

# Statistical strategies for averaging EC50 from multiple dose–response experiments

Xiaoqi Jiang · Annette Kopp-Schneider

Received: 4 July 2014 / Accepted: 25 August 2014 / Published online: 8 October 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** In most dose–response studies, repeated experiments are conducted to determine the EC50 value for a chemical, requiring averaging EC50 estimates from a series of experiments. Two statistical strategies, the mixed-effect modeling and the meta-analysis approach, can be applied to estimate average behavior of EC50 values over all experiments by considering the variabilities within and among experiments. We investigated these two strategies in two common cases of multiple dose–response experiments in (a) complete and explicit dose–response relationships are observed in all experiments and in (b) only in a subset of experiments. In case (a), the meta-analysis strategy is a simple and robust method to average EC50 estimates. In case (b), all experimental data sets can be first screened using the dose–response screening plot, which allows visualization and comparison of multiple dose–response experimental results. As long as more than three experiments provide information about complete dose–response relationships, the experiments that cover incomplete relationships can be excluded from the meta-analysis strategy of averaging EC50 estimates. If there are only two experiments containing complete dose–response information, the mixed-effects model approach is suggested. We subsequently provided a web application for non-statisticians to implement the proposed meta-analysis strategy of averaging EC50 estimates from multiple dose–response experiments.

**Keywords** Log-logistic model · EC50 · Mixed-effects modeling · Meta-analysis · Data visualization

## Introduction

The half-maximal effective concentration (EC50) as an important measure used in toxicology refers to the concentration of a chemical that produces half-maximal response. In vitro dose–response data typically exhibit a sigmoidal relationship between concentrations of a chemical and response measurements in the assay, which can be fitted by the four-parameter log-logistic (4pLL) model (e.g., Finney 1976) that includes EC50 directly as one of the model parameters. The 4pLL model function describing the dose–response data ( $x, y$ ) is given by

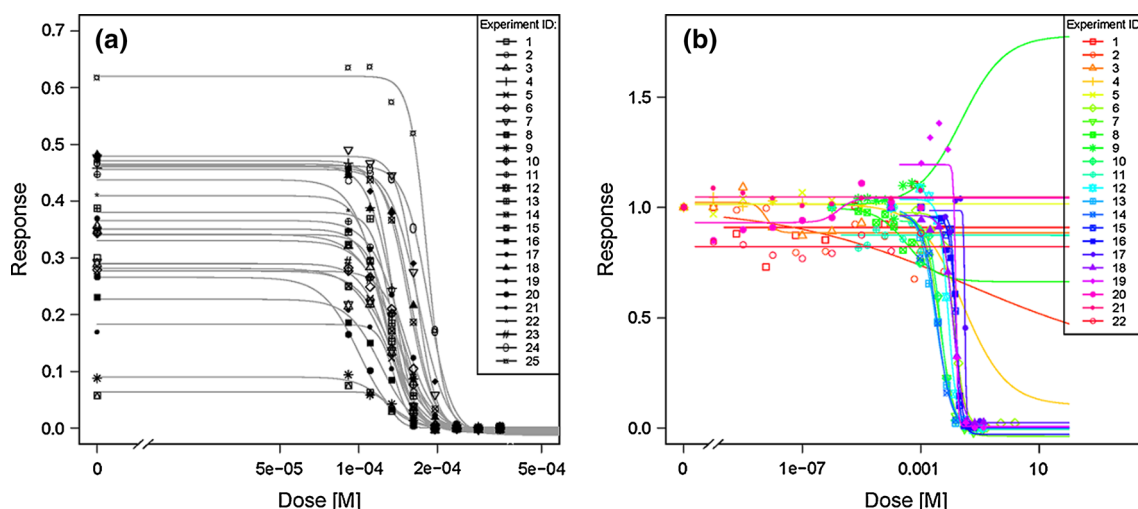
$$y = f(x) = \phi^{(c)} + \frac{\phi^{(d)} - \phi^{(c)}}{1 + \exp\{\phi^{(b)}[\log(x) - \log(\phi^{(e)})]\}} \quad (1)$$

The parameters  $\phi^{(c)}$  and  $\phi^{(d)}$  correspond to the lower and upper horizontal limits for mean response, respectively. The parameter  $\phi^{(e)}$  is the effective concentration EC50. The parameter  $\phi^{(b)}$  determines the slope of the curve. The positive sign of the parameter  $\phi^{(b)}$  indicates a decreasing curve, and the negative sign indicates an increasing curve.

Often in practical dose–response studies, more than one experiment is performed for a chemical to estimate EC50 values. The individual EC50 estimates vary to some extent from experiment to experiment, even under an identical experimental design. In this case, summarizing EC50 values from a series of experiments requires statistical support to obtain an average EC50 value with a corresponding confidence interval over all experiments. Sometimes, for a given experiment, the maximum tested concentration level

X. Jiang · A. Kopp-Schneider (✉)  
Division of Biostatistics, German Cancer Research Center  
(DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany  
e-mail: kopp@dkfz.de

X. Jiang  
e-mail: xiaoqi.jiang@dkfz.de



**Fig. 1** Illustrations of two types of multiple dose–response experiments. The data were selected from real dose–response experiments. Dose–response curves of each experiment were fitted using four-parameter log-logistic models. The various symbols in each plot

might not be high enough to observe the lower (or upper) plateau of a decreasing (or an increasing) dose–response curve. The ranges of tested concentrations can differ in multiple dose–response experiments. This may lead to problems with the estimation of individual EC50 values in experiments in which a dose–response relationship cannot be observed completely. Based on whether this situation is present, a dose–response study of multiple experiments can be classified into two types, as illustrated in Fig. 1: (a) complete dose–response relationships in all experiments; and (b) the combination of complete and incomplete dose–response relationships in multiple experiments.

For type (a) complete dose–response relationships, Jiang and Kopp-Schneider (2014) compared two statistical methods for averaging EC50 values: the mixed-effects modeling and the meta-analysis approaches. Using the mixed-effects model, the average EC50 value over all experiments is estimated by considering the variabilities within and among experiments simultaneously. For  $k$  experiments, the mixed-effects version of the 4pLL model function for the repeated response  $y_{ij}$  measured in the experiment  $i$  ( $i = 1, \dots, k$ ) at the  $j$ th ( $j = 1, \dots, n_j$ ) concentration  $x_{ij}$  is expressed as

$$y_{ij} = f(x_{ij}) = \phi_i^{(c)} + \frac{\phi_i^{(d)} - \phi_i^{(c)}}{1 + \exp\{\phi_i^{(b)}[\log(x_{ij}) - \log(\phi_i^{(e)})]\}} \\ = (\beta^{(c)} + b_i^{(c)}) + \frac{(\beta^{(d)} + b_i^{(d)}) - (\beta^{(c)} + b_i^{(c)})}{1 + \exp\{(\beta^{(b)} + b_i^{(b)})[\log(x_{ij}) - \log(\beta^{(e)} + b_i^{(e)})]\}} \quad (2)$$

where  $\phi_i^{(b)}$ ,  $\phi_i^{(c)}$ ,  $\phi_i^{(d)}$ , and  $\phi_i^{(e)}$  are experiment-specific parameters and the fixed effects  $\beta^{(b)}$ ,  $\beta^{(c)}$ ,  $\beta^{(d)}$ , and  $\beta^{(e)}$  represent the mean values of the corresponding experiment-specific

represent the mean responses for every concentration in each experiment. (a) Complete dose–response relationships in multiple experiments; (b) combination of complete and incomplete dose–response relationships in multiple experiments

parameters. The random effects  $b_i^{(b)}$ ,  $b_i^{(c)}$ ,  $b_i^{(d)}$ , and  $b_i^{(e)}$  represent the deviations of the experiment-specific parameter from their mean values. The random effects are assumed to be independent for different experiments and distributed normally with mean 0 and variance–covariance matrix  $\Psi$ . The model can be reduced by fixing some of the four random effects to zero, if the corresponding experiment-specific parameters do not vary from experiment to experiment. Averaging the EC50 values based on the mixed-effects model function (2) corresponds to estimation of the fixed-effects parameter  $\beta^{(e)}$  and its standard error (se). The 95 % Wald-type confidence interval can be constructed to report uncertainty of the mean EC50 value over all experiments. As degrees of freedom for t distribution  $\sum_{i=1}^k n_i - k - 3$  (Pinheiro and Bates 2000) can be used.

Fitting the four-parameter log-logistic mixed-effects (4pLLME) model (2) is computationally demanding and often does not result in parameter estimates, which restricts its application in practical use. Instead of the strict mixed-effects modeling approach, the two-step meta-analysis strategy is a viable alternative. In the case that complete dose–response relationships are observed in more than three experiments, Jiang and Kopp-Schneider (2014) identified the meta-analysis strategy as a robust method of averaging EC50 estimates from multiple experiments. The first step of the meta-analysis is to estimate individual EC50 values  $\phi_1^{(e)}, \dots, \phi_k^{(e)}$  and their variances  $\sigma_1^2, \dots, \sigma_k^2$  for each experimental data set. The number of experiments is denoted by  $k$ . In the second step, the between-experiment variance  $\hat{\tau}^2$  is estimated. Jiang and Kopp-Schneider (2014) suggested using the Hedges estimator (Hedges and Olkin 1985):

$$\tau_{\text{HE}}^2 = \frac{1}{k-1} \sum_{i=1}^k \left( \phi_i^{(e)} - \frac{1}{k} \sum_{i=1}^k \phi_i^{(e)} \right)^2 - \frac{1}{k} \sum_{i=1}^k \hat{\sigma}_i^2 \quad (3)$$

Then, the weighted average EC50 value  $\hat{\mu}$  over  $k$  experiments is calculated by

$$\hat{\mu} = \frac{\sum_{i=1}^k \hat{w}_i \hat{\phi}_i^{(e)}}{\sum_{i=1}^k \hat{w}_i}, \text{ where } \hat{w}_i = \frac{1}{\hat{\tau}_{\text{HE}}^2 + \hat{\sigma}_i^2}.$$

To construct the 95 % confidence interval for  $\hat{\mu}$ , the method proposed by Hartung and Knapp (2001a, b) is suggested as follows:

$$\left[ \hat{\mu} - t_{k-1,0.975} \sqrt{\frac{\sum_{i=1}^k \hat{w}_i (\hat{\phi}_i^{(e)} - \hat{\mu})^2}{(k-1) \sum_{i=1}^k \hat{w}_i}}, \hat{\mu} + t_{k-1,0.975} \sqrt{\frac{\sum_{i=1}^k \hat{w}_i (\hat{\phi}_i^{(e)} - \hat{\mu})^2}{(k-1) \sum_{i=1}^k \hat{w}_i}} \right],$$

where  $t_{k-1,0.975}$  corresponds to the quantile of the  $t$  distribution.

The performance of different analysis strategies has not yet been investigated regarding the combination of complete and incomplete dose–response relationships [type (b)]. This question will be the first topic of the present paper. Moreover, we provide a new plot pattern for the visualization of dose–response data collected from a large number of experiments (e.g.,  $\geq 10$ ). Finally, we suggest systematic analysis strategies for averaging EC50 from multiple dose–response experiments.

## Materials and methods

### Exemplary case study

The 3T3 mouse fibroblasts and neutral red uptake (NRU) assay were carried out to evaluate the basal cytotoxicity of chemicals. Data, published by Clothier et al. (2013), from 22 experiments performed independently in six laboratories for the test chemical sodium valproate were selected for the analyses. The assay was performed at eight concentration levels of the test chemical across each 96-well plate, in six replicate wells per concentration. In addition, twelve solvent control measurements were available for each experiment.

### Simulation study

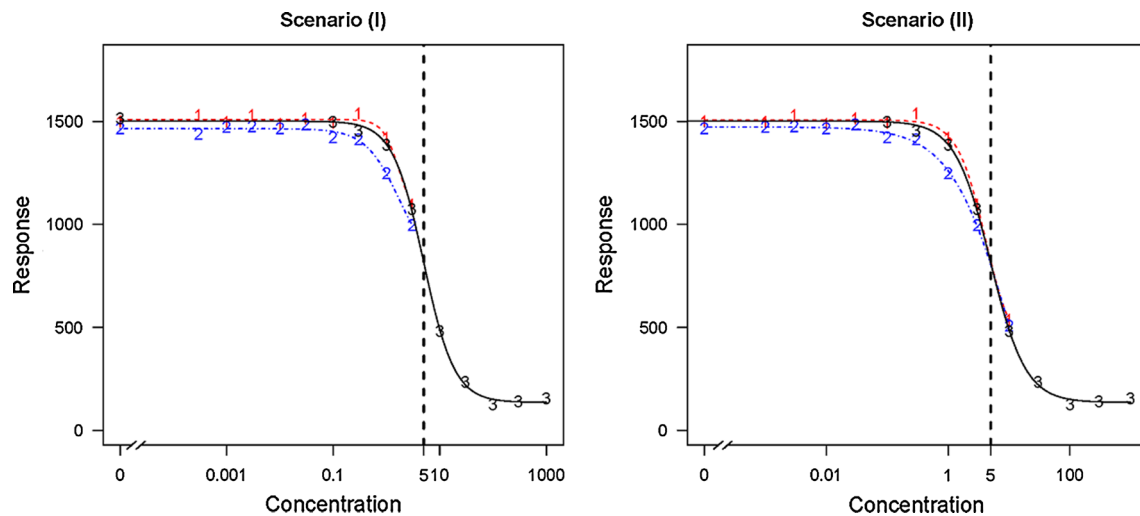
To analyze the combination of complete and incomplete dose–response relationships in multiple experiments, we performed a simulation study. Two scenarios of the experiment that contains an incomplete dose–response relationship were considered: (I) where the true average EC50 value over all experiments was larger than the maximum concentration tested in this experiment, and (II) where

the true overall average EC50 value was within the range of tested concentrations in this experiment, but the lower limit of the decreasing curve was still not reached. The total number of experiments ( $k$ ) was restricted to 3 or 6. For  $k = 3$ , one experiment had a complete dose–response relationship and two had incomplete relationships; for  $k = 6$ , two experiments had complete dose–response relationships and four had incomplete relationships. Figure 2 illustrates the simulation scenarios (I) and (II) for the example of 3 experiments. The data were generated based on the 4LLME model (2) using uncorrelated random effects. The true average EC50 value combined from  $k$  experiment was 5. The concentration levels that cover complete relationships chosen for the experiments were as follows: 0, 0.1, 0.3, 1, 3, 10, 30, 100, 300, and 1,000. For the experiments having incomplete relationships, the concentration levels in scenario (I) were 0, 3e–4, 1e–3, 3e–3, 1e–2, 3e–2, 1e–1, 3e–1, 1, and 3; and in scenario (II) 0, 1e–3, 3e–3, 1e–2, 3e–2, 1e–1, 3e–1, 1, 3, and 10. Twelve replicates of response measurements for concentration 0 and six replicates for the other concentration levels were generated.

To evaluate whether experiments having incomplete dose–response relationships should be excluded from further analyses even in the case of a small total number of experiments, the average EC50 value and its 95 % confidence interval were estimated in the following two situations: (1) including all experiments; and (2) including only those experiments with complete dose–response relationships. The simulation procedure described above was repeated 1,000 times for each scenario and both values of  $k$ , 3 and 6. The average EC50 values obtained using the two approaches in various scenarios were evaluated with respect to the bias and mean square error (MSE) of the estimates. The closer the MSE is to zero, the higher the accuracy of an estimate. Their 95 % confidence intervals (CIs) were compared in terms of length and coverage probability (CP). The CP reflects how often the CIs capture the true EC50 value and should be around 95 % for a 95 % CI.

### Statistical methods and software

The restricted maximum likelihood (REML) estimation method based on the LME (linear mixed-effects) approximation (Lindstrom and Bates 1990) was applied in the 4pLLME model. In the meta-analysis strategy, seven heterogeneity estimators were used: Hedges (HE) (Hedges and Olkin 1985), Hunter–Schmidt (HS) (Hunter and Schmidt 2004), DerSimonian–Laird (DL) (DerSimonian and Laird 1986), Sidik–Jonkman (SJ) (Sidik and Jonkman 2005), (restricted) maximum likelihood (ML) (REML) (Viechtbauer 2005), and Empirical Bayes (EB) (Morris 1983). The average EC50 value was estimated by applying both



**Fig. 2** The simulation scenarios (I) and (II) for three experiments showing a combination of complete and incomplete dose–response relationships. The numbers indicate the mean responses for every

concentration level in each experiment. The dashed line indicates the true overall average EC50 value

weighted and unweighted (UW) versions. For details see Jiang and Kopp-Schneider (2014).

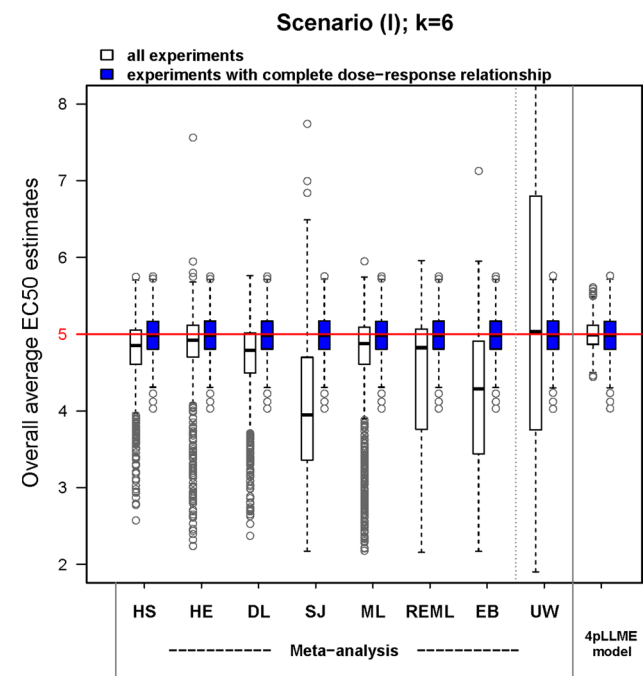
All calculations and visualizations were performed with the statistical software R version 3.0.3 (Team R Development Core 2013). The 4pLL model was fitted using the R package drc (Ritz and Streibig 2005). The R package nlme (Pinheiro and Bates 2000) was used for fitting the 4pLLME model and the R package metafor (Viechtbauer 2010) for computing the heterogeneity estimators in the meta-analysis. A web application for the implementation of the meta-analysis approach and data visualization was established (available at <http://biostatistics.dkfz.de/mdra/>).

## Results

The simulation study compared the results of averaging EC50 estimates by applying mixed-effects modeling and meta-analysis approaches where complete dose–response relationships were observed in fewer than three experiments. The case study of exemplary data presents the results of averaging EC50 estimates from a large number of experiments with the combination of complete and incomplete dose–response relationships by using visualization methods and the meta-analysis approach.

### Simulation results

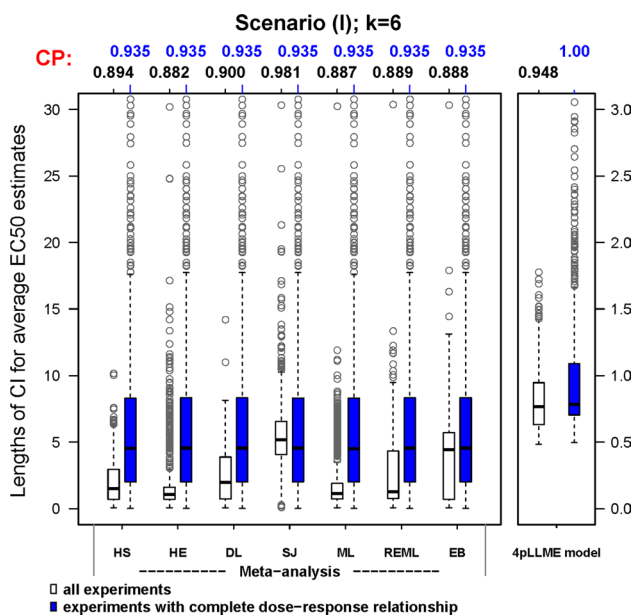
Similar results were obtained for scenarios (I) and (II) in the simulation study. To depict the results, scenario (I) for  $k = 6$  was selected. Figure 3 shows boxplots of the overall average EC50 estimates using all experiments versus



**Fig. 3** Boxplots comparing the range and distribution of the average EC50 estimates over all experiments and excluding those with incomplete dose–response relationships

excluding experiments with incomplete dose–response relationships for different estimation methods. In analogy, Fig. 4 shows the distribution of lengths of CIs for the average EC50 estimates as well as the CPs.

In the case of  $k = 6$ , when using the meta-analysis approach, the average EC50 estimates over all experiments



**Fig. 4** Boxplots comparing the range and distribution of lengths of confidence intervals for the average EC50 estimates over all experiments and excluding those with incomplete dose–response relationships. The corresponding coverage probabilities (CPs) are displayed on top of each boxplot

were very biased (relative bias ranging from  $-19.72$  to  $-1.91$  %), and the unweighted estimation varied strongly (MSE of 7.99). Nevertheless, the average EC50 values estimated from two experiments that covered complete dose–response relationships were unbiased (bias ranging from  $-0.28$  to  $-0.27$  %), no matter which heterogeneity estimator was applied in meta-analysis (e.g., see blue boxplots in Fig. 3). In terms of CPs, the exclusion of experimental data sets with incomplete dose–response relationships obviously improved the results of the CI for average EC50 estimates except when using the SJ heterogeneity estimator, but the CIs were wider due to the reduced number of experiments. For example, in scenario (I), the CPs were lower ( $\leq 0.90$ ) when using all experimental data sets to conduct the meta-analysis and rose to 0.935 after the elimination of data sets with incomplete dose–response relationships, regardless of which heterogeneity estimator was applied (e.g., see blue boxplots in Fig. 4). When using the 4pLLME model, the average EC50 estimates on the basis of all experimental data sets were more accurate than those based on data sets that contained complete dose–response relationships, resulting in the smallest bias of  $-0.0025$  with MSE of 0.0344 in scenario (I) and bias of  $-0.0011$  with MSE of 0.0572 in scenario (II). The 4pLLME modeling of all experimental data resulted in the narrowest CIs for average EC50 estimates as well as their CPs, which were close to 0.95; for example, in scenario (I), the mean length of CIs was 0.809 and the CP was 0.948.

In the case of  $k = 3$ , the average EC50 values estimated from all experiments using the meta-analysis approach were less accurate than the individual EC50 values estimated from a single experimental data set with a complete dose–response relationship. The CIs for average EC50 estimates based on a single experiment with complete dose–response relationship did not meet the requirement for coverage of the true value; the CP was 0.710. The 4pLLME modeling of all experimental data produced the smallest bias in the overall average EC50 estimates, which was  $-0.0012$  with MSE of 0.0711. Although the CP of CIs based on the 4pLLME model of three experiments was too high (0.993), the lengths of CIs were on average narrower than those obtained from meta-analysis; for example, in scenario (I), the mean length of CIs based on the 4pLLME model was 1.733, whereas the smallest mean length of CIs based on meta-analysis using the HS heterogeneity estimator was 133.34.

To sum up, exclusion of experimental data sets that provide inadequate dose–response information from the meta-analysis approach could improve the accuracy of averaging EC50 estimates. If complete dose–response relationships were observed in only fewer than three experiments, the 4pLLME modeling of all experimental data sets led to more accurate results of averaging EC50 estimates than the meta-analysis restricted to the experiments containing adequate dose–response information.

## Results for exemplary case study

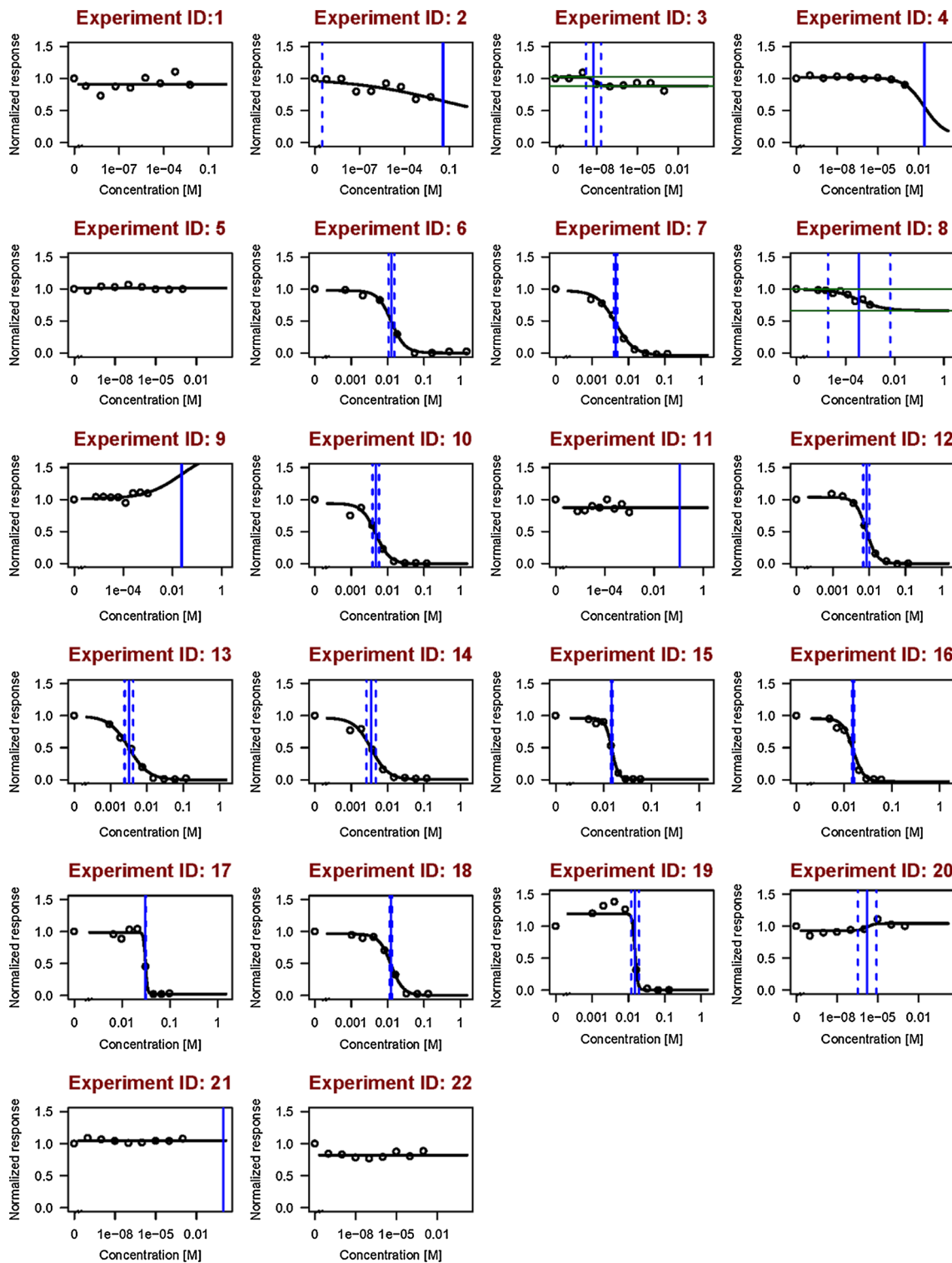
### Visualization of exemplary data

Figure 5 illustrates the exemplary dose–response data sets, each in a separate plot, including dose–response curve fits, mean response values with corresponding concentration levels, as well as individual EC50 estimates and their 95 % CIs. Some experiments exhibit complete and explicit dose–response relationships (e.g., experiment ID: 6), but some show no effects at all in the range of tested concentrations (e.g., experiment ID: 5). Peculiar curve fits (e.g., experiment IDs: 3 and 9) are obtained for some experimental data sets.

To display information about all dose–response experimental data sets simultaneously, we generated a summary plot: the dose–response screening (DORES) plot, shown in Fig. 6.

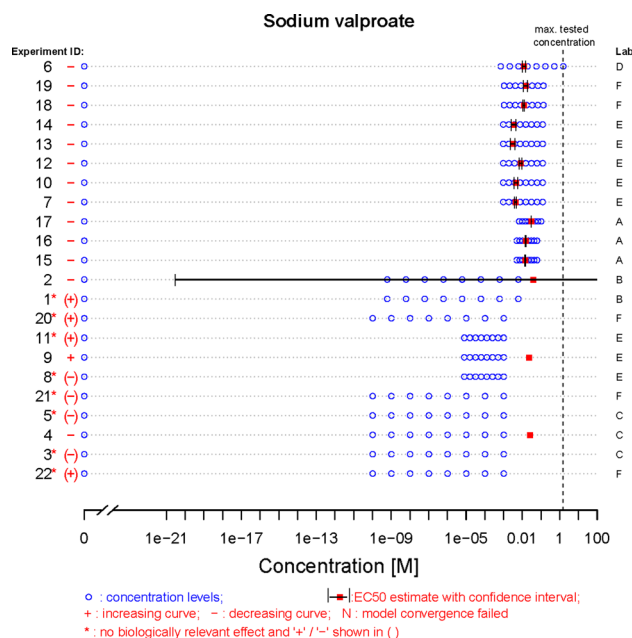
### Using information from the DORES plot to screen data

The first left-hand column of the plot lists the experiment IDs arranged according to the maximum tested concentration level of each experiment from largest to smallest (from top to bottom of the plot). The vertical dashed line indicates



**Fig. 5** Dose–response curves fitted separately to each experimental data set for testing sodium valproate using four-parameter log-logistic model. The vertical lines indicate the EC50 values (solid lines) and

95 % confidence intervals (dashed lines) estimated from each experiment. The data points in each plot represent the mean responses for every concentration level in each experiment



**Fig. 6** DORES plot of the dose–response data from 22 experiments for testing chemical sodium valproate. The blue circles in each row represent the concentration levels tested in each single experiment. The EC50s obtained based on the 4pLL model are depicted as red squares bounded by their confidence intervals. Experiments with  $|(\phi^{(d)} - \phi^{(c)})/\phi^{(d)}| < r\%$  ( $r = 40$ ) were considered as having no biologically relevant effect

the maximum tested concentration across all experiments in the dose–response study. Each row contains dose–response information for one experiment. The blue circles display the tested concentration levels. Various ranges of tested concentration levels for each experiment can be easily compared. Denoting in general a concentration that causes  $x\%$  change of the maximal effect by  $EC_x$ , any quantile  $EC_x$  of the dose–response relationship can be displayed in the DORES plot by a red square bounded by its CI. For example, for  $x = 50$   $EC_{50}$  is displayed in Fig. 6. The graph is plotted on a base-10 logarithmic scale, so that concentration levels are equally spaced on the log scale for serial dilutions and the CIs are symmetrical around the estimated  $EC_x$  values. Comparison of the red square with the blue circles indicates whether the estimated  $EC_x$  is outside of the concentration range. An experiment in which a tested concentration range does not cover a complete dose–response relationship will likely cause problems with  $EC_x$  estimation.

The rightmost column of the plot includes information such as laboratories, technicians, dates of experiment, and any other potential factors that may influence experimental results. For instance, the experimental data presented in Fig. 6 were conducted in six different laboratories (A, B, C, D, E, and F).

The second column from the left provides information about 4pLL model fits. If the dose–response model fails to

converge for a given experiment, then “N” as abbreviation for “Not converged” is displayed. (The non-convergence situation did not occur in the exemplary data.) When no convergence problems occur in the model, the plus sign “+” indicates an increasing curve, while the minus sign “-” represents a decreasing curve. Even if a model fit is obtained for a data set, the change in response over the whole concentration range may be minor and of no biological relevance. The lack of relevance can be quantitatively assessed by inspecting the difference between the lower and upper limits  $\phi^{(c)}$  and  $\phi^{(d)}$ . Specifically, for a predefined percentage  $r\%$ , if  $|(\phi^{(d)} - \phi^{(c)})/\phi^{(d)}| < r\%$  (i.e., the maximum change of mean observed effect is less than  $r\%$ ), then the observed dose–response relationship may be considered not biologically relevant. A meaningful choice of  $r$  would be 40 for the exemplary data, but the value of  $r$  depends on the specific situation. The asterisk “\*” behind an experiment ID number indicates that this experimental data lack biological relevance. If any experiment has no relevance, the other information, such as an increasing or a decreasing curve, an estimated  $EC_x$  value, or its CI, can be ignored. Hence, in these cases, the plus or minus sign is embedded in brackets and the red square is not displayed.

Experimental data sets not containing useful information for estimating the effective concentration of interest should be excluded from averaging analyses. Ideally, the individual experimental data set should have a complete and reasonable dose–response relationship, and a point estimate located within the range of tested concentration levels should be included in the averaging analysis of  $EC_x$  values (e.g., experiment ID: 6 in Figs. 5 and 6). A data set can be excluded from averaging effective concentrations, if one or more of the following characteristics occur:

- “N”: the applied dose–response model fails to converge.
- “\*”: the dose–response relationship exhibits no biological relevance. For example, see experiment IDs: 3 and 8 in Figs. 5 and 6, where the difference between the upper  $\phi^{(d)}$  and lower  $\phi^{(c)}$  limits of their fitted curves is too small.
- For biological reasons, decreasing responses is expected for increasing dose levels, but the model fitting shows increasing curve behavior and vice versa. For example, see experiment ID: 9 in Figs. 5 and 6, where an increasing dose–response relationship is observed in contradiction to what is expected.

#### Averaging results

According to the aforementioned criteria, the experimental data sets with experiment IDs: 1, 3, 5, 8, 9, 11, 20, and 22 were excluded from the  $EC_{50}$  averaging analysis. From the DORES plot, two experiment IDs (2 and 4) were identified

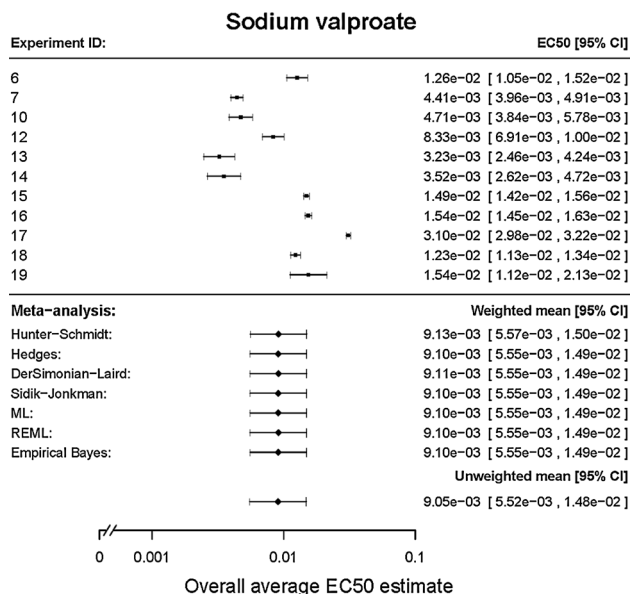
with concentration ranges too narrow to observe the complete dose–response relationship, resulting in estimated EC50 values larger than the maximum tested concentration levels. Additionally, for experiment ID 2, the CI for EC50 estimate was extremely large, and for experiment ID 4, the variance of EC50 could not be estimated. Jiang and Kopp-Schneider (2014) concluded therefore that the meta-analysis strategy would be suitable for averaging EC50 estimates where complete dose–response relationships were observed in more than three experiments. Since more than three experimental data sets contained full dose–response information, the data sets of experiment IDs 2 and 4 were excluded from the EC50 averaging analysis as well.

The individual EC50 estimates with their 95 % CIs are illustrated in the upper half of the forest plot (Fig. 7). The variation among the experiments was obvious. The average EC50 estimates based on eleven selected experiments were performed by using the meta-analysis approach. The results are shown in the lower half of the forest plot. Using different heterogeneity estimators did not lead to differences in overall average EC50 estimates.

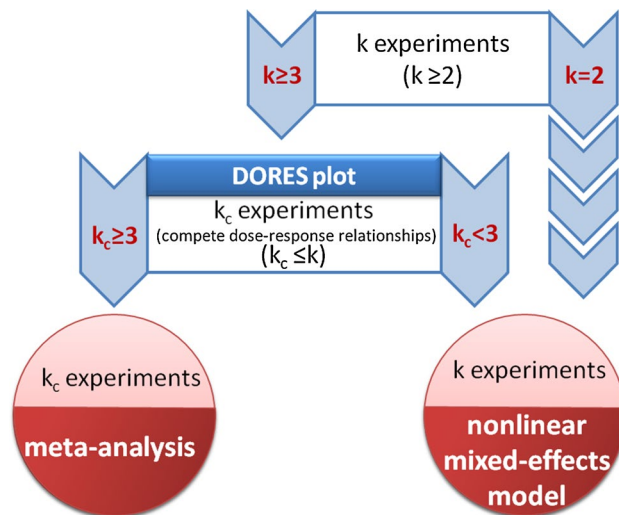
**Discussion and recommendation**

The present study investigates analysis strategies for averaging EC50 estimates from multiple dose–response experiments performed on different concentration levels and ranges from experiments that do not exhibit clear and complete dose–response relationships. We propose the DORES plot to visualize multiple dose–response experimental data sets in a single plot. It provides a convenient and efficient way to compare all experiments and to select experimental data sets for further analyses. The DORES plot can display information from a study involving a large number of experiments: that is, which concentration levels and ranges are chosen for each experiment; whether all experiments exhibit complete and biologically relevant dose–response relationships; and what the individual EC50 estimates and their confidence intervals are.

The systematic analytical strategies for averaging EC50 estimates from multiple dose–response experiments are summarized in the flowchart in Fig. 8. The first step is to conduct a dose–response analysis for each individual experimental data set, in order to investigate whether all experiments contain the necessary information and to display the results of the preliminary analyses in the DORES plot. The second step is to select those experimental data sets with complete dose–response relationships to be included in the averaging analysis. The last step is to average the EC50 estimates. If complete dose–response relationships are observed in at least three experiments, the meta-analysis approach is recommended to average EC50



**Fig. 7** Forest plot showing the individual EC50 estimates with 95 % confidence intervals for 11 experiments for cytotoxicity of sodium valproate. The overall average EC50 values with corresponding 95 % confidence intervals estimated using the meta-analysis approach are presented at the bottom of the figure



**Fig. 8** Flowchart summarizing the analytical strategies for averaging EC50 estimates from multiple dose–response experiments. The total number of experiments is denoted by  $k$ , and  $k_c$  indicates the number of experiments involving complete dose–response relationships

estimates. The Hedges estimator is suggested to estimate the between-experiment variance, which can even be calculated by hand [see formula (3)]. If only a few experiments (<3) with complete dose–response relationships remain, the best method is to carry out more experiments choosing the concentration range in which the complete dose–response



relationship can be observed. However, in reality, this is often not possible due to time, costs, and other restrictions. In this situation, a nonlinear mixed-effects model based on all experimental data sets is more appropriate than the meta-analysis approach. In the case of no convergence problems, the mixed-effects model as the exact analysis strategy should always be applied to averaging EC50 estimates. The advantage of the mixed-effects modeling approach is that it integrates the between-experiment information to obtain a joint distribution for all parameterized effects available. As a consequence, the confidence interval for mean EC50 over all experiments tends to be smaller as compared to that obtained from the meta-analysis approach.

The proposed strategies can also be used to average other quantiles of effective concentrations (e.g., EC10). In our study, we applied the four-parameter log-logistic model to fit dose–response data. The proposed approach and strategies can be generalized to other dose–response models, e.g., the hormesis model, log-normal models, and Weibull models. We have created a web application to generate the DORES plot and to perform the proposed averaging analysis strategies of multiple dose–response experiments. For details please access <http://biostatistics.dkfz.de/mdra/>.

**Acknowledgments** We thank our colleagues Dr. Tim Holland-Letz, Dr. Manuela Hummel, and Ann Thüringer for providing valuable comments and proof reading. The work of Xiaoqi Jiang was supported by the European Union’s Seventh Framework Collaborative Large-Scale Integrating Project Predict-IV under Grant agreement No. 202222.

## References

- Clothier R, Gómez-Lechón M, Kinsner-Ovaskainen A et al (2013) Comparative analysis of eight cytotoxicity assays evaluated within the ACuteTox Project. *Toxicol Vitro* 27(4):1347–1356
- Core TRD (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN: 3-900051-07-0
- DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Control Clin Trials* 7(3):177–188
- Finney DJ (1976) Radioligand assay. *Biometrics* 32:721–740
- Hartung J, Knapp G (2001a) On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Stat Med* 20:1771–1782
- Hartung J, Knapp G (2001b) A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat Med* 20(24):3875–3889
- Hedges LV, Olkin I (1985) *Statistical methodology in meta-analysis*. Academic Press, New York
- Hunter JE, Schmidt FL (2004) *Methods of meta-analysis: correcting error and bias in research findings*. Sage, Thousand Oaks
- Jiang X, Kopp-Schneider A (2014) Summarizing EC50 estimates from multiple dose–response experiments: a comparison of a meta-analysis strategy to a mixed-effects model approach. *Biometric J* 56:493–512
- Lindstrom MJ, Bates DM (1990) Nonlinear mixed effects models for repeated measures data. *Biometrics* 46:673–687
- Morris CN (1983) Parametric empirical Bayes inference: theory and applications. *J Am Stat Assoc* 78(381):47–55
- Pinheiro JC, Bates DM (2000) *Mixed effects models in S and S-PLUS*. Springer, Berlin
- Ritz C, Streibig JC (2005) Bioassay analysis using R. *J Stat Softw* 12(5):1–22
- Sidik K, Jonkman JN (2005) Simple heterogeneity variance estimation for meta-analysis. *J R Stat Soc Ser C Appl Stat* 54(2):367–384
- Viechtbauer W (2005) Bias and efficiency of meta-analytic variance estimators in the random-effects model. *J Educ Behav Stat* 30(3):261–293. doi:10.3102/10769986030003261
- Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. *J Stat Softw* 36(3):1–48