Pharmaceutical Research, Vol. 26, No. 9, September 2009 (© 2009)
DOI: 10.1007/s11095-009-9931-1

Research Paper

Semiparametric Distributions With Estimated Shape Parameters

Klas J. F. Petersson,[1,2] Eva Hanze,[1] Radojka M. Savic,[1] and Mats O. Karlsson[1]

Received February 10, 2009; accepted June 16, 2009; published online July 1, 2009

Purpose. To investigate the use of adaptive transformations to assess the parameter distributions in population modeling.
Methods. The logit, box-cox, and heavy tailed transformations were investigated. Each one was used in conjunction with the standard (exponential) transformation for PK and PD parameters. The shape parameters of these transformations were estimated to allow the parameter distributions to more accurately resemble a wider range of parameter distributions. The transformations were tested both in simulated settings where the true distributions were known and in 30 models developed from real data.
Results. In the simulated setting the transformations were better than the standard lognormal distribution at characterizing the true distributions. Improvement could also be seen in objective function value (OFV) and in simulation based diagnostics. In the real datasets, significant model improvement based on OFV could be seen in 22, 18, and 22 out of the 30 models for the three transformations respectively.
Conclusion. Transformations with estimated shape parameters are a promising approach to relax the often erroneous assumption of a known shape of the parameter distribution. They offer a simple and straightforward way of handling and characterizing parameter distributions.

KEY WORDS: estimation; normality assumption; parameter distributions; population modeling; transformations.

INTRODUCTION

In analysis of repeated measures data from clinical trials, parametric population (nonlinear mixed effects) modeling methods are often used. In these methods, the shapes of the distributions of parameter values in the population are generally assumed known, whereas the magnitudes of variability are estimated as model parameters. Three distributions are commonly used to describe parameters: first, the normal distribution as the identity transformation of the underlying random effects variable, η, which is a zero mean variable with estimated variance ω^2; second, the lognormal distribution, which often is used for parameters that are bounded to be non-negative; and third, logit transformations, which are often used for parameters that have theoretical upper and lower bounds, for example fractions.

However, the assumptions made regarding distribution shapes may not always hold. The distributions may deviate from assumed shapes for a number of reasons. Polymorphisms in metabolic enzymes may give rise to deviations from lognormality of the clearance distribution. Physiological boundaries may limit many parameters to display the full range of values predicted by theoretical distributions. Distribution shapes may also be influenced by study design, e.g. exclusion criteria in screening for a clinical trial that excludes subjects with too low a baseline value for some biomarker creates a truncated distribution. In these cases, when random effects distributions are skewed, broad, narrow, or heavily tailed, an approach to assess this is needed. The most common way of relaxing the assumption of a known distribution is to use nonparametric estimation methods. These methods estimate the entire probability distributions defined at a number of parameter values equal to the number of individuals in the datasets, completely relaxing distribution assumptions. The nonparametric methods have some drawbacks though: model building, model evaluation, and model utilization techniques are less well developed compared to parametric methods, and software are less accessible and versatile.

A related way to address complex distributions is through mixture models, where the full parametric distribution is assumed to be the sum of parametric distributions from a number of subpopulations, each with its own distribution characteristics. Mixture models take an intermediate position between parametric and nonparametric methods. To date, they have mainly been used when there is some mechanistic reason to believe that there are existing, distinctly different subpopulations.

Another way of handling non-standard distributions is to use a transformation function where "shape" parameters of the transformation are estimated along with the rest of the model parameters. These estimated distributions do not fully relax the assumptions on the distributions but allow a wider range of distributions to be adequately described. Further, despite being named "semiparametric" (1), they still retain the advantages of parametric, over nonparametric, methods. The

[1] Department of Pharmaceutical Biosciences, Division of Pharmacokinetics and Drug Therapy, Uppsala University, Box 591SE-751 24, Uppsala, Sweden.
[2] To whom correspondence should be addressed. (e-mail: Klas.petersson@farmbio.uu.se)

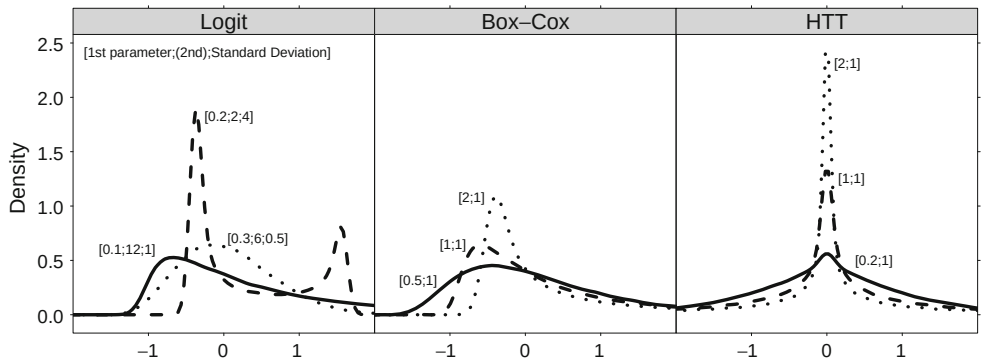

Fig. 1. Examples of a normal distribution transformed by the different transformations. The shape parameter values used to create the distributions and the standard deviation of the original distribution are indicated.

first semiparametric transformation was developed and implemented in the NLMIX software by Davidian and Gallant in 1993 (2). This method uses polynomials as transformation functions. Another developed method uses spline functions (1,3). Neither the polynomial nor the spline transformations appear to have been used in published papers outside the citations above. Likely reasons for this are instability, complex implementation, and lack of supported software.

In this work we will focus on other semiparametric transformations that are easy to implement in existing software. Our aims are as follows:

(i) By simulation and estimation investigate the transformations' ability to better describe an underlying distribution by estimating shape parameters in the transformations.
(ii) Investigate the possibility of improving the model fit to data using these new transformations on already developed models based on real data.

METHODS AND MATERIALS

Transformations

The semiparametric transformations in this work are expected to be used in the following manner: first, the standard, fixed transformation (e.g. identity, lognormal, or logit transformation) is applied to the parameter. Thereafter, the random effects parameter, η_i, in that fixed transformation is replaced by a flexible transformation, $\eta_{i,transformed}$, where at least one parameter related to the shape of the transformation is estimated. When selecting transformation functions, three features were sought: (i) ability to take on the identity transformation ($\eta_{i,transformed}$ taking on a normal distribution), allowing the need for these transformations to be tested by the likelihood ratio test; (ii) ability to facilitate the interpretation of typical individual parameters by assuring that $\eta_i=0$ will result in $\eta_{i,transformed}=0$; and (iii) ability to retain, approximately, the correlation structure among random effects by assuring that the rank orders of η_i and $\eta_{i,transformed}$ are the same. This means that if η_i is ordered from lowest to highest, $\eta_{i+1} > \eta_i$, this will also be true post transformation, $\eta_{i,transformed+1} > \eta_{i,transformed}$. By ensuring this outcome, the implantation of a transformation should not to any large extent affect any estimate of the correlations between random effects.

Three transformations were investigated: logit, box-cox, and heavy-tailed (HT).

(i) The logit transformation (Eq. 1) uses two estimated shape parameters to transform a normal distribution into a potentially left or right skewed distribution or a bimodal one, see Fig. 1. The parameter values need to have boundaries, θ_1 needs to be between 0 and 1, and θ_2 needs to be a positive value. θ_1 governs the skewness of the distribution and θ_2 the width. When ω is small, θ_2 is large, and θ_1 is 0.5, the logit transformation approaches the identity transformation, resulting in a normal distribution of $\eta_{i,transformed}$.

$$\eta_{i\_Transformed} = \left(e^{\left[LOG\left(\frac{\theta_1}{(1-\theta_1)}\right)+\eta_i \right]} \cdot \left(1 + e^{\left[LOG\left(\frac{\theta_1}{(1-\theta_1)}\right)+\eta_i \right]}\right)^{-1} - \theta_1 \right) \cdot \theta_2 \quad (1)$$

(ii) The box-cox transformation (Eq. 2) uses one parameter to transform a normal distribution into a left or right skewed distribution. The parameter needs no boundaries, but the transformation function is not defined at $\theta_1=0$. As $\theta_1 \to 0$ the box-cox transformation approaches the identity transformation.

$$\eta_{i\_Transformed} = \frac{\left((e^{\eta_i})^{\theta_1} - 1\right)}{\theta_1} \quad (2)$$

(iii) The heavy tailed transformation needs one parameter that determines the distributions shape (Eq. 3). This transformation can, as the name implies, create distributions that are more heavily tailed than the

Table I. Mixtures Used to Simulate Skewed Distribution

Prob.	Typical value (L/h)	Variance
0.4	20	0.0256
0.3	30	0.1089
0.2	40	0.25
0.1	50	0.36

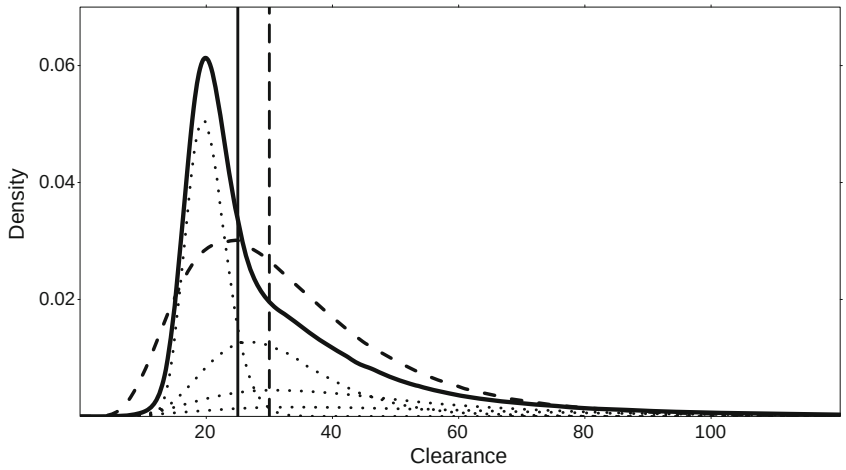

Fig. 2. The distribution of clearance that was used to simulate the datasets (solid line). The dotted distributions are the four sub-distributions that make up the total distribution. Vertical lines are the median of the true (solid) and an example of an estimated lognormal distribution (dashed).

normal distribution, but can also create a symmetric bimodal distribution. For η=0, this function is not defined at θ_1=0, and as $\theta_1 \rightarrow 0$ the HT transformation also approaches the identity transformation. The parameter needs no boundary but one could restrict it to a positive value to get only tailed distributions, not bimodal.

$$\eta_{i\_Transformed} = \eta_i \cdot |\eta_i|^{\theta_1} \quad (3)$$

Table II. The Models in Which the Distribution Transformations were Implemented

Model No. (compound)	No. ID	No. Obs.	Administration	Disposition/Description	No. ETAs	Reference
PK1—Levosimedan	24	359	IV[b]	2-comp.	3	(8)
PK2—Antibody ATM-027	14	413	IV	2-comp.	5	(9)
PK3—Cladribine	161	488	IV/PO[a]/SC	3-comp.	5	(10)
PK4—Ximelagatran	596	3595	IV/PO	1-comp.	4	(11)
PK5—Antibody X	70	559	IV	2-comp.	4	
PK6—Voriconazole	83	1274	IV/PO	2-comp.	3	
PK7—Desmopressin	28	373	PO	1-comp.	3	(12)
PK8—Desmopressin	28	373	PO	1-comp.	5	(12)
PK9—Desmopressin	72	139	PO	1-comp.	3	(12)
PK10—Desmopressin	72	139	PO	1-comp.	5	(12)
PK11—Desmopressin	100	512	PO	1-comp.	4	(12)
PK12—Desmopressin	100	512	PO	1-comp.	5	(12)
PK13—Moxonidine	74	1022	PO	1-comp.	4	(13)
PK14—Moxonidine	74	1022	PO	1-comp.	5	(13)
PK15—Glibenclamide	8	287	IV/PO	2-comp.	7	(14)
PK16—Glibenclamide	8	287	IV/PO	2-comp.	8	(14)
PK17—Pefloxacin	74	337	IV	1-comp.	2	(15)
PK18—Gefitinib	34	705	PO.	1-comp.	3	(16)
PK19—Tobramycin	97	322	IV	2-comp.	2	(17)
PK20—Prazosin	64	887	PO	1-comp	3	(18)
PK21—Pyrazinamide	227	3092	PO	1-comp.	3	(19)
PD1—Levodopa	19	851	PO	Direct Imax.	4	(20)
PD2—Chemotherapies	636	3549	IV	Semimechanistic	4	(21)
PD3—Tesaglitazar	413	4035	PO	Indirect effect, Emax	3	(22)
PD4—Tesaglitazar	413	4663	PO	Mechanism based, built upon PD3	5	(22)
PD5—Cladribine	59	332	*in vitro*	EMAX model	2	(23)
PD6—Moxonidine	97	1942	PO	Imax.	3	(24)
PD7—Moxonidine	97	1944	PO	Direct inhibitory Emax.	2	(24)
PD8—Digoxin	225	787	IV	Linear effect	3	(25)
PD9—Tesaglitazar	94	1337	PO	1-comp conc dependent elim.	5	(26)

[a] PO: Per oral
[b] IV: Intravenous

Simulations example

In these simulations, the distribution of interindividual random effects for clearance was simulated using a mixture of four sub-distributions with varying variances, see Table I. The random effects were transformed with a fixed exponential transformation. The distributions were chosen so that the end result was a skewed unimodal distribution (see Fig. 2). A one-compartment disposition first order absorption and elimination model was used to simulate 100 datasets of sparse data with 25 individuals with 3 or 7 observations each, 100 datasets with 50 individuals with 3 observations each, and 100 datasets of richer data with 500 individuals with 7 observations each. K_a and volume of distribution were set to have typical values of 1 h^{-1} and 250 L, respectively, both with an interindividual variability of 30%. To the datasets were then fitted models that included a lognormal distribution model without or with each of the three transformations with estimable shape parameters. The true mixture model was also fitted to the datasets for comparison. The estimated parameter distributions for all models were compared to the simulated distribution by calculating the relative error from the parameter values at the 10, 20, 30, 40, 50, 60, 70, 80, and 90th percentiles (see Fig. 2 and Eq. 4). The difference in objective function values (ΔOFV) between models with and without flexible transformations was also calculated for each model and dataset. The ΔOFV is expected to be approximately chi-square distributed, with degrees of freedom equal to the number of estimated shape parameters of the flexible transformation, and can be used in the likelihood ratio test for hypothesis tests. The objective function value (OFV) is equal to minus 2 times the log likelihood of the data given the parameters. The cut-off values used for inclusion of a transformation were 3.84 for one parameter transformations and 5.99 for the two parameter transformations. Wälhby *et al.* has shown that the type I error rates and cut-off values for the likelihood ratio test under certain conditions do not follow the χ^2 distribution (4,5). Therefore, the type I error rates and cut-off values for significant inclusion for the shape parameters were evaluated and found to concur with the nominal values. (Appendix I).

The simulation properties of one model with a fixed lognormal parameter distribution and one where an estimated box-cox transformation was used was also illustrated by performing visual predictive checks (VPC). 1000 dataset were simulated using a model estimated with a standard lognormal distribution, and 1000 datasets were simulated using an estimated box-cox transformed distribution. This simulated data was used to construct 95% confidence intervals for the 2.5 and 97.5 percentiles as well the median. These confidence intervals could then be plotted together with the observed 2.5, 50, and 97.5 percentiles. This was done with a model fitted to a dataset with 500 individuals and 7 observations.

Used for these calculations was R 2.4.0 or higher; the NONMEM version used was NONMEM VI (Iconus, Hanover, MD). The programs PsN (psn.sf.net) and Xpose (xpose.sf.net) were used to perform the VPCs (6,7).

$$RPE = \frac{P_{true} - P_{estimated}}{P_{true}} \tag{4}$$

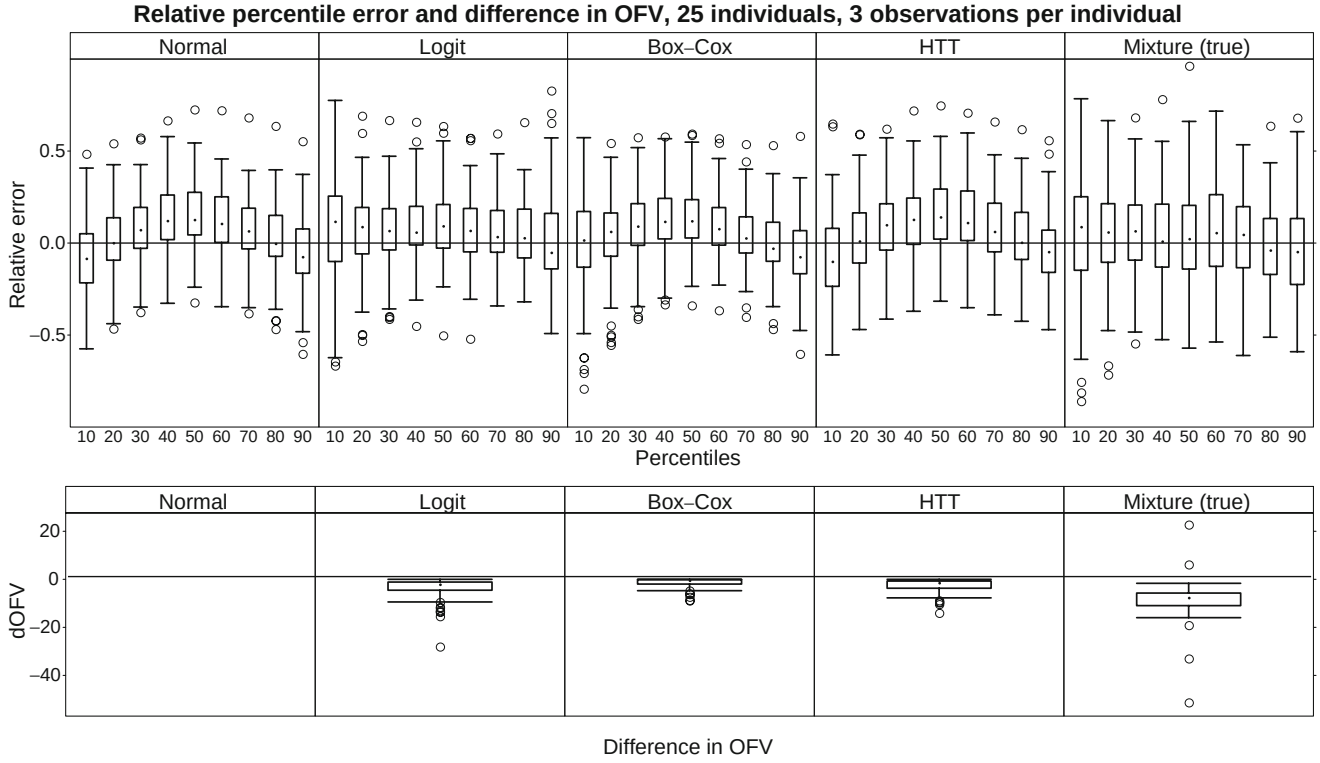

Fig. 3. The relative errors of the investigated percentiles and the differences in OFV between fixed lognormal distribution (Normal) and the other models for the rich data with 25 individuals and 3 observations per individual.

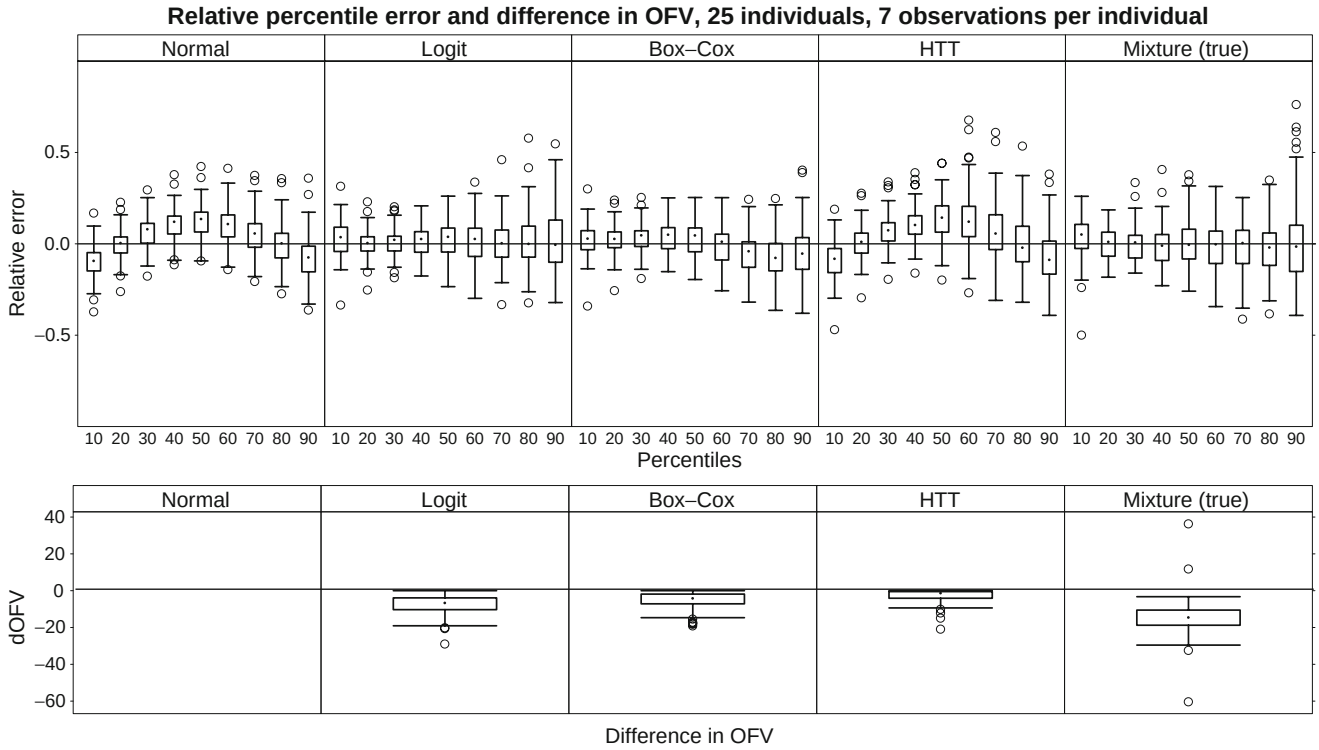

Fig. 4. The relative errors of the investigated percentiles and the differences in OFV between normal distribution and transformed ones. This is from the data with 25 individuals and 7 observations per individual.

Real Models Examples

30 population models were evaluated using the three different transformations of the Gaussian distribution. The models evaluated had a wide range of number of individuals and observations, from sparse to rich data (2–45 observations per subject). Both PK and PD models were represented among these models. The models are summarized in Table II.

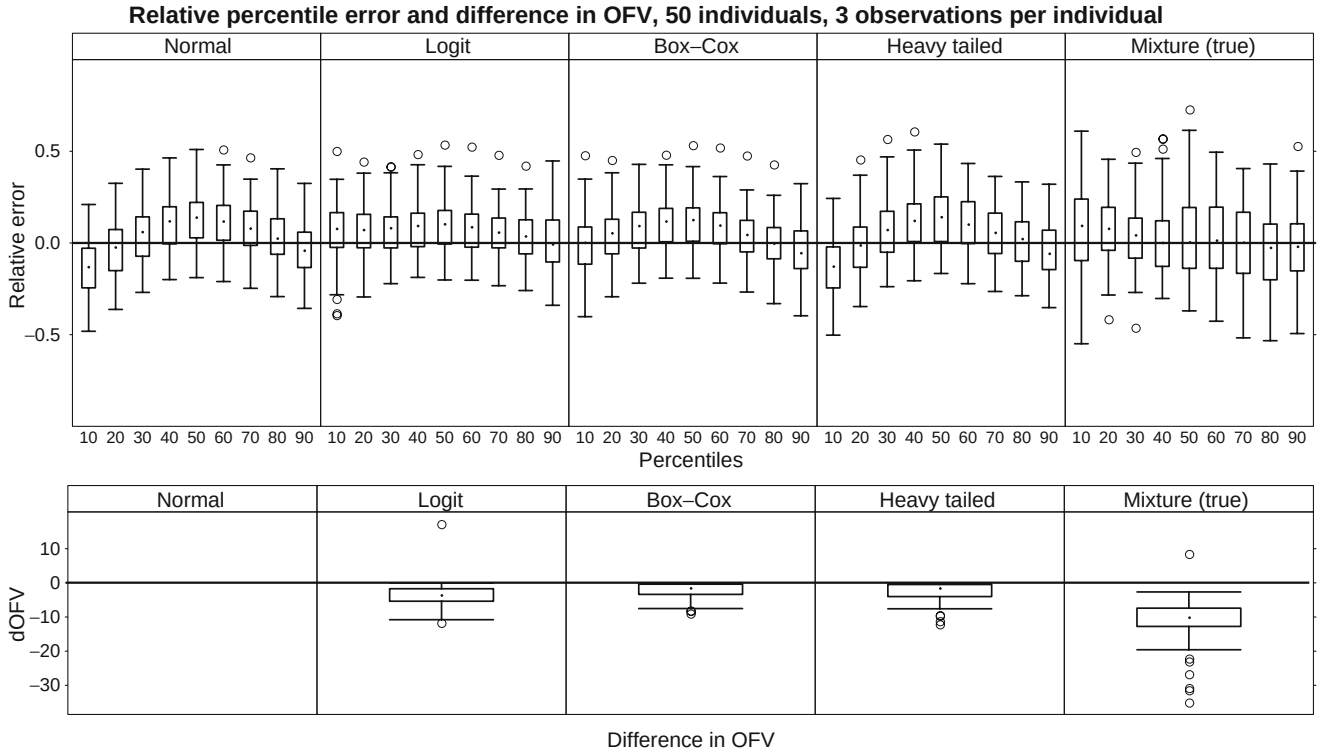

Fig. 5. The relative errors of the investigated percentiles and the differences in OFV between normal distribution and transformed ones. This is from the data with 50 individuals and 3 observations per individual.

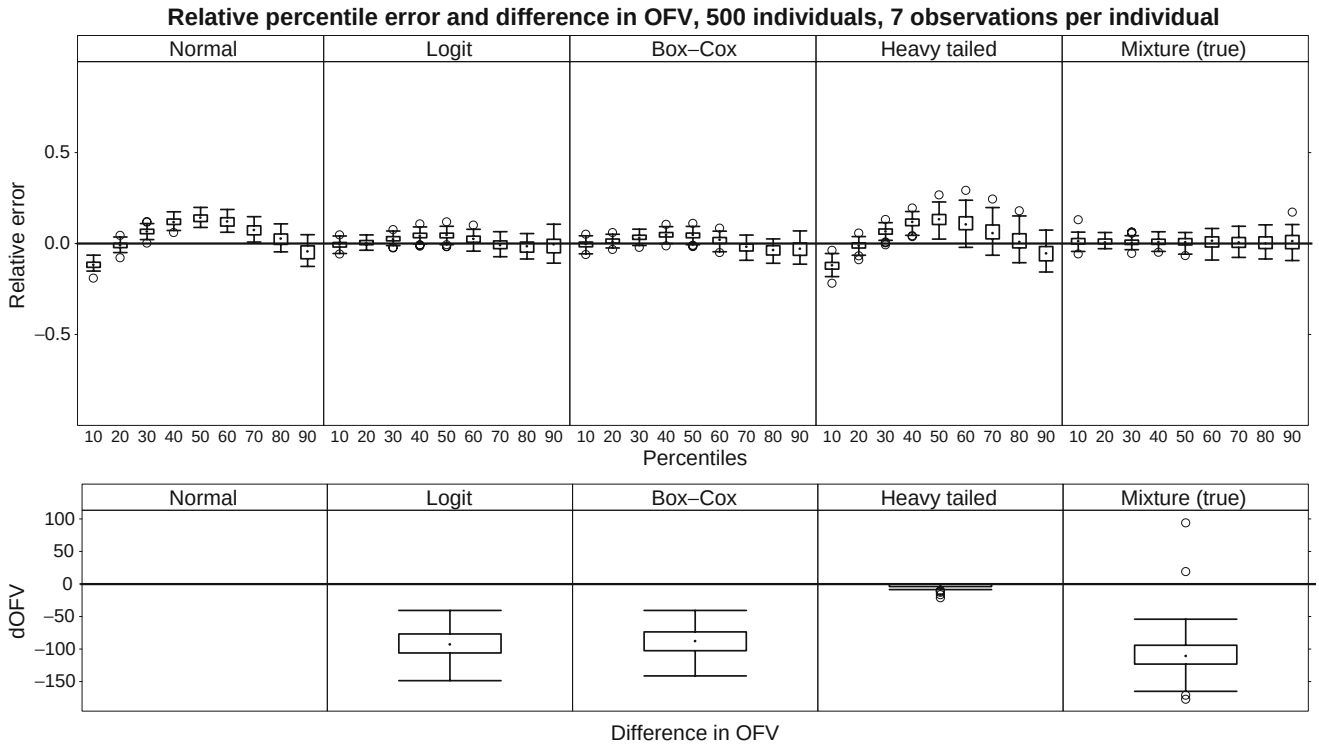

Fig. 6. The relative errors of the investigated percentiles and the differences in OFV between normal distribution and transformed ones. This is from the data with 500 individuals and 7 observations per individual.

For the purpose of clarity, the models are hereafter identified with numbers rather than by names; references are indicated if available. The transformations were added to one ETA-distribution at a time, and, if the inclusion was found to be significant ($p<0.05$) based on the likelihood ratio test, the transformation was included on additional parameters within the same model. The three different transformations were tested separately and were not applied to the same model simultaneously.

RESULTS

Simulations Example

In the simulation study, the inclusion of an estimated transformation that allows a skewed distribution reduced the biases at the percentiles investigated compared to the fixed exponential transformation, both with rich and sparse data (Figs. 3, 4, 5 and 6). Compared to the true model, the logit

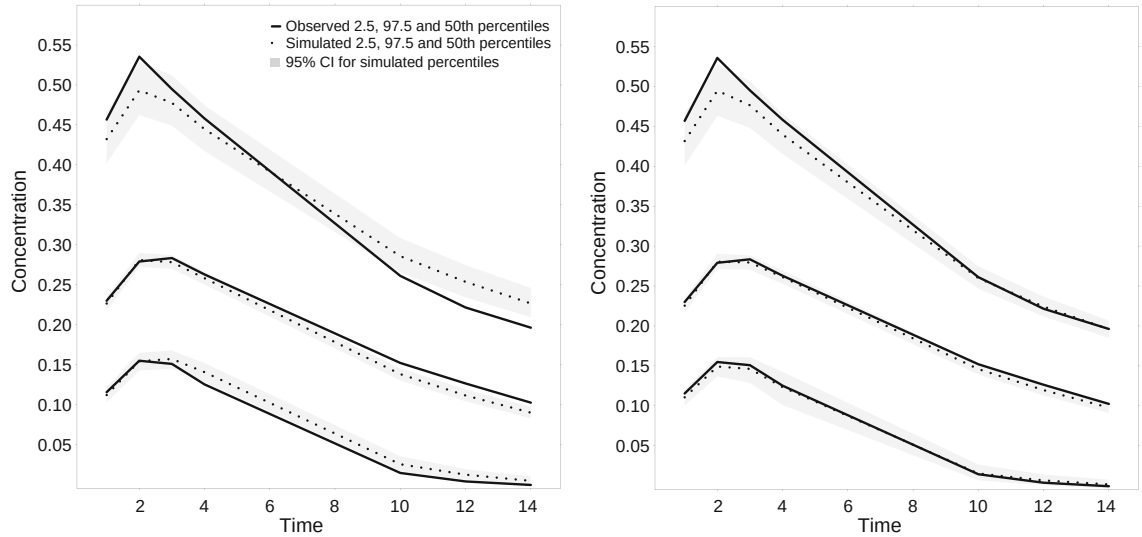

Fig. 7. Visual predictive checks for a misspecified lognormal clearance distribution (left) and after inclusion of a box-cox transformation with estimated shape parameter (right). 95 percentiles and median of observations are compared to the 95% confidence intervals of simulated 95% prediction intervals (PI) and median.

Table III. Logit Transformation Resulting in Significant ($p<0.05$) Improvement of Models for Real Data. Where Several Parameters are Transformed in the Same Model, ΔOFV is for Adding the Additional Parameter. Parameter Estimates are from the Final Model with all Significant Transformations

Model No.	ΔOFV	Parameter(s)	Θ_1	Θ_2	ω^2
PK1	−13.8	CL	0.37	0.832	3.01
PK2	−17.5	BASE[b]	0.0033	2.57	46.9
PK3	−23.4	CL	0.99	92.1	0.374
	−13.0	V3	0.99	0.789	84.3
PK4	−16.9	Ka	0.99	269	0.536
PK5	−6.8[a]	CL	0.99	179	0.191
	−13.8	V2	0.94	1.78	4.61
PK6	−12.5	CL	0.86	7.48	0.266
	−20.3	Km	0.0020	399	5.82
	−18.1	F	0.93	18.3	0.365
PK7	−6.6[a]	Ka	0.64	3.6	4.85
PK9	−8.3	V	0.49	2.5	1.39
PK10	−6.0[a]	Ka	0.77	6.02	0.26
PK13	−26.6	Ka:	0.011	8.25	9.17
	−14.0	IOV[c] K*a*	0.013	189	0.717
PK15	−7.7[a]	V2	0.96	0.37	26.2
PK16	−7.6[a]	V2	0.96	0.363	26.1
	−8.3[a]	MTT[d]	0.4	0.634	39.6
PK18	−6.2[a]	CL	0.24	3.45	0.804
PK20	−9.5	Ka	0.005	321	1.43
PK21	−7.0	V	0.46	1.73	0.117
PD1	−53.9	EC50	0.85	0.761	4.22
PD2	−19.7	MTT	0.95	8.89	0.0861
	−17.2	SL[e]	0.91	17	0.0729
PD3	−38.0	BASE	0.36	0.607	1.3
	−15.0	PLAC[f]	0.33	2.21	0.572
	−23.7	EC50	0.91	17.4	4.79
	−79.5	RV[g]	0.15	1.07	5.13
PD4	−94.9	RV1	0.045	1.1	11.5
	−144.3	RV2	0.041	1.04	5.05
PD5	−6.7[a]	EMAX	0.021	195	0.147
	−22.1	BASE	0.21	1.05	1.17
PD7	−17.5	BASE	0.98	15.2	0.198
	−16.4	BASE	0.73	5.13	34.1
PD8	−26.4	BAST[h]	0.69	1.09	0.691

[a] Drop might be borderline significant given number of individuals and/or empirical cut-off values
[b] BASE: Baseline
[c] IOV: Inter occasion variability
[d] MTT; Mean transit time
[e] SL: Slope
[f] PLAC: Placebo
[g] RV: Residual variability
[h] BAST: Time varying baseline

transformation performs worse under the more information rich circumstance but better when data is sparse. Testing the models' simulation performances for an individual data set using VPCs showed that a misspecification of the parameter distribution by using a fixed lognormal distribution leads to a visibly poorer agreement with observations compared to a model based on box-cox transformations (Fig. 7). Including a transformation that can adapt to the skewed distribution corrected for this over-prediction. In these examples the stability was high for the standard lognormal, logit, box-cox, and HT transformations when data was rich, with 100, 98, 100, and 97% successful minimizations, respectively. For the mixture model 89% minimized successfully. When data was sparser the corresponding values were 98, 92, 97, 95, and 92%.

Real Models Examples

When the transformations were implemented into existing models, a significant drop in OFV was seen in 22, 18, and 22 of the models for logit, box-cox, and HTT, respectively (Tables III, IV, and V and Figs. 8 and 9). The largest drops in OFV could be seen when transforming distributions of inter-individual random effects on the residual error (ETA-on-EPS) with the transformations allowing skewed distributions. For the logit transformation 14 of 35 transformed parameters the estimated variance will yield a transformed distribution with a distinct bimodal shape. 13 of the HT transformed parameters had negative shape parameters, i.e. bimodality, and 18 had positive values. Parameter distributions were about as often left skewed as right skewed.

Table IV. Box-cox Transformation Resulting in Significant ($p<0.05$) Improvement of Models for Real Data. Where Several Parameters are Transformed in the Same Model, ΔOFV is for Adding the Additional Parameter. Parameter Estimates are from the Final Model with all Significant Transformations

Model No.	ΔOFV	Parameter(s)	Θ_1	ω^2
PK2	−9.4	BASE[b]	7	0.074
PK3	−23.5	CL	−1.03	0.33
PK4	−17.0	Ka	−0.74	0.96
PK5	−6.7	CL	−1.11	0.15
	−13.8	V2	−3.76	0.070
PK6	−18.4	F	−0.72	0.42
PK7	−4.9	V	−0.66	0.26
PK9	−13.1	CL	−0.65	0.21
PK10	−12.1	CL	−1.34	0.16
PK12	−4.0[a]	Ka	−0.48	0.94
PK13	−35.9	Ka	0.77	2.45
PK18	−4.7	CL	0.66	0.24
PD1	−48.3	Ke	0.12	0.24
PD2	−19.7	MTT[c]	−2.18	0.016
	−17.2	SL[d]	−0.65	0.13
	−13.0	GAM[e]	3.37	0.017
PD3	−22.6	BASE	1.85	0.018
	−9.1	PLAC[f]	−0.45	1.04
	−75.7	RV[g]	2.97	0.072
PD4	−30.3	RV1	−1.11	0.13
	−68.1	RV2	−0.12	0.039
PD5	−6.4	EMAX	3.81	2.3
	−19.6	BASE	0.25	0.025
PD7	−17.5	BASE	−3.57	0.014
PD8[i]	−27.2	BAST[h]	−0.68	22.6

[a] Drop might be borderline significant given number of individuals and/or empirical cut-off values
[b] BASE: Baseline
[c] MTT; Mean transit time
[d] SL: Slope
[e] GAM:Hill factor
[f] PLAC: Placebo
[g] RV: Residual variability
[h] BAST: Time varying baseline
[i] This model is omitted in Figs. 8 and 9

Table V. Heavy Tailed Transformation Resulting in Significant ($p < 0.05$) Improvement of Models for Real Data. Where Several Parameters are Transformed in the Same Model, ΔOFV is for Adding the Additional Parameter. Parameter Estimates are from the Final Model with all Significant Transformations

Model No.	ΔOFV	Parameter(s)	Θ_1	ω^2
PK2	−8.5	BASE[b]	0.67	0.80
	−11.5	CL	0.42	0.46
	−5.0[a]	V	0.28	0.69
PK3	−12.8	V3	1.36	0.50
PK4	−15.8	Ka	−0.33	1.32
	−42.8	RV[c]	0.29	0.051
PK5	−4.6[a]	CL	0.43	0.30
	−5.6	V2	0.30	0.27
PK6	−35.2	Km	−0.57	6.10
	−7.2	RV	−0.25	0.14
PK7	−9.4	Ka	−0.28	4.51
PK8	−13.1	CL	0.57	0.33
PK9	−10.1	CL	0.25	0.27
PK12	−10.0	N[d]	0.58	2.90
PK15	−9.3	Ke	0.93	0.097
PK16	−7.4[a]	Ke	−0.66	0.13
	−9.1	V2	1.30	0.00015
	−4.8[a]	MTT[e]	−0.61	0.0014
PK17	−4.7[a]	CL	0.93	0.40
	−7.0	V	0.31	0.23
PK18	−6.48	V	−0.31	0.22
PK19	−47.0	CL	−0.066	0.079
PK21	−21.6	V	0.21	0.038
PD1	−5.2[a]	Ke	−0.18	1.09
PD2	−4.6[a]	BASE	0.13	0.086
PD3	−8.4	BASE	−0.32	0.0096
	−10.1	EC50	−0.17	0.47
	−18.8	PLAC[f]	−0.32	0.76
PD4	−4.4[a]	KIN[g]	−0.25	0.0076
	−11.3	RV	0.10	0.067
PD5	−5.9	EMAX	0.41	1.8
PD6	−24.2	CL	−0.32	1
PD9	−12.9	V	0.41	0.12

[a] Drop might be borderline significant given number of individuals and/or empirical cut-off values
[b] BASE: Baseline
[c] RV: Residual variability
[d] N; Number of transit comp.
[e] MTT:mean transit time
[f] PLAC: Placebo
[g] KIN: Red blood cell release

DISCUSSION

Implementation of one of the transformations evaluated in this paper is an easy way to allow deviations from the standard, fixed parameter distributions that are traditionally used in parametric mixed effects population modeling. The implementation of the transformations into the NONMEM code is straightforward. Improvement of the model's fit to real data could be seen in about two thirds of the models for each of the three transformations. Considering individual parameters, there were 112 parameter distributions in the 30 models; 35, 25, and 33 could be improved with the logit, box-cox, and HT transformation.

Often both transformations that allow a skewed distribution (logit and box-cox) show significant drops in OFV for the same models and same parameters, and the estimated distributions had similar shape. This implies that the distributions estimated are closer to the true ones than the fixed transformations. The estimated distribution shapes of models with few individuals should be interpreted with caution. For such small data sets the likelihood ratio test using nominal cut-off values may not always be correct (see Appendix) resulting in too frequent selection of the more complex model.

The simulated skewed distribution was better approximated with a box-cox or logit transformation that could quite adequately adapt to the deviation from lognormality. As expected the HT transformation, which is only able to transform into symmetrically heavy-tailed distributions, performed no better than the identity transformation. Box-cox and logit transformations and behaves reasonably well compared to the true mixture model considering that they use 5 and 4 fixed parameters and 3 random parameters less than the mixture model. The mixtures estimated by the mixture model generally showed little resemblance with the ones simulated, especially when data was sparse. Model stability was also lower and runtimes longer for the mixtures. Thus is the, in this case, more mechanistic mixture model not well suited to explain the phenomenon of a unimodal, skewed distribution. The more empirical transformation approach appears more appropriate. When dealing with truly bimodal or multimodal distributions, or when prior parameter information exists, the mixture model would be preferred even if two of the flexible transformations are able to create bimodal distributions. The mean relative errors at the investigated percentiles are zero already at 25 individuals with rich data for the logit transformation, but with sparse data there is some bias. Further investigations (results not shown) show that this is the case with 7 observations and 50 individuals too. This implies that to correctly characterize the distribution there needs to be sufficient information on the individual etas, otherwise the eta distribution shape probably becomes distorted or there is shrinkage. Still, even when individual data is sparse, the errors look better than using the normal, and the drops in OFV indicate that the estimated distributions are offering a better fit than the standard lognormal. The improvement in simulation properties seen in the VPC may seem small, but the distribution used to simulate did not deviate as much from lognormality as some distributions estimated from real data. Concerning simulation studies, the logit transformation offers a nice feature. In these studies, truncated parameter distributions are sometimes used. With the estimated logit transformation the parameter distribution is automatically truncated at $-\theta_1 * \theta_2$ (lower bound) and $(1- \theta_1) * \theta_2$ (upper bound). For the other transformations, no such automatic truncation occurs.

Based on these results it appears as if PD models improve more by introducing a transformation that allows the distributions to be skewed as the largest drops in OFV can be seen with box-cox and logit transformations. A larger number of PK models, on the other hand, seem to benefit more from the transformation that allows a heavy tailed distribution. Also, baseline parameters often appear to benefit from transformation inclusion in the PD models. This could, for example, be due to parameter distributions being truncated by study inclusion criteria. For PK models, no parameter stands out in the same way.

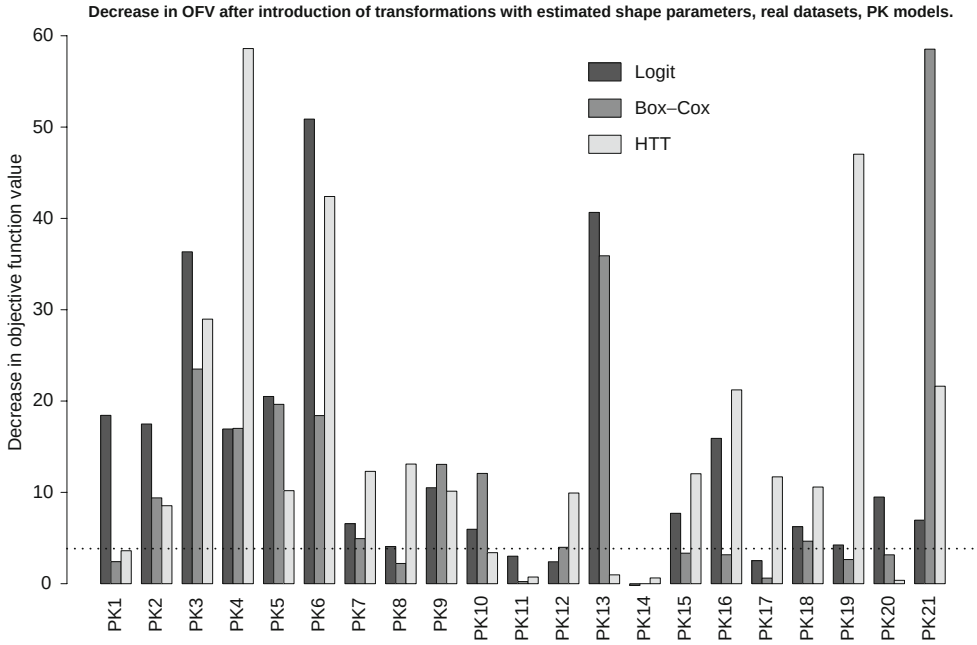

Fig. 8. Improvement in goodness-of-fit expressed as a decrease in OFV after implementation in all real PK models. The dotted line represents a significance level of 0.05 for inclusion of one model parameter.

Even if runtimes were not directly investigated in this work, we did not notice any significant effect when implementing these estimated transformations. One should also note that in the simulation example the estimated logit transformation on one occasion and the mixture model on three occasions ended up in local minima, here seen as dOFVs higher than 0 (Figs. 3, 4, 5 and 6).

There are tendencies that models with larger numbers of individuals have larger decreases in OFV with estimated transformations. This suggests that most individuals benefit from the new distributions, not only outlying individuals. Some tests made (results not shown) where individual OFVs were calculated support this statement (27). More individuals give better defined distributions and give more precise estimates of the shape parameters.

That a significant decrease can be seen for a parameter both with transformations that allow skewed distributions and with the HT transformation, which is symmetrical, might seem counterintuitive, but the subsequent fixed exponential transformation can transform these HT distributions into quite skewed distributions. This could also explain some models where the HT shape parameters were negative, but

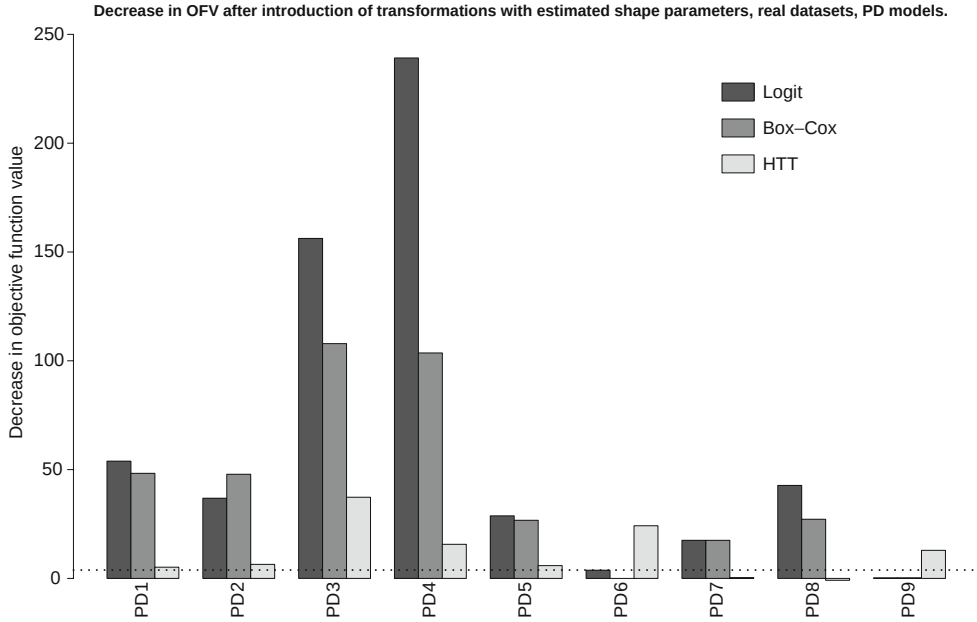

Fig. 9. Improvement in goodness-of-fit expressed as a decrease in OFV after implementation in all real PD models. The dotted line represents a significance level of 0.05 for inclusion of one model parameter.

there were no tendencies in the empirical bayes estimates that the distribution was bimodal. The combination of HT and exponential transformations creates a unimodal distribution that resembles the true one significantly better than the lognormal. The HT transformation could be applied to either only positive or only negative ηs to investigate if the shape is truly symmetrically heavy tailed or only one sided.

The shapes of underlying parameter distributions are not easy to assess using standard goodness-of-fit diagnostics, and neither are the misspecifications of the applied distributions easy to diagnose. Empirical bayes estimates are often not informative, due to η-shrinkage, unless all subjects have rich information about the parameter in question. These transformations can therefore be used as a tool to investigate parameter distribution assumptions in the final stage of model development. A lack of improvement when using estimated transformations would be a strong indication that the selected fixed transformations are appropriate. When used for this purpose, these transformations would be a useful complement to the already existing nonparametric methods which have shown to be powerful for detecting non-standard shapes. For describing such distributions, however, these transformations offer several advantages compared to nonparametric methods in that all the tools available for model building, evaluation, and utilization that apply to parametric methods also apply to these models. It seems suitable to try these transformations subsequent to or simultaneously to covariate modeling. In this study, we did, for illustrative purposes, not mix different types of transformations within one model. For other applications, the best model may well be obtained with a mixture of transformation types, logit, box-cox, and/or HT. In the case where more than one transformation type falls out as significant for the same parameter, OFV and simulation properties should guide model selection.

The transformations may induce a change in variability, such that the estimate of parameter variability is different after introduction of a semiparametric transformation. The SDs of the estimated transformed distributions compared to the estimated SDs using the standard lognormal distribution are shown in Fig. 10. It appears as the SD of the estimated distributions, lognormal or transformed, are quite similar in the majority of cases. This shows that whether the estimated shape of the distribution is correct or not, using the transformations does not to any large extent change the estimation of the variability. Stated differently, this indicates that even in cases when standard parametric modeling fails to appropriately recognize the shape of a parameter distribution, the estimated magnitude of variability is generally adequate. In the cases where there is a larger difference in variability between the lognormal and the transformed distributions, the estimated shape parameters deviate markedly from the unity transformation values, or the variances for either the lognormal or transformed distributions are large.

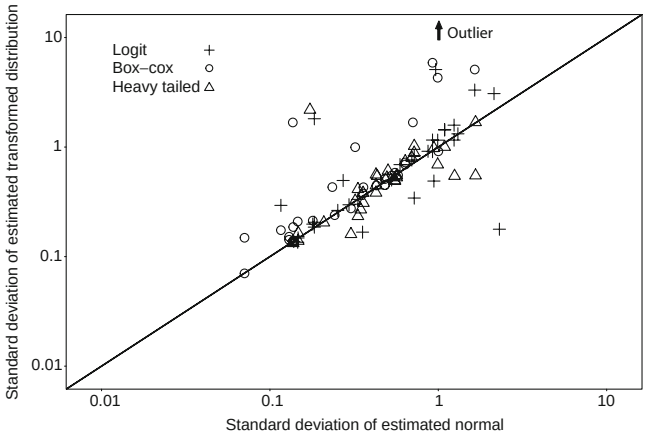

Fig. 10. Comparison of the estimated variability of a transformed distribution *versus* the estimated variability of a corresponding lognormal distribution. One extreme outlier is omitted.

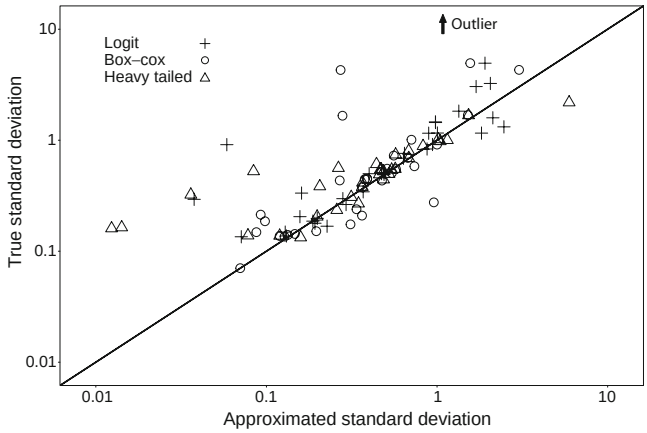

Fig. 11. The true variability in transformed distributions *versus* an approximation of that variability. The approximations are the estimated standard deviations before transformation for box-cox and HT transformations and $(1\text{-}\theta_1)*\theta_1*\theta_2$ times the standard deviation before transformation for logit. One extreme outlier is omitted.

The transformations may make it harder to assess the magnitude of parameter variability directly from parameter estimates. Fortunately, there are approximations that may be suitable for an assessment. For the logit transformation, SD ($\eta_{i,transformed}$) can be approximated by multiplying the estimated ω with $(1\text{-}\theta_1)\theta_1\theta_2$. This approximation seems to be reasonable in the ranges of ω found in the models based on real data (Fig. 11). For the box-cox transformation, ω and SD($\eta_{i,transformed}$) are approximately equal whenever the former is small. For example, SD($\eta_{i,transformed}$) will be 0.319 when θ_1 is 1 or −1 and ω is 0.3. However, this approximation will be less appropriate as estimated ω increases and θ_1 is more different from zero. For the heavy-tailed transformations ω and SD($\eta_{i,transformed}$) are approximately equal whenever θ_1 is in the range −0.5 to 0.5; it becomes less appropriate with ω larger than 1, and negative θ_1 generally means a worse approximation.

The fact that for the box-cox and the logit transformation, the transformation only approaches the identity transformation as the shape parameters approaches their corresponding values tells that the models are not truly nested. However, there exist in the shape parameter space values where the difference between $\eta_{i,}$ and $\eta_{i,transformed}$ is so small that there is no change in OFV. This gives the transformations the property of being practically nested from a numerical point of view rather than strictly analytically. The behavior of the transformations when put trough the likelihood ratio test also implies that they can be assumed to be nested models as the statistic follow a χ^2-distribution. This assumption would break down as the number of individuals approaches infinity. Even if the models are not strictly

mathematically nested the cut-off values can be used as guides for statistical inclusion. If one would consider the models not to be nested, the Akaike information criterion would be an appropriate tool to distinguish the better model

In conclusion, the shapes of parameter distributions can often be better approximated by flexible transformations with estimated shape parameters. Implementation into real models caused significant drops in OFV in two thirds of the models, implying that in model building, more often than not, the parameter distribution assumptions are violated.

ACKNOWLEDGEMENTS

Moa Grahm for technical assistance and Johnson & Johnson Pharmaceutical Research & Development for sponsoring the doctoral studies of Klas Petersson.

APPENDIX I. ASSESSMENT OF THE TYPE I ERROR RATES

When building models using maximum likelihood as in NONMEM, the likelihood ratio test is often used to test if the inclusion of a new parameter that is to be estimated into the model is statistically significant. Under asymptotic conditions the likelihood ratio statistic is $\chi 2$ distributed, which implies that a drop in OFV needs to be larger than 3.84 for an inclusion of a new parameter, corresponding to one degree of freedom, to be statistically significant at $p<0.05$ (28). Monte Carlo simulations were performed to investigate if the likelihood ratio statistic of the shape parameters estimated in these transformations are $\chi 2$ distributed with the same number of degrees of freedom as additional parameters estimated, which would be in accordance with statistical theory. This kind of investigation would also test the robustness of the likelihood ratio test as it has been done with other model misspecifications (4,5).

Datasets were simulated from a model with a constant rate infusion at steady state, thus parameterized only with one structural parameter; CL, modeled with interindividual variability, this simple model was chosen for runtime reasons. Simulations were performed with transformations included, but the shape parameters set to values that would give no change in the shape of the distributions. These values were 0.00001 for box-cox, 0.5 and 4 for the two parameters of the logit transformation, and 0 for HTT. The simulation settings were altered and combined with respect to number of individuals and number of observations. The number of individuals ranged between 25 and 500, and observations ranged between 2 and 19. For each combination, 1000 datasets were simulated. The datasets were then estimated both with the full model with transformation as well as with the reduced model, i.e. the standard lognormal distribution. The OFV values from the estimation with both models were then compared for each simulated dataset, and the drop in OFV which produced an error rate of 5% was calculated. Expected values would be 3.84 for box-cox and HTT and 5.99 for logit because of its two parameters.

The results of the assessment of the type I error rates through simulations showed that the cut-off values for inclusion of a transformation of a parameter distribution were drops in OFV of 7, 4, and 5 for the logit, box-cox, and HT transformations, respectively. These values do not differ to any large extent from the nominal values of 3.84 and 5.99. These error rates were considered to be stable from 50 individuals and up. At lower numbers of individuals (25) the values are slightly elevated (see Table VI, which shows the average cut-off values of from varying the number of observations per individual).

Table VI. Average Empirical OFV Cut-off Values Across Different Number of Individuals

No. of Ind.	Logit	Box-cox	Heavy tailed
25	8.2	4.1	6.6
50	7.1	4.0	5.3
250	7.0	3.8	4.6
500	7.4	3.9	4.4

APPENDIX II EXAMPLES OF NM-TRAN CODE

Logit transformation

```
TVCL=THETA(1)
LGPAR1 = THETA(2)
LGPAR1 = THETA(3)
PHI = LOG(LGPAR1/(1-LGPAR1))
PAR1 = EXP(PHI+ETA(1))
ETATR = (PAR1/(1+PAR1)-TVTH)*LGPAR2
CL=TVCL*EXP(ETATR)
```

Box-Cox transformation

```
TVCL=THETA(1)
BXPAR=THETA(2)
PHI = EXP(ETA(1))
ETATR = (PHI**BXPAR-1)/BXPAR
CL=TVCL*EXP(ETATR)
```

Heavy tailed transformation

```
TVCL=THETA(1)
HTPAR=THETA(2)
ETATR=ETA(1)*SQRT(ETA(1)*ETA(1))**HTPAR
CL=TVCL*EXP(ETATR)
```

REFERENCES

1. Bressolle F, Gomeni R. Predictive performance of a semi-parametric method to estimate population pharmacokinetic parameters using NONMEM. J Pharmacokinet Biopharm. 1998;26:349–61.
2. Davidian M, Gallant AR. Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine. J Pharmacokinet Biopharm. 1992; 20:529–56.
3. Fattinger KE, Sheiner LB, Verotta D. A new method to explore the distribution of interindividual random effects in non-linear mixed effects models. Biometrics. 1995;51:1236–51.
4. Wahlby U, Jonsson EN, Karlsson MO. Assessment of actual significance levels for covariate effects in NONMEM. J Pharmacokinet Pharmacodyn. 2001;28:231–252.

5. Wahlby U, Bouw MR, Jonsson EN, Karlsson MO. Assessment of type I error rates for the statistical sub-model in NONMEM. J Pharmacokinet Pharmacodyn. 2002;29:251–269.
6. Lindbom L, Ribbing J, Jonsson EN. Perl-speaks-NONMEM (PsN)–a Perl module for NONMEM related programming. Comput Methods Programs Biomed. 2004;75:85–94.
7. Jonssonand EN, Karlsson MO. Xpose–an S-PLUS based population pharmacokinetic/pharmacodynamic model building aid for NONMEM. Comput Methods Programs Biomed. 1999;58:51–64.
8. 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;55:44–551.
9. 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.
10. Lindemalm S, Savic RM, Karlsson MO, Juliusson G, Liliemark J, Albertioni F. Application of population pharmacokinetics to cladribine. BMC Pharmacol. 2005;5:4.
11. 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.
12. Osterberg O, Savic RM, Karlsson MO, Simonsson US, Norgaard JP, Walle JV, *et al.* Pharmacokinetics of desmopressin administrated as an oral lyophilisate dosage form in children with primary nocturnal enuresis and healthy adults. J Clin Pharmacol. 2006;46:1204–11.
13. 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 Pharmacokinet Biopharm. 1998;26:207–46.
14. 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.
15. Iliadis MC, Bruno R, Lacarelle B, Cosson V, Mandema JW, Le Roux Y, *et al.* Evaluation of Bayesian estimation in comparison to NONMEM for population pharmacokinetic data. J Pharmacokinet Biopharm. 1992;20:653–69.
16. 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;98:14–1723.
17. Vozeh S, Aarons L, Wenk M, Weiss P, Follath F. Population pharmacokinetics of tobramycin. Br J Clin Pharmacol. 1989;28: 305–14.
18. Karlssonand MO, Sheiner LB. The importance of modeling interoccasion variability in population pharmacokinetic analyses. J Pharmacokinet Biopharm. 1993;21:735–50.
19. 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.
20. 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.
21. 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.
22. 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).
23. 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;88:324–2.
24. McNay JL, Brynne L, 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.
25. 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.
26. Hamrén, B. Safety and Efficacy Modelling in Anti-Diabetic Drug Development., *Department of Pharmaceutical Biosciences, Division of Pharmacokinetics and Drug Therapy*, Vol. Doctor, Uppsala University, Uppsala, 2008, p. Paper II.
27. Sadray S, Jonsson EN, Karlsson MO. Likelihood-based diagnostics for influential individuals in non-linear mixed effects model selection. Pharm Res. 1999;16:1260–5.
28. Sheiner LB, Beal SL. NONMEM Users Guide. (1992).