | 2.5 ± 0.4 | 2.4 ± 0.4 | 2.5 ± 0.3 |
| $EC_{20}$-ratio | 2.6 ± 0.3 | 2.7 ± 0.3 | 2.4 ± 0.1 |
| $EC_{50}$-ratio | 2.8 ± 0.3 | 3.1 ± 0.5 | 3.0 ± 0.2 |
| $EC_{80}$-ratio | 3.1 ± 0.3 | 4.1 ± 1.5[a] | NA |
| Slope ratio | NA | NA | 2.6 ± 0.2 |
| S2 (overall mean ± sd) | | 42 ± 16 | |
| Inverse prediction | 69 ± 10 | 69 ± 10 | 59 ± 6 |
| $EC_{20}$-ratio | 45 ± 6 | 42 ± 6 | 46 ± 3 |
| $EC_{50}$-ratio | 34 ± 5 | 30 ± 7 | 39 ± 3 |
| $EC_{80}$-ratio | 26 ± 7 | 20 ± 13 | NA |
| Slope ratio | NA | NA | 30 ± 2 |
| S3 (overall mean ± sd) | | 32 ± 13 | |
| Inverse prediction | 19 ± 3 | 18 ± 3 | 21 ± 2 |
| $EC_{20}$-ratio | 50 ± 8 | 43 ± 6 | 56 ± 4 |
| $EC_{50}$-ratio | 32 ± 5 | 32 ± 6 | 39 ± 3 |
| $EC_{80}$-ratio | 20 ± 6 | 23 ± 9 | NA |
| Slope ratio | NA | NA | 30 ± 2 |
| S4 (overall mean ± sd) | | 3 ± 1 | |
| Inverse prediction | 4.5 ± 0.7 | 4.3 ± 0.7 | 4.7 ± 0.5 |
| $EC_{20}$-ratio | 3.4 ± 0.6 | 3.0 ± 0.6 | 3.0 ± 0.2 |
| $EC_{50}$-ratio | 2.3 ± 0.4 | 2.0 ± 0.7 | 2.6 ± 0.2 |
| $EC_{80}$-ratio | 1.6 ± 0.6 | 1.7 ± 2.0 | NA |
| Slope ratio | NA | NA | 3.4 ± 0.3 |
| S5 (overall mean ± sd) | | 18 ± 10 | |
| Inverse prediction | 11 ± 2 | 12 ± 2 | 11 ± 1 |
| $EC_{20}$-ratio | 12 ± 1 | 12 ± 1 | 10 ± 0.6 |
| $EC_{50}$-ratio | 19 ± 2 | 21 ± 3 | 20 ± 1 |
| $EC_{80}$-ratio | 31 ± 6 | 45 ± 15[a] | NA |
| Slope ratio | NA | NA | 14 ± 0.8 |

NA: not applicable.
[a] |*z*-score| > 2.

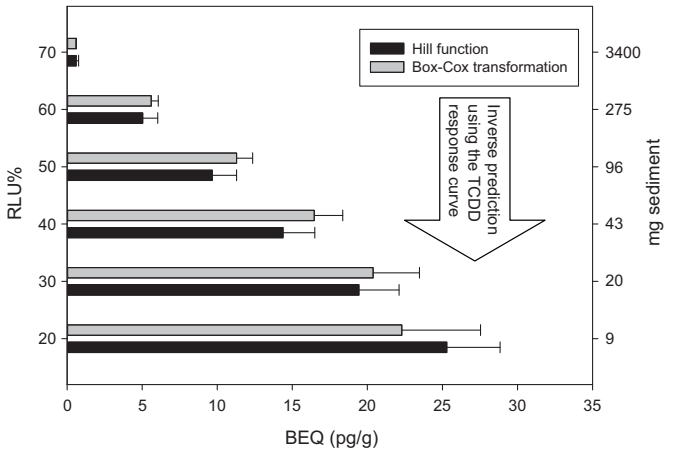


**Fig. 8.** BEQ assessment for sample S3. The extracts were diluted to provide a signal response that lies between 20 and 70% of the maximal induced TCDD response. The sample response were subsequently converted into BEQ (pg/g) by inverting the TCDD dose–response curve corresponding to a Hill fit and/or a Box–Cox transformation. The error bar represents the propagated uncertainty on the BEQ estimate (Eq. (5)).

## 4. Conclusions

It is a must that statistical procedures used for bio-analytical methods provide the same appropriately developed information on precision and accuracy of the measurements as for instrumental analysis. They should reflect the technical advantages of the methods, especially the ability to cover a great range of concentration equivalents and include information about detection capabilities. This study investigated the CALUX calibration methodology for equivalence estimation, and provides the following guidance for the cell line H1L6.1c3:

(a) The reproducibility standard deviations for TCDD and sample dose–response curves can be depicted by power laws, which may serve as a performance criterion for internal quality controls.
(b) The assay-LOD was based upon the analysis of solvent blank distributions according to IUPAC definitions. This unbiased assessment gives rise to a detection capability (DC) by inverse modelling. DC exhibits unimodal distributions with a median amounting to 1.6 pM, which can be seen as a threshold value beyond which one may assume that is not simply due to random variation (i.e. noise). To be noted that this assay-LOD, valid for a standard curve, does not correspond to a sample related LOD.
(c) The log transformation of the concentration or dose–variable in the sigmoidal model is useless and must be discouraged since it deteriorates the estimation behaviour of the regression and ultimately the estimated value.
(d) Sampling distributions of EC50 are normal with an overall mean ~30 pM and 95% confidence intervals ranging from 20 to 40 pM TCDD. The EC50 value differs for different calibrations with a median uncertainty of ~7%. These statistics derived from a set of TCDD calibration curves carried out over a 1-year period may serve as performance criteria for internal quality controls.
(e) Alternative calibration functions were implemented using Box–Cox transformations of the response variable. These linearized models exhibited a comparable accuracy in the determination of EC-values to that of the sigmoidal model, but an overall better precision in BEQ assessment.

Conclusions (a), (c) and (e) are very generic and can be applied to other cell lines. Conclusions (b), and (d) are application specific and under study for the new generation H1L7.5c1 cell line. Conclusion (e) offers the opportunity to improve the confidence that can be placed on the result in a TCDD concentration/dose range of 1.6–100 pM TCDD. It is a useful alternative for dose–responses curves, which do not exhibit a lower and/or upper plateau, and for which the Hill fit is inaccurate. Under these conditions the BEQ assessment can be performed using inverse prediction or using the slope ratio method; the latter being more robust as based on multi-point estimates. The slope ratio method after Box–Cox transformations may also be suitable if considerations of cost, time, or samples limit the total number of experiments that can be performed (e.g. analysis of food and feed matrices as well as of human fluids). For example, a dose–response curve, fitted with a four parameter Hill equation required 3 replicates × 8 points, thus a minimum of 48 wells per plate are needed to estimate only one sample potency. With the same amount of wells and a fitting procedure based on a Box–Cox transformation (3 replicates × 4 points), the potency of 3 samples could at least be assessed.

## Acknowledgements

This work was financed by the Belgian Science Policy, Interuniversity Attraction Poles Programme TIMOTHY-P6/13. Further, we would like to thank the European Commission for the financial support of the work of the European Union Reference Laboratory for Dioxins and PCBs in Feed and Food, Freiburg (Germany).

## References

[1] T. Kosjek, E. Heath, M. Petrović, D. Barceló, Trends Anal. Chem. 26 (2007) 1076–1085.
[2] C. Blasco, Y. Picó, Trends Anal. Chem. 28 (2009) 745–757.
[3] US EPA, Science and Decisions: Advancing Risk Assessment, 2008.
[4] WHO, Initiative to Estimate the Global Burden of Foodborne Diseases, 2008, http://www.who.int/foodsafety.
[5] A.K.D. Liem, P. Fürst, C. Rappe, Food Addit. Contam. 17 (2000) 241.
[6] S.M. Hays, L.L. Aylward, Regul. Toxicol. Pharmacol. 37 (2003) 202.
[7] I. Windal, S. Vandevijvere, M. Maleki, S. Goscinny, C. Vinkx, J.F. Focant, G. Eppe, V. Hanot, J. Van Loco, Chemosphere 7 (2010) 584.
[8] W. Baeyens, F. Verstraete, L. Goeyens, Talanta 63 (2004) 1095.
[9] J.L.C.M. Dorne, L.R. Bordajandi, B. Amzal, P. Ferrari, P. Verger, Trends Anal. Chem. 28 (2009) 695.
[10] G. Streck, Trends Anal. Chem. 28 (2009) 635.
[11] M.S. Denison, B. Zhao, D.S. Baston, G.C. Clark, H. Murata, D. Han, Talanta 63 (2004) 1123.
[12] P.M. Garrison, K. Tullis, J.M.M.J.G. Aarts, A. Brouwer, J.P. Giesy, M.S. Denison, Fundam. Appl. Toxicol. 30 (1996) 194.
[13] J.J. Whyte, C.J. Schmitt, D.E. Tillitt, Crit. Rev. Toxicol. 34 (2004) 1.
[14] L. Hoogenboom, L. Goeyens, S. Carbonnelle, J. van Loco, H. Beernaert, W. Baeyens, W. Traag, T. Bovee, G. Jacobs, G. Schoeters, Trends Anal. Chem. 25 (2006) 410.
[15] N. Van Wouwe, I. Windal, H. Vanderperren, G. Eppe, C. Xhrouet, A.-C. Massart, N. Debacker, A. Sasse, W. Baeyens, E. De Pauw, F. Sartor, H. Van Oyen, L. Goeyens, Talanta 63 (2004) 1157.
[16] M. Nording, M.S. Denison, D. Baston, Y. Persson, E. Spinnel, P. Haglund, Environ. Toxicol. Chem. 26 (2007) 1122.
[17] H. Sanctorum, I. Windal, V. Hanot, L. Goeyens, W. Baeyens, Arch. Environ. Contam. Toxicol. 52 (2007) 317.
[18] I. Windal, N. Van Wouwe, G. Eppe, C. Xhrouet, V. De Backer, W. Baeyens, E. De Pauw, L. Goeyens, Environ. Sci. Technol. 39 (2005) 1741.
[19] A. De Lean, P.J. Munson, D. Rodbard, Am. J. Physiol. 235 (1978) 97.
[20] J. Neter, M.H. Kutner, C.J. Nachtsheim, W. Wasserman, Applied Linear Statistical Models, fourth ed., McGraw-Hill, 1996.
[21] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics 20A, Elsevier, 1997.
[22] F. Mosteller, J.W. Tukey, In Data Analysis and Regression, Addison-Wesley, 1977.
[23] D. Han, S.R. Nagy, M.S. Denison, Biofactors 20 (2004) 11–22.
[24] I. Van Overmeire, J. Van Loco, P. Roos, S. Carbonnelle, L. Goeyens, Talanta 63 (2004) 1241.
[25] D.L. Villeneuve, A.L. Blankenship, J.P. Giesy, Environ. Toxicol. Chem. 19 (2000) 2835.

[26] D.J. Finney, Probit Analysis, Cambridge University Press, Cambridge, 1964.
[27] L.A. Currie, Pure Appl. Chem. 67 (1995) 1699–1723.
[28] W. Horwitz, J. Assoc. Off. Anal. Chem. 66 (1983) 1295–1301.
[29] M. Thompson, Analyst 125 (2000) 385–386.
[30] W. Reh, B. Sheffler, Comp. Stat. Data Anal. 22 (1996) 449.
[31] M. Elskens, A. de Brauwere, W. Baeyens, C. Beucher, R. Corvaisier, N. Savoye, P. Treguer, Mar. Chem. 106 (2007) 272–286.
[32] D.J. Finney, Int. Stat. Rev. 47 (1977) 1–12.