

---

## Research Article

---

# Importance of Shrinkage in Empirical Bayes Estimates for Diagnostics: Problems and Solutions

Radojka M. Savic<sup>1,2</sup> and Mats O. Karlsson<sup>1</sup>

Received 11 February 2009; accepted 21 July 2009; published online 1 August 2009

**Abstract.** Empirical Bayes (“post hoc”) estimates (EBEs) of  $\eta$ s provide modelers with diagnostics: the EBEs themselves, individual prediction (IPRED), and residual errors (individual weighted residual (IWRES)). When data are uninformative at the individual level, the EBE distribution will shrink towards zero ( $\eta$ -shrinkage, quantified as  $1\text{-SD}(\eta_{\text{EBE}})/\omega$ ), IPREDs towards the corresponding observations, and IWRES towards zero ( $\varepsilon$ -shrinkage, quantified as  $1\text{-SD}(\text{IWRES})$ ). These diagnostics are widely used in pharmacokinetic (PK) pharmacodynamic (PD) modeling; we investigate here their usefulness in the presence of shrinkage. Datasets were simulated from a range of PK PD models, EBEs estimated in non-linear mixed effects modeling based on the true or a misspecified model, and desired diagnostics evaluated both qualitatively and quantitatively. Identified consequences of  $\eta$ -shrinkage on EBE-based model diagnostics include non-normal and/or asymmetric distribution of EBEs with their mean values (“ETABAR”) significantly different from zero, even for a correctly specified model; EBE–EBE correlations and covariate relationships may be masked, falsely induced, or the shape of the true relationship distorted. Consequences of  $\varepsilon$ -shrinkage included low power of IPRED and IWRES to diagnose structural and residual error model misspecification, respectively. EBE-based diagnostics should be interpreted with caution whenever substantial  $\eta$ - or  $\varepsilon$ -shrinkage exists (usually greater than 20% to 30%). Reporting the magnitude of  $\eta$ - and  $\varepsilon$ -shrinkage will facilitate the informed use and interpretation of EBE-based diagnostics.

**KEY WORDS:** empirical Bayes estimate; model building; model evaluation; NONMEM; shrinkage.

## INTRODUCTION

Model-based analysis has lately become a method of choice for analyzing the outcome from clinical trials in drug development. The model(s) explaining underlying pharmacokinetics (PK) and pharmacodynamics (PD) of a therapeutic agent as well as disease progression is used for data description, interpretation, and simulation of future scenarios, as well as decision-making. Thus, the choice of an adequate model is a crucial but often not straightforward procedure. The model building process is difficult and involves testing, evaluating, and diagnosing a range of plausible models with a major aim to make an adequate inference from the observed data, summarized as parameter estimates, variability estimates, and significant covariates (1). There are several tools assisting modelers in a model building process, including graphical diagnostics (1–3).

Graphical displays are extensively used during the model building process and are considered an essential tool for data visualization, inspection of model adequacy, and assumption testing. Graphical diagnostics are considered a powerful and intuitive visual method to be used not only by modelers but

also as a communicating tool. The US Food and Drug Administration and the European Committee for Medicinal Products for Human use guidances explicitly mention the utility of graphical diagnostics. Specifically mentioned in both guidances are graphical diagnostics based on individual parameter estimates (4–6). The individual parameter estimates in non-linear mixed effects are estimated using the Bayesian methodology, and they are generally referred to as an empirical Bayes estimates (EBEs). The use of the phrase ‘empirical Bayes’ emphasizes that the parameters for the prior distribution are estimated from the data and are used as if they were known to obtain the posterior distribution. At one extreme, with no observations available, the patient will be regarded as a typical patient. At the other extreme, when data for an individual goes towards infinity, the prior will have marginal impact; in between these extremes, both factors will contribute, and depending on the relative variability (including both, between, and within subject variability), individual estimates could be closer either to the population mean or to the true individual parameter value (7–9).

Although the advantages and strengths of graphical diagnostics have been justified and stressed on several occasions, shortcomings have not been systematically explored (10). In this work, we investigate the informative value of the EBE-based diagnostics and how their usefulness depends on the individual data richness. Additionally, circumstances which are likely to misguide modelers towards making erroneous deci-

---

<sup>1</sup> Division of Pharmacokinetics and Drug Therapy, Department of Pharmaceutical Biosciences, Faculty of Pharmacy, Uppsala University, P.O. Box 591, 75124, Uppsala, Sweden.

<sup>2</sup> To whom correspondence should be addressed. (e-mail: Rada.Savic@farmbio.uu.se)

sions in model development, relating to choice of structural, covariate, and stochastic model components, are underlined.

**METHODS**

**Empirical Bayes Estimation and Notation**

Derivation of individual parameter estimates (EBEs), denoted for an individual  $j$  as a vector of  $s$  model parameters  $P_j=(p_{j1},p_{j2}...p_{js})$ , requires quantities such as (1) typical individual estimates denoted as a vector of  $\theta=(\theta_1,... \theta_s)$ ; (2) the magnitude of inter-individual variability of parameters in question, variances ( $\omega_1^2, \dots, \omega_s^2$ ), and covariances of the  $\Omega$ -matrix; as well as (3) the magnitude (variance) of residual variability ( $\sigma^2$ ) to be known. The number of variance parameters ( $\omega^2$ ) does not necessarily need to be the same as the number of typical parameter estimates ( $\theta$ ). Then, in individual  $j$ , the  $k_{th}$  parameter's deviation from the typical parameter value in the population can be denoted  $\eta_{jk}$ . In parametric non-linear mixed effect modeling, the shape of the distribution of the  $\eta$ s is assumed to be normal. Commonly, to avoid negative parameter values, the parameter distribution is log-normally transformed using the relationship described in Eq. 1 where  $\eta_{jk}$  enters nonlinearly into the expression for the parameter in order to appropriately mimic the parameter distribution shape in the population. Also, as Eq. 1 is straightforward, the  $\eta_{jk}$  value is often referred to as an individual parameter. Thus, considering Eq. 1, the vector  $\eta_k$  will be used as an indicator for the individual parameter throughout the remainder of this communication.

$$p_{jk} = \theta_k \cdot e^{\eta_{jk}} \tag{1}$$

For individual  $j$ , individual parameter  $\eta_{jk}$ ,  $\sim 1 \leq k \leq s$  can be estimated from the observed data vector  $y_{ji}=(y_{j1}, y_{j2}, \dots, y_{jn})$ ,  $n$  being the number of observations within an individual and known prior parameter distribution. Detailed derivation of this estimation procedure is available elsewhere (7,11).

With  $\hat{y}_{ji}$ , we denote a model prediction of  $y_{ji}$ , defined as a function of parameter vector  $P_j$  and  $X_{ji}$ ,  $X_{ji}$  being the vector of independent variables (such as time and dose covariates) related to  $y_{ji}$  (Eq. 2).

$$\hat{y}_{ji} = f(P_j, X_{ji}). \tag{2}$$

**Definition of Diagnostic Variables**

In this communication,  $\hat{y}_{ji}$  is referred to as an individual prediction (IPRED) and to  $\eta_{EBE,jk}$  as the EBE. Another useful term, the individual weighted residual (IWRES), is calculated following Eq. 3. By using the standard deviation of the residual variability as a weighting factor, the IWRES distribution is standardized to have a zero mean and unit variance expectation when the model and parameters are correct at the individual level.

$$IWRES = \frac{(y_{ij} - \hat{y}_{ij})}{\sigma}. \tag{3}$$

**Definition of  $\eta$ - and  $\varepsilon$ -Shrinkage**

If the population model is adequate, the quality of the individual parameter estimate will depend heavily on the observed data available. When data are sparse and less informative on individual parameters, it is expected that the EBEs will deviate less from the population mean, which will result in a difference in the EBE and priors distribution, mainly in terms of decreased variance of EBEs but also with possible distortions of the distribution shape. In an extreme case of no data available on a particular individual, the individual's EBE will be equal to the population value. So, the variance of EBE distribution is shrinking towards zero as the quantity of information at the individual level diminishes, a phenomenon defined as  $\eta$ -shrinkage ( $sh_\eta$ ). Similarly, as individual parameter estimate tend towards the true individual parameter in case of informative data, IWRES distribution approaches a normal distribution with zero mean and unit variance. Conversely, as data diminishes, the IWRES distribution shrinks towards zero. We define this phenomenon as  $\varepsilon$ -shrinkage ( $sh_\varepsilon$ ) and is sometimes also called "overfitting".

$\eta$ - and  $\varepsilon$ -shrinkage are calculated using Eqs. 4 and 5. A shrinkage magnitude of zero, i.e., no shrinkage, corresponds to the situation when the model is correct and individual data is sufficiently abundant so that individual parameter estimates mimic the true individual parameter estimates, and a magnitude of one is the opposite case when data contain virtually no information about the parameters in question and the individual parameter values approach the typical parameter value.

$$sh_\eta = 1 - \frac{SD(\eta_{EBE})}{\omega} \tag{4}$$

$$sh_\varepsilon = 1 - SD(IWRES) \tag{5}$$

**Simulations**

*General Description*

The influence of shrinkage on EBE-based diagnostics was assessed via ordinary ( $n \leq 10$ ) or Monte Carlo (MC;  $n = 100$ ) PK or PD model simulation–estimation procedures using non-linear mixed effects modeling (NONMEM; 12). The model summary is given in Table I. Data sets with different information content were created by altering the number (sample frequency) and location (sampling time) of samples. Each scenario defined with a certain model, and study design underwent multiple simulations. If not specified otherwise below, each simulated dataset contained 1,000 subjects. Thereafter, the EBEs were estimated using either the true and/or an intentionally misspecified model for each dataset using Bayesian estimation (i.e., MAXEVAL=0 option) in NONMEM. In this estimation, the prior information, i.e., population parameter estimates, variances, covariances, and residual variability were fixed to the same values used for simulations of the data sets, whenever misspecification was not introduced.

**Table I.** Summary of Models used for Simulations

Model indexing	Model description
Model 1	One compartment i.v. bolus model
Model 2	One compartment i.v. bolus model with Michaelis–Menten elimination
Model 3	Two compartment i.v. bolus model
Model 4	One compartment model with first order absorption
Model 5	One compartment model with zero order absorption
Model 6	One compartment model with transit compartment absorption
Model 7	Two compartment model with first order absorption
Model 8	E <sub>max</sub> model
Model 9	Sigmoidal E <sub>max</sub> model
Model 10	Indirect response model

i.v. intravenous

Based on estimated EBEs, IPREDs, and IWRES, graphical diagnostics were created and subsequently inspected for trends or indications of model misspecification. Also, the extent of  $\eta$ - and  $\varepsilon$ -shrinkage was calculated for these estimated EBE or IWRES distributions. Overall, the relationship between the extent of shrinkage and informativeness of diagnostics was assessed both qualitatively, i.e., by visual inspection of whether or not the resulting plot indicated the expected pattern, and quantitatively, i.e., to determine at what shrinkage magnitude a particular diagnostics lose its utility.

Commonly used EBE diagnostics were explored. A brief description of these diagnostics as well as the methodology used to study their informativeness for each case is given below.

### Specific Description

#### 1. Shape of EBE distribution

*Description.* A visual diagnostic often used to inspect the EBE distribution and to confirm the normality assumption is either a histogram or, preferably, a quantile–quantile-plot (qq-plot), where deviations from normality can be detected.

*Methodology.* Multiple datasets ( $n=10$  for each unique design) were simulated from a PK model (one compartment model with first order absorption, model 4) using different study designs with respect to sampling frequency (2–10) and sampling times. The EBEs were estimated using the true model for each dataset. EBE and IWRES distributions were visualized and compared to the true normal distribution using qq-plots.

#### 2. ETABAR outcome

*Description.* A hypothesis test if the mean value of EBEs is significantly different from zero is performed at each run, and the  $p$  value of this test is reported in the so-called “ETABAR outcome” from NONMEM. A significant ETABAR is considered as an indication for a potential model misspecification.

*Methodology.* Data sets were simulated from one compartment intravenous bolus model (model 1) and one compartment model with first order absorption (model 4) to study this diagnostic. The simulated datasets comprised two and three samples each for models 1 and 2, respectively. Two conditions were studied: (1) dependence of  $\eta_{CL}$  ETABAR on the time of the last sample in model 1 and (2) dependence of the  $\eta_{ka}$  ETABAR on the time of the first sample in model 2. For condition 1, seven (7) MC simulation–estimation studies were performed using different study designs in which the last sample was positioned at the time equal to the 0.67, 1, 2, 3, 4, 6, and 8 of the disposition half-life ( $t_{1/2}=3$  h) for seven MC studies. Similarly, for condition 2, eight MC simulation–estimation studies was performed in which the first samples was positioned to be equal to the 0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, and 4 absorption half-lives ( $t_{1/2abs}=1$  h). Each MC involved 100 replicates of simulated datasets. The true model was fitted to the simulated data using First Order Conditional Estimation method (FOCE) in NONMEM VI.

#### 3. EBE versus EBE plots

*Description.* This diagnostic is often used for inspection of the correlation between individual parameter estimates. If the plot indicates a markedly non-zero correlation for a model where the population correlation value was fixed to zero, the correlation is generally tested in the model for its significance. If the plot indicates zero correlation for a model where the population correlation value was also zero, the correlation is often not tested for its significance.

*Methodology.* Two conditions were studied: the value of EBEs to (1) indicate correlation if it is truly present and/or (2) correctly determining when the correlation is not present. Condition 1 was studied using model 1, in which the true correlation between individual parameters of  $\eta_{CL}$  and  $\eta_V$  was 0.75. Datasets were simulated using this model and different study designs. For each design, ten replicates were simulated. The model with true parameter estimates but no correlation was used to estimate EBEs. The expectation is that EBE versus EBE plot will indicate parameter correlation. Correlation between EBEs values for  $\eta_{CL}$  and  $\eta_V$  was computed for each EBE versus EBE plot. Condition 2 was studied using model 4 and model 9 (an E<sub>max</sub> model). In these models, the true correlation between all individual parameters was zero. All study conditions were equivalent to the study described above; however, the true model with no correlation was used to estimate EBEs.

#### 4. EBE versus covariate plots

*Description.* This diagnostic is used for inspection of relationships between parameters and covariates. When a large number of covariates are available, this diagnostic enables a quick screening of relationships. Only indicated relationships are often subsequently estimated directly in the population model. EBEs are also used in automated procedures such as the generalized additive modeling (13) and to investigate the shape of the covariate-parameter relationship.

*Methodology.* The utility of EBEs to correctly indicate the EBE-covariate relationship when (1) it is truly present and (2) when the covariate relationship is not present were studied using model 4. A direct relationship between typical

value of volume of distribution and weight was used for simulations. Multiple datasets were simulated under different study designs, where early observations were censored sequentially. A model with no covariate relationship was fitted to the data and used to estimate EBEs which were subsequently plotted *versus* the covariate (weight). Correlation between EBEs and weight as well as the slope of EBE *versus* weight relationships were calculated for each data set.

#### 5. Dependent variable (DV) *versus* IPRED plot

*Description.* This is a commonly used diagnostic for detection of (structural) model misspecification. Its usefulness, compared to the DV *versus* population prediction (PRED) plot, comes from separating different sources of variability.

*Methodology.* A range of models was used to simulate multiple datasets, which were subsequently analyzed with structurally misspecified models. The examples involve the following pairs of models (first model in a pair is used for simulation, second for estimation): (1) sigmoidal Emax model–Emax model (models 9 and 8), (2) transit compartment absorption model–first order absorption model (model 6 and 4), (3) two compartment model–one compartment model (models 3 and 1), (4) one compartment model with first order absorption–one compartment model with zero order absorption (models 4 and 5), and (5) one compartment model with Michaelis–Menten elimination–one compartment model with linear elimination (models 2 and 1). For each condition, multiple datasets with different number of observations were simulated. IPREDs were derived from the misspecified model.

#### 6. |IWRES| *versus* IPRED plot

*Description.* This is a commonly used diagnostic for assessing the residual error model. An appropriate residual error model should be associated with a horizontal regression line with no slope.

*Methodology.* Multiple datasets based on different study designs where observations were censored randomly were simulated using model 8 with combined proportional and additive error model and subsequently fitted with the misspecified model containing proportional error model only. Absolute values of IWRES were plotted *versus* IPRED for determination of residual model misspecification. The expectation was that IWRES values would indicate model misspecification visualized as a linear regression line being not horizontal (e.g., having slope different than 0). The slope of this regression line was calculated for each case.

### Relationship between $\eta$ - and $\varepsilon$ -Shrinkage

Relationship between  $\eta$ - and  $\varepsilon$ -shrinkage was studied using model 7. First, rich datasets were simulated (number of observations=25). Then, sparse datasets were created by censoring observations within subjects either randomly or uniformly (same observation deleted for all subjects), keeping the total number of observations the same for all subjects. This procedure was repeated until two observations per

subject remained. For each dataset, EBEs were estimated using the true model and true parameter estimates.  $\eta$ - and  $\varepsilon$ -shrinkage were calculated and inspected for relationships.

## RESULTS

In general, in the absence of shrinkage, all EBE-based diagnostics are powerful model evaluation tools. These diagnostics separate different sources of variability, thus graphs become easier to interpret. However, in the presence of shrinkage, all the discussed diagnostics begin losing their informativeness and could potentially become misleading as indicated in our detailed evaluation of (1) consequences of  $\eta$ -shrinkage, (2) consequences of  $\varepsilon$ -shrinkage, and (3) relationship between  $\eta$ - and  $\varepsilon$ -shrinkage.

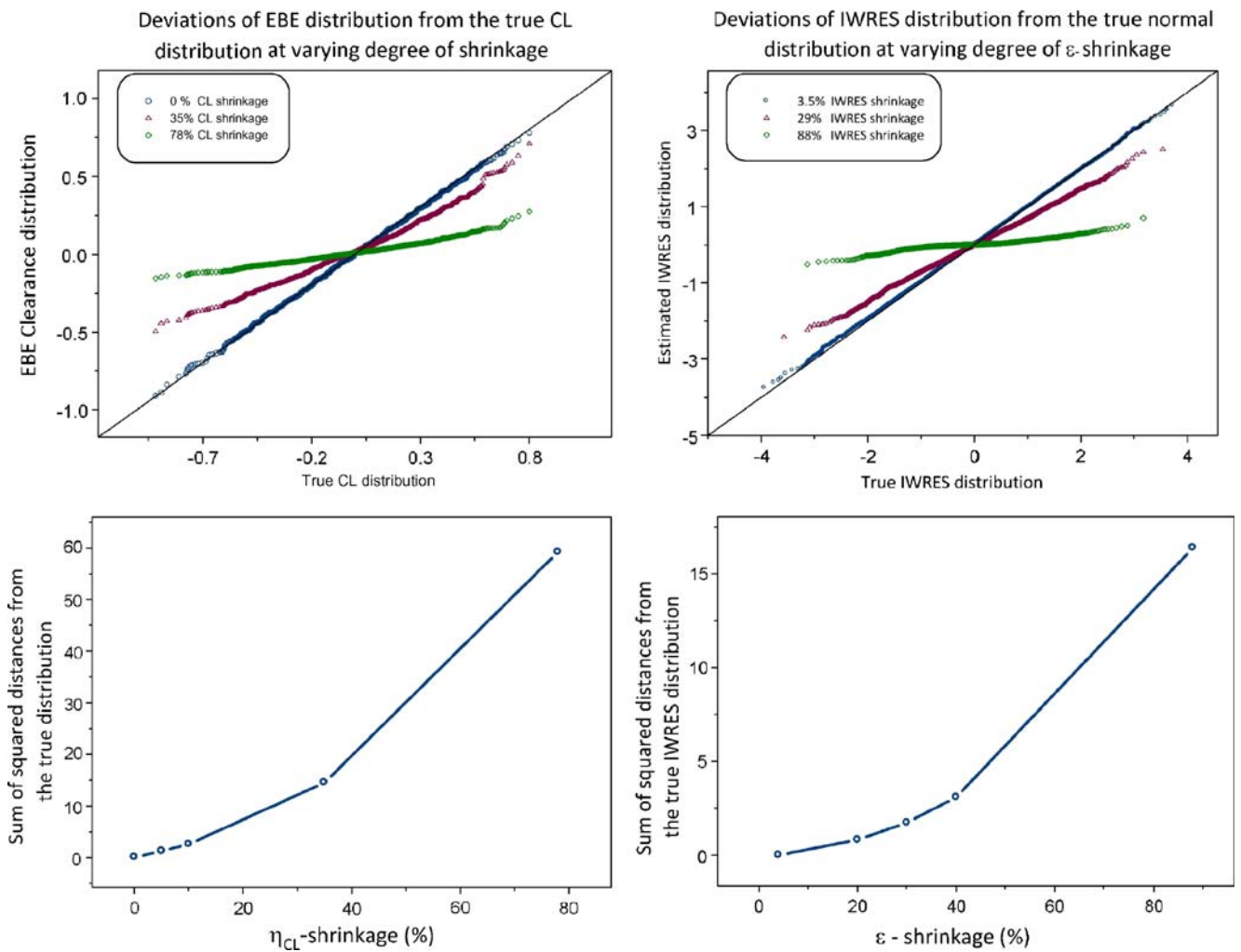
### Consequences of $\eta$ -Shrinkage

EBEs may, in addition to shrinkage, show non-normal distribution even when the true underlying  $\eta$  distribution is normal. A pattern of increasing deviations from normality in the qq-plot with increasing shrinkage, exemplified in Fig. 1 (left panels) for EBE of  $\eta_{CL}$ , was evident for all parameters of the studied model (model 4). The larger the shrinkage, the larger are the observed deviations from normality.

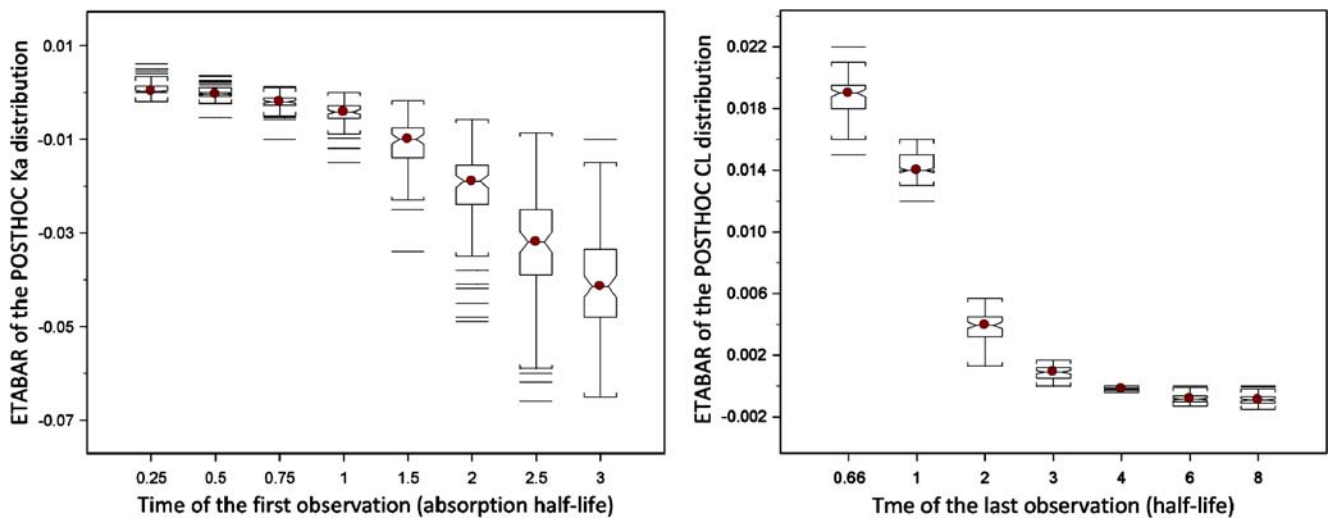
As a result of asymmetric  $\eta$ -shrinkage, mean values of EBEs (“ETABAR”) may be significantly different from zero, even for a correctly specified model. The relationship between ETABAR for  $\eta_{ka}$  and  $\eta_{CL}$  distributions and the time of the first or last sample, respectively, is shown in Fig. 2. For  $\eta_{ka}$ , when the first sample is taken early enough, i.e., informative on this parameter, ETABAR is close to zero, as expected. If the first sample is taken later in time, this sample is not particularly informative for individuals that have fast absorption because their absorption process is essentially complete by that time, but it will still be informative for individuals that have slow absorption; thus, shrinkage is expected to happen mainly for individuals with high  $k_a$  values, which leads to asymmetric shrinkage resulting in ETABAR values significantly different from zero. A similar phenomenon is evident for the  $\eta_{CL}$  EBE distribution (right panel of Fig. 2).

When EBEs are used for inspections for parameter correlations, they can indicate correlation when correlation truly does not exist (upper panel of Fig. 3), and oppositely, correlation when truly present would sometimes not be detected by EBEs (lower panel of Fig. 3). The upper-left corner and lower-left corner panels of Fig. 3 represent the true parameter relationships, while other panels show relationships based on the EBEs estimated under the true model for different study designs. As Fig. 4 indicates, commonly induced correlations include those between  $\eta_{EC50} \sim \eta_{Emax}$  and  $\eta_{ka} \sim \eta_v$ . The quantitative relationship between shrinkage extent and informativeness of these plots is shown in Fig. 4. Both phenomena, induced and hidden correlations, become apparent when shrinkage is higher than 20–30%.

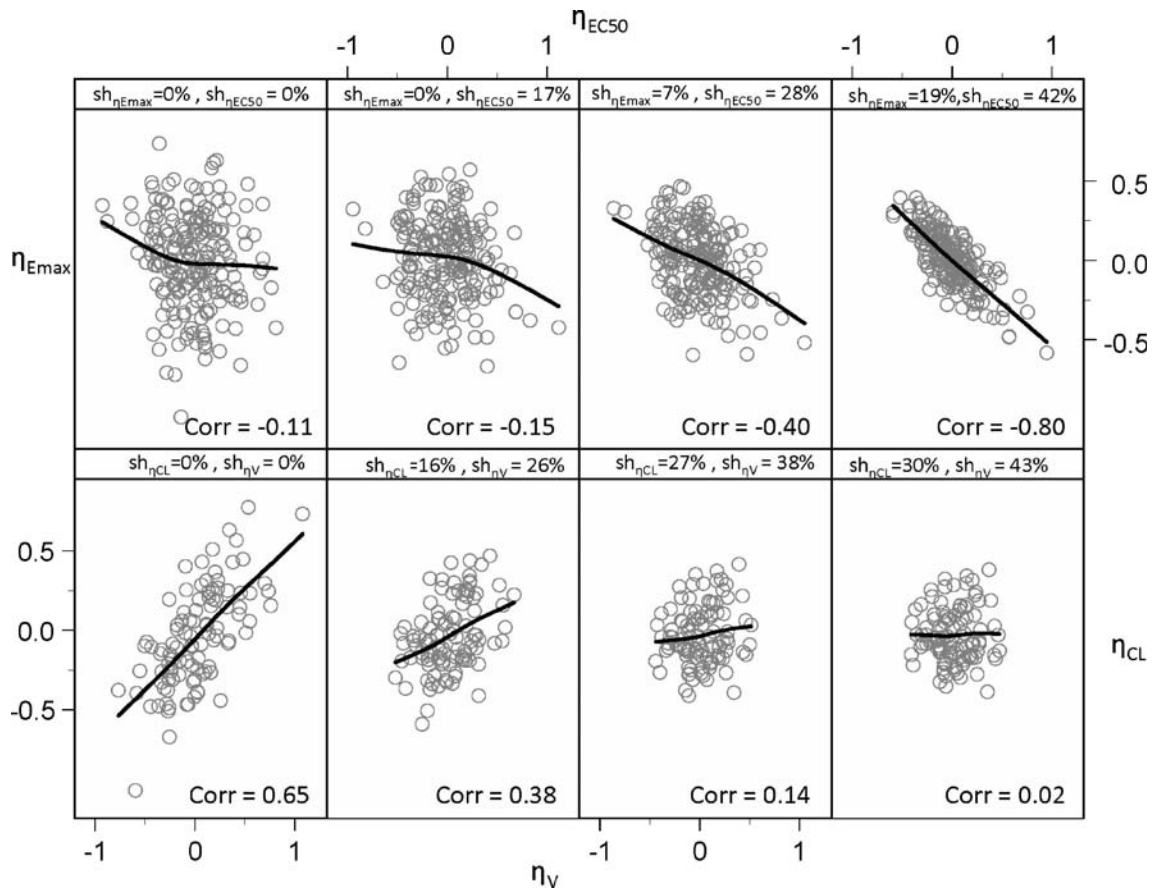
Similarly, when EBEs are used for a search of significant covariates, they can obscure relationships, show distorted shape, or even in certain circumstances falsely indicate relationship, when relationship truly does not exist (Figs. 5 and 6). The left panel of Fig. 6 shows falsely induced



**Fig. 1.** Deviation of empirical Bayes estimates (EBE) and individual weighted residual (IWRES) distribution from normality due to  $\eta$ - and  $\varepsilon$ -shrinkage, respectively, shown as a qq-plot (upper panels) and sum of squared distances from the true distributions (lower panels). This is a representative example of three simulations. EBE and IWRES were estimated using the true model



**Fig. 2.** Dependence of ETABAR on the time of the first and last sample for  $\eta_{ka}$  and  $\eta_{CL}$  distribution, respectively. Deviations of ETABAR from zero is a result of asymmetric empirical Bayes estimates shrinkage

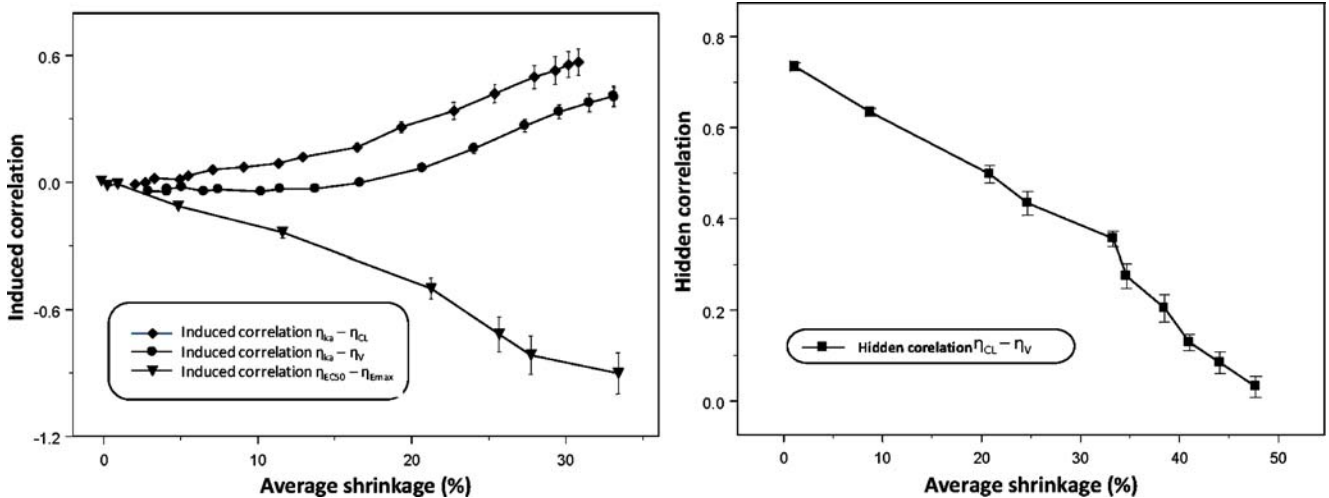


**Fig. 3.** Empirical Bayes estimates (EBE) versus EBE plot indicating parameter correlation due to  $\eta$ -shrinkage when correlation is not truly present (*upper panel*) and hiding parameter correlation when correlation is truly present (*lower panel*)

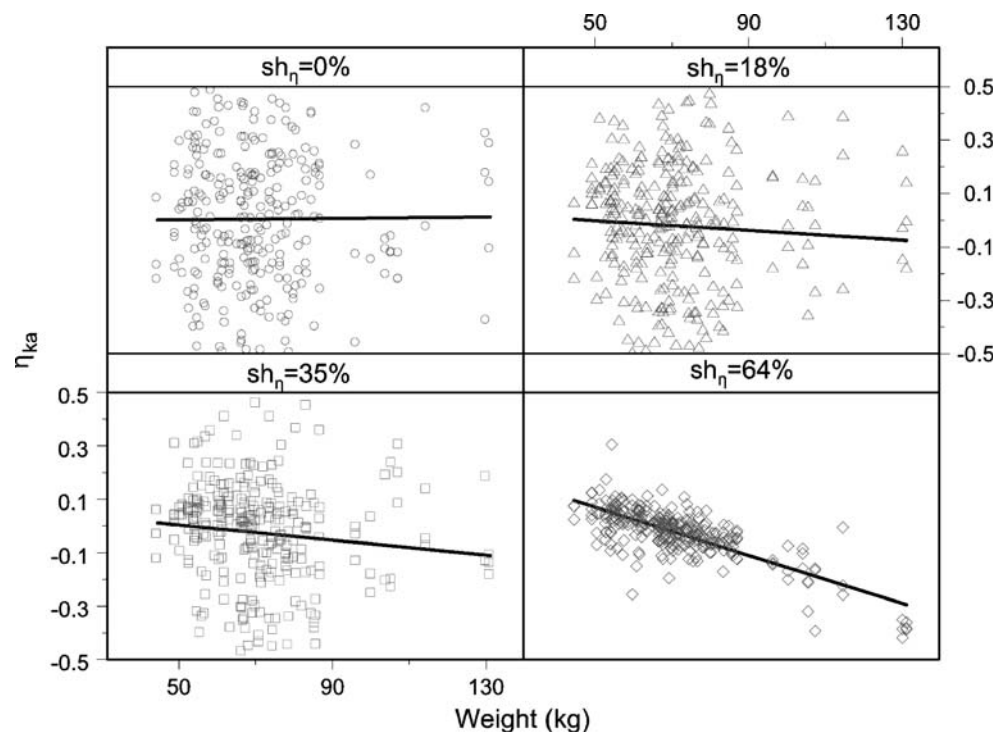
relationship between  $\eta_{ka}$  and weight, when the true covariate relationship was between volume and weight. The right panel of Fig. 6 shows how an existing relationship between weight and  $\eta_V$  is less evident, the higher the shrinkage in  $\eta_V$ . The observed pattern appears due to induced correlation between  $\eta_{ka}$  and  $\eta_V$ .

**Consequences of  $\epsilon$ -Shrinkage**

The IWRES distribution may in the presence of  $\epsilon$ -shrinkage show non-normal distribution even when the underlying model is correct (Fig. 1, right panel).



**Fig. 4.** *Left panel:* relationship between empirical Bayes estimates with increased shrinkage. The average of the shrinkage in two studied parameters (e.g.,  $\eta_{Emax}$  and  $\eta_{EC50}$ ) is shown on the x-axes. Each symbol represents mean (SE) based on ten simulation-estimation procedures

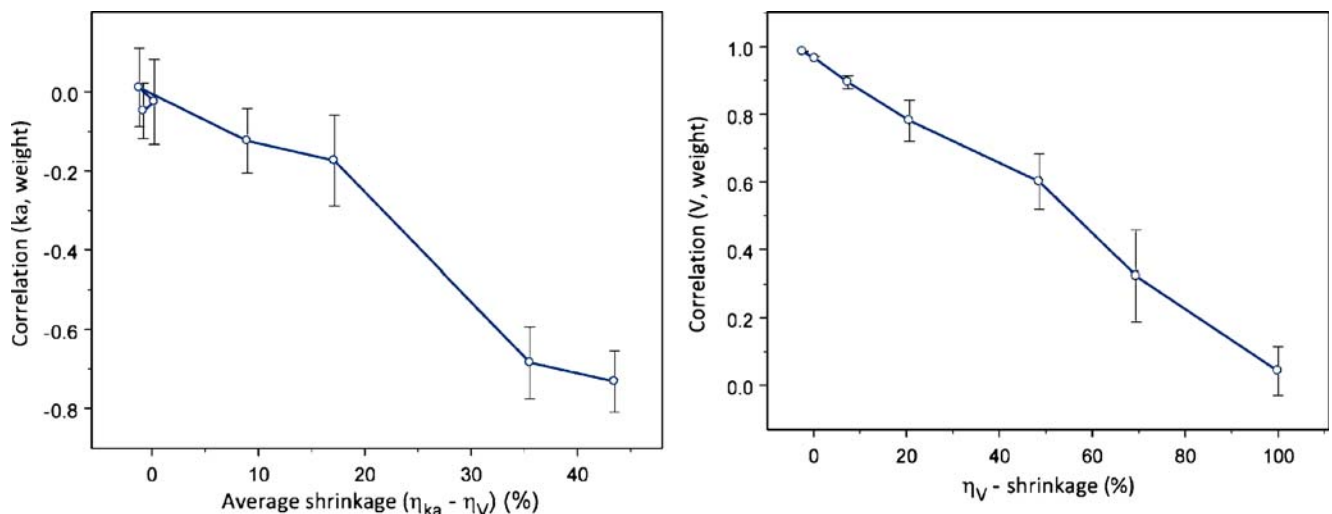


**Fig. 5.** Empirical Bayes estimates *versus* covariate plot indicating parameter-covariate relationship due to  $\eta$ -shrinkage when relationship is not truly present

The power of IPREDs to detect model misspecification is reduced in the presence of  $\varepsilon$ -shrinkage. Figure 7 shows the DV *versus* IPRED diagnostic when a misspecified model (zero order absorption, model 5) was applied to data simulated with first order absorption (model 5) in two scenarios: with rich data resulting with low  $\varepsilon$ -shrinkage and with sparse data resulting in high  $\eta$ - and  $\varepsilon$ -shrinkage. When  $\varepsilon$ -shrinkage is low, the plot clearly indicates the model misspecification; however, when shrinkage is substantial, signs of model misspecification are absent, and the plot

indicates a perfect fit. Similar phenomena were observed with all other scenarios tested. Already, at the  $\varepsilon$ -shrinkage extent of 20–30%, it was hard to visually detect model misspecification.

The power of IWRES to detect residual model misspecification diminishes with an increase in  $\varepsilon$ -shrinkage. Figure 8 (left panel) shows IWRES diagnostics produced based on misspecified model (proportional residual error only) fitted to the data simulated with the model that had combined error model (additive + proportional residual



**Fig. 6.** Relationship between shrinkage extent and induced covariate (*left panel*) and hidden covariate relationship (*right panel*). The average of the  $\eta_{ka}$ - and  $\eta_V$ -shrinkage is shown on the x-axes of the *left panel*

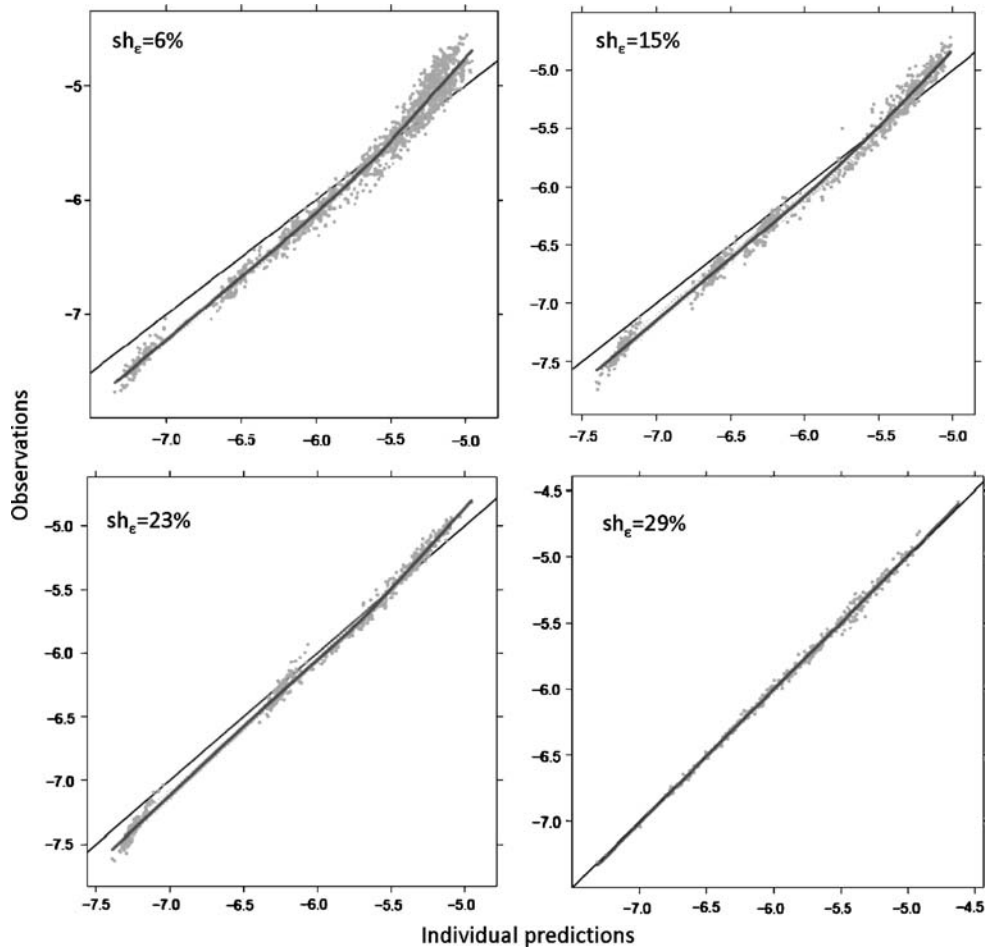


Fig. 7. Individual prediction versus dependent variable plot for detection of structural model misspecification

error). A quantitative relationship between  $\epsilon$ -shrinkage and model misspecification indication, expressed as the slope of the regression line for  $|IWRES|$  versus IPRED, is shown in Fig. 8 (right panel).

**Relationship between  $\eta$ - and  $\epsilon$ -Shrinkage**

Relationships between  $\eta$ - and  $\epsilon$ -shrinkage are shown in Fig. 9. As expected, these two terms are positively correlated.

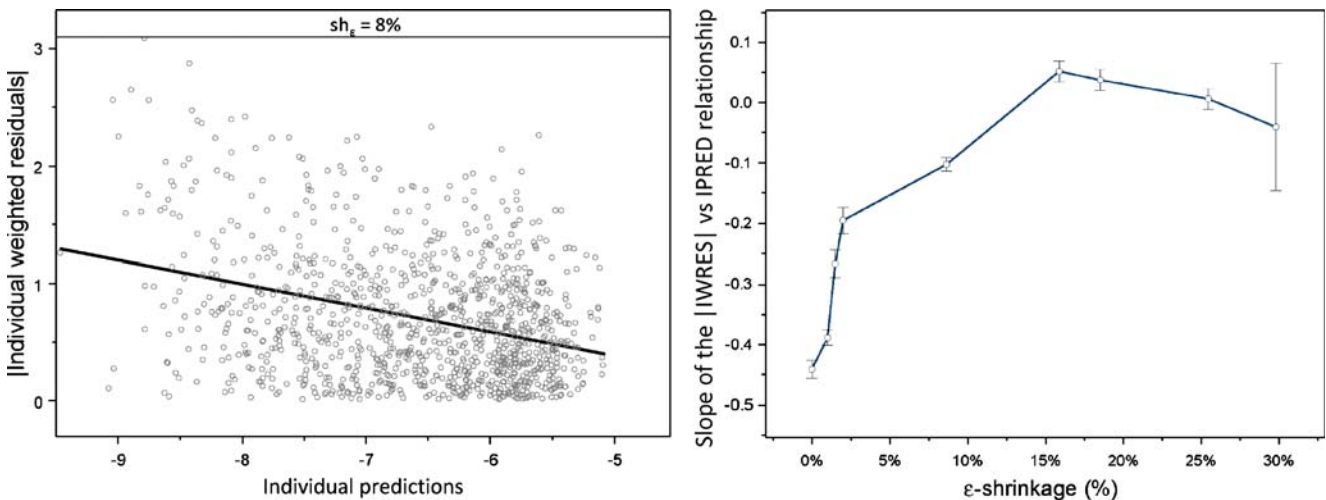


Fig. 8. Power of individual weighted residual ( $|IWRES|$ ) versus individual prediction (IPRED) to detect residual model misspecification. In absence of shrinkage, the regression line of  $|IWRES|$  versus IPRED clearly indicates residual model misspecification (left panel). With increased shrinkage, the slope of the regression line is diminishing (right panel)



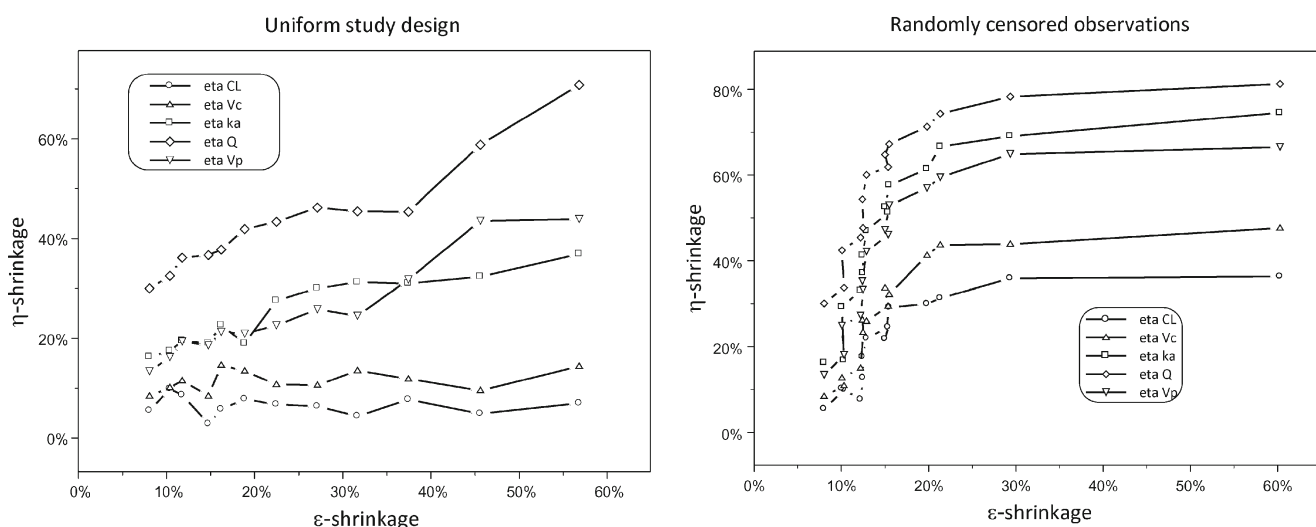


Fig. 9. Relationship between  $\eta$ - and  $\varepsilon$ -shrinkage under different study designs

The shape of this relationship is however different for different study designs.

## DISCUSSION

In this paper, we have evaluated the usefulness of EBE-based diagnostics in different circumstances and their dependencies on the information content in the datasets. These diagnostics are powerful when the individual's data provide a substantial amount of information on the model parameters. However, with decreased information content per individual, these diagnostics become not only uninformative but potentially also misleading. The EBE-based diagnosis is a central part of the model building process, but if a modeler would accept such diagnostics without considering shrinkage, it may impede decision-making, increase time for data analysis, decrease trust in adequate models, and accept inappropriate models.

Although it is difficult to give a rule of thumb at what shrinkage extent diagnostics becomes misleading, we observed in our examples that already at a level of 20–30% for both,  $\eta$ - and  $\varepsilon$ -shrinkage, EBE-based diagnostics generally lost their power, and false indications started to appear. Therefore, it seems rational to provide values for shrinkage whenever EBE-based diagnostics are used for communicating model quality. It is important to clarify that all mentioned patterns are associated with the graphical diagnostics solely. When both models were tested directly in NONMEM, correct and misspecified, the correct model would be selected for every example, based on standard model building procedures using the objective function values. Thus, significant shrinkage in most of the cases does not indicate any problem with the dataset or with the model; it only affects the diagnostics based on EBE. Thus, whenever shrinkage is present, a model building process involving more direct testing and less or no reliance on EBE-based diagnostics should be used. Additionally, other types of diagnostics ought to be used in these cases, for example simulation-based diagnostic (10) or conditional weighted residuals (14).

An interesting finding was that EBEs may in the presence of shrinkage indicate false relationships or hide true relationships when used for covariate screening. Here, this was a consequence of the induced correlation between two parameters,  $\eta_{ka}$  and  $\eta_v$ . (in general, sign of such an induced correlation may depend on the design). This finding is of special interest when EBEs are used for selection of covariates that will be further tested in the model for their significance. If only certain parameters are screened for covariates, it may happen that EBEs would indicate false parameter-covariate relationships, which may even turn out to be significant when tested directly in the model, while the covariate was truly related to other parameter. Therefore, whenever this is the case (i.e., single parameter is screened for covariate relationships and a certain covariate appears to be significant), the modeler should also test the relationship between the significant covariate and other parameters in order to determine the appropriate relationship.

We have observed that  $\eta$ - and  $\varepsilon$ -shrinkages are correlated, which is intuitive. However, while  $\eta$ -shrinkage may appear substantial even when observations per subject are numerous,  $\varepsilon$ -shrinkage remains low in these circumstances.  $\varepsilon$ -Shrinkage becomes substantial when the number of observations is equal or less than the number of random effects. As a general principle, the lower the residual variability, the more informative are observations, and the lower will the  $\eta$ -shrinkage be. Also, as a general principle, a higher number of between-subject random effect and higher variability in these will lead to a higher flexibility in the individual model to approach to observed values and therefore lead to higher  $\varepsilon$ -shrinkage.

We have so far only discussed shrinkage magnitudes between zero and one. Based on theoretical considerations,  $\eta$ - and  $\varepsilon$ -shrinkage should, in the absence of model misspecification, be positive (15). Indeed, whenever the shrinkage was computed under the correct model, shrinkage values were in this range. For a number of scenarios when data were simulated and model parameters re-estimated with misspecifications of the parameter distribution, shrinkage was non-negative. However, shrinkage may also take negative values

**Table II.** Summary of the 29 Models Used to Assess the Shrinkage in Published Models with Real Datasets

Model	Type	Number of obs	Number of ind	Number of obs/ind (average)	Number of $\theta$	Number of $\eta$	Shrinkage (min)	Shrinkage (max)	Shrinkage CL
CdA (18)	PK	1,102	161	6.84	12	7	6.66	84.17	6.66
Pefloxacin (19)	PK	337	74	4.55	2	2	5.16	37.68	5.16
Glibenclamide lag time (20)	PK	287	8	35.88	9	5	-4.38	13.14	-4.38
Glibenclamide transit (21)	PK	287	8	35.88	10	7	-4.05	12.72	-4.05
Desmopressin kid lag time (22)	PK	139	72	1.93	5	4	5.91	42.35	5.91
Desmopressin kid transit (22)	PK	373	72	1.93	6	5	8.61	66.97	8.61
Desmopressin adult lag time (22)	PK	373	28	13.32	6	4	2.17	5.83	2.36
Desmopressin adult transit (22)	PK	373	28	13.32	6	7	0.73	51.21	0.99
Desmopressin joint lag time (22)	PK	512	100	5.12	10	4	11.51	32.15	11.51
Desmopressin joint transit (22)	PK	512	100	5.12	8	6	13.80	49.47	13.80
Levosimendan (23)	PK	359	24	14.96	5	4	0.95	33.41	0.95
Antibody Atm-027 (24)	PK	413	14	29.50	8	4	-1.39	17.37	1.26
Moxonidine lag time (25)	PK	1,022	74	13.81	6	6	17.76	54.18	32.38
Moxonidine transit (21)	PK	1,022	74	13.81	7	9	14.33	97.10	36.62
Gefitinib (26)	PK	705	34	20.74	7	5	-0.58	5.55	-0.35
Melagatran (27)	PK	3,595	596	6.03	8	3	12.63	38.03	12.63
Pyrazinamide (28)	PK	3,092	227	13.62	10	6	13.79	85.82	16.31
Prazosin (29,30)	PK	887	64	13.86	8	8	14.44	56.25	44.86
Antibody X	PK	559	70	7.99	6	4	16.85	43.56	16.85
Voriconazole (31)	PK	1,274	83	15.35	11	8	2.04	92.58	71.77
Tobramycin (32)	PK	322	97	3.32	5	1	8.46	90.77	8.46
Levodopa (33)	PD	851	19	44.79	7	7	-1.80	38.64	NA
Digoxin (34)	PD	787	225	3.50	5	2	7.10	71.94	NA
CdA (35)	PD	341	59	5.78	5	3	2.44	25.46	NA
Docetaxel (36)	PD	3,549	636	5.58	11	4	18.29	46.64	NA
Moxonidine (blood pressure; 37)	PD	1,942	97	20.02	8	3	3.72	34.94	NA
Moxonidine (heart rate; 37)	PD	1,944	97	20.04	5	2	3.81	31.73	NA
Tesaglitazar (FPG; 38)	PD	4,035	412	9.79	9	4	6.26	38.26	NA
Tesaglitazar (HbA1C measurements; 39)	PD	4,663	412	11.32	13	5	1.54	42.39	NA

The number of observations (dependent variables), individuals, and random effects are summarized as well as maximum and minimum  $\eta$ -shrinkage values. Please see referenced articles for further information

*obs* observations, *ind* individuals, *NA* not applicable, *PK* pharmacokinetic, *PD* pharmacodynamic, *FPG* fasting plasma glucose, *CL* clearance

in certain circumstances; we found negative shrinkage in some real data examples, typically with rich data and a small number of subjects. Also, in situations where a parameter variance is fixed to a lower value than the true value, negative shrinkage may occur. Thus, negative shrinkage may be a useful signal for possible model misspecification.

This exercise was performed on simulation examples, and one can argue that such scenarios would not occur often in model building processes using data from real clinical trials. However, when inspecting 29 published PK and PD models, we identified that shrinkage was present in all of them (Table II). Substantial shrinkage is often present in PD parameters and PK parameters describing absorption and distribution. Even the most well-informed parameter, clearance (CL), showed shrinkage above 20% in about one third of the PK studies. A large number of observations do not guarantee the absence of  $\eta$ -shrinkage as it can occur even in cases when the number of samples is as large as 20 observations per individual. This is simply because the studies are not designed to provide perfect information on all model parameters. In this work, we explored implications of EBE shrinkage for model diagnostics. However, EBEs are also used in other aspects of non-linear mixed effect analysis, for example, in FOCE, sequential PK-PD analysis or nonparametric estimation in NONMEM (14,16,17). Shrinkage would affect these estimation procedures too; however, this is outside the scope of the present investigation.

In conclusion, when shrinkage is higher than about 20–30%, EBE-based diagnostics lack informativeness and may be misleading. Therefore, it is desirable to report the extent of  $\varepsilon$ - and  $\eta$ -shrinkage to assess the relevance of diagnostics employing EBEs, IPRED, and IWRES. When shrinkage is high, other diagnostics and more direct population model estimation need to be employed in model building and evaluation.

## ACKNOWLEDGEMENTS

The authors would like to thank Dr. Siv Jönsson for valuable comments on the manuscript, Paul Baverel for his help with reviewing shrinkage in real models and datasets, and the anonymous reviewers for their valuable comments on the manuscript.

## REFERENCES

- Williams PJ, Ette EI, editors. *Pharmacometrics: the science of quantitative pharmacology*. New York: Wiley; 2007.
- Ette EI, Ludden TM. Population pharmacokinetic modeling: the importance of informative graphics. *Pharm Res*. 1995;12:1845–55.
- Jonsson EN, Karlsson MO. Xpose—an S-PLUS based population pharmacokinetic/pharmacodynamic model building aid for NONMEM. *Comput Methods Programs Biomed*. 1999;58:51–64.
- FDA. *Guidance for industry, population pharmacokinetics*. Washington, DC: US Department of Health and Human Services; 1999.
- Wade JR, Edholm M, Salmonson T. A guide for reporting the results of population pharmacokinetic analyses: A Swedish perspective. *AAPS J*. 2005 Oct 5;7(2):45.
- Bhattaram VA, Booth BP, Ramchandani RP, Beasley BN, Wang Y, Tandon V, *et al*. Impact of pharmacometrics on drug approval and labeling decisions: a survey of 42 new drug applications. *AAPS J*. 2005;7:E503–12.
- Davidian M, Giltinan DM. *Nonlinear models for repeated measurement data*. London: Chapman & Hall/CRC; 1995.
- Sheiner LB, Beal S, Rosenberg B, Marathe VV. Forecasting individual pharmacokinetics. *Clin Pharmacol Ther*. 1979;26:294–305.
- Sheiner LB, Beal SL. Bayesian individualization of pharmacokinetics: simple implementation and comparison with non-Bayesian methods. *J Pharm Sci*. 1982;71:1344–8.
- Karlsson MO, Savic RM. Diagnosing model diagnostics. *Clin Pharmacol Ther*. 2007;82:17–20.
- Wang Y. Derivation of various NONMEM estimation methods. *J Pharmacokinetic Pharmacodyn*. 2007;34:575–93.
- Beal SL, Sheiner LB, Boeckmann AJ. *NONMEM users guides*. Ellicott City, MD: Icon Development Solutions; 1989–2006.
- Jonsson EN, Karlsson MO. Automated covariate model building within NONMEM. *Pharm Res*. 1998;15:1463–8.
- Hooker AC, Staatz CE, Karlsson MO. Conditional weighted residuals (CWRES): a model diagnostic for the FOCE method. *Pharm Res*. 2007;24:2187–97.
- Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data*. New York: Springer; 2000. p. 80–1.
- Zhang L, Beal SL, Sheiner LB. Simultaneous vs. sequential analysis for population PK/PD data I: best-case performance. *J Pharmacokinetic Pharmacodyn*. 2003;30:387–404.
- Savic RM, Kjellsson MC, Karlsson MO. Evaluation of the nonparametric estimation method in NONMEM VI. *Eur J Pharm Sci*. 2009;37:27–35.
- Lindemalm S, Savic RM, Karlsson MO, Juliusson G, Liliemark J, Albertioni F. Application of population pharmacokinetics to cladribine. *BMC Pharmacol*. 2005;5:4.
- Bruno R, Iliadis MC, Lacarelle B, Cosson V, Mandema JW, Le Roux Y, *et al*. Evaluation of Bayesian estimation in comparison to NONMEM for population pharmacokinetic data analysis: application to pefloxacin in intensive care unit patients. *J Pharmacokinetic Biopharm*. 1992;20:653–69.
- Rydberg T, Jonsson A, Karlsson MO, Melander A. Concentration-effect relations of glibenclamide and its active metabolites in man: modelling of pharmacokinetics and pharmacodynamics. *Br J Clin Pharmacol*. 1997;43:373–81.
- Savic RM, Jonker DM, Kerbusch T, Karlsson MO. Implementation of a transit compartment model for describing drug absorption in pharmacokinetic studies. *J Pharmacokinetic Pharmacodyn*. 2007;34:711–26.
- Osterberg O, Savic RM, Karlsson MO, Simonsson US, Norgaard JP, Walle JV, *et al*. Pharmacokinetics of desmopressin administered as an oral lyophilisate dosage form in children with primary nocturnal enuresis and healthy adults. *J Clin Pharmacol*. 2006;46:1204–11.
- Jonsson EN, Antila S, McFadyen L, Lehtonen L, Karlsson MO. Population pharmacokinetics of levosimendan in patients with congestive heart failure. *Br J Clin Pharmacol*. 2003;55:544–51.
- Zingmark PH, Edenius C, Karlsson MO. Pharmacokinetic/pharmacodynamic models for the depletion of Vbeta5.2/5.3 T cells by the monoclonal antibody ATM-027 in patients with multiple sclerosis, as measured by FACS. *Br J Clin Pharmacol*. 2004;58:378–89.
- Karlsson MO, Jonsson EN, Wiltse CG, Wade JR. Assumption testing in population pharmacokinetic models: illustrated with an analysis of moxonidine data from congestive heart failure patients. *J Pharmacokinetic Biopharm*. 1998;26:207–46.
- Li J, Karlsson MO, Brahmer J, Spitz A, Zhao M, Hidalgo M, *et al*. CYP3A phenotyping approach to predict systemic exposure to EGFR tyrosine kinase inhibitors. *J Natl Cancer Inst*. 2006;98:1714–23.
- Cullberg M, Eriksson UG, Wahlander K, Eriksson H, Schulman S, Karlsson MO. Pharmacokinetics of ximelagatran and relationship to clinical response in acute deep vein thrombosis. *Clin Pharmacol Ther*. 2005;77:279–90.
- Wilkins JJ, Langdon G, McIlleron H, Pillai GC, Smith PJ, Simonsson US. Variability in the population pharmacokinetics of pyrazinamide in South African tuberculosis patients. *Eur J Clin Pharmacol*. 2006;62:727–35.
- Karlsson MO, Sheiner LB. The importance of modeling interoccasion variability in population pharmacokinetic analyses. *J Pharmacokinetic Biopharm*. 1993;21:735–50.

30. Mandema JW, Verotta D, Sheiner LB. Building population pharmacokinetic–pharmacodynamic models. I. Models for covariate effects. *J Pharmacokinet Biopharm.* 1992;20:511–28.
31. Karlsson MO, Lutsar I, Milligan PA. Population pharmacokinetic analysis of voriconazole plasma concentration data from pediatric studies. *Antimicrob Agents Chemother.* 2009;53:935–44.
32. Aarons L, Vozeh S, Wenk M, Weiss P, Follath F. Population pharmacokinetics of tobramycin. *Br J Clin Pharmacol.* 1989;28:305–14.
33. Troconiz IF, Naukkarinen TH, Ruottinen HM, Rinne UK, Gordin A, Karlsson MO. Population pharmacodynamic modeling of levodopa in patients with Parkinson's disease receiving entacapone. *Clin Pharmacol Ther.* 1998;64:106–16.
34. Hornestam B, Jerling M, Karlsson MO, Held P. Intravenously administered digoxin in patients with acute atrial fibrillation: a population pharmacokinetic/pharmacodynamic analysis based on the digitalis in acute atrial fibrillation trial. *Eur J Clin Pharmacol.* 2003;58:747–55.
35. Lindemalm S, Liliemark J, Gruber A, Eriksson S, Karlsson MO, Wang Y, *et al.* Comparison of cytotoxicity of 2-chloro-2'-arabino-fluoro-2'-deoxyadenosine (clofarabine) with cladribine in mononuclear cells from patients with acute myeloid and chronic lymphocytic leukemia. *Haematologica.* 2003;88:324–32.
36. Friberg LE, Henningsson A, Maas H, Nguyen L, Karlsson MO. Model of chemotherapy-induced myelosuppression with parameter consistency across drugs. *J Clin Oncol.* 2002;20:4713–21.
37. Brynne L, McNay JL, Schaefer HG, Swedberg K, Wiltse CG, Karlsson MO. Pharmacodynamic models for the cardiovascular effects of moxonidine in patients with congestive heart failure. *Br J Clin Pharmacol.* 2001;51:35–43.
38. Hamren B, Ericsson H, Samuelsson O, Karlsson MO. Mechanistic modelling of tesaglitazar pharmacokinetic data in subjects with various degrees of renal function—evidence of interconversion. *Br J Clin Pharmacol.* 2008;65:855–63.
39. Hamren B, Bjork E, Sunzel M, Karlsson M. Models for plasma glucose, HbA1c, and hemoglobin interrelationships in patients with type 2 diabetes following tesaglitazar treatment. *Clin Pharmacol Ther.* 2008;84(2):228–35.