
Research Article

Model-Based Residual Post-Processing for Residual Model Identification

Moustafa M. A. Ibrahim,^{1,2} Rikard Nordgren,¹ Maria C. Kjellsson,¹ and Mats O. Karlsson^{1,3}

Received 9 March 2018; accepted 7 June 2018; published online 2 July 2018

Abstract. The purpose of this study was to investigate if model-based post-processing of common diagnostics can be used as a diagnostic tool to quantitatively identify model misspecifications and rectifying actions. The main investigated diagnostic is conditional weighted residuals (CWRES). We have selected to showcase this principle with residual unexplained variability (RUV) models, where the new diagnostic tool is used to scan extended RUV models and assess in a fast and robust way whether, and what, extensions are expected to provide a superior description of data. The extended RUV models evaluated were autocorrelated errors, dynamic transform both sides, inter-individual variability on RUV, power error model, t-distributed errors, and time-varying error magnitude. The agreement in improvement in goodness-of-fit between implementing these extended RUV models on the original model and implementing these extended RUV models on CWRES was evaluated in real and simulated data examples. Real data exercise was applied to three other diagnostics: conditional weighted residuals with interaction (CWRESI), individual weighted residuals (IWRES), and normalized prediction distribution errors (NPDE). CWRES modeling typically predicted (i) the nature of model misspecifications, (ii) the magnitude of the expected improvement in fit in terms of difference in objective function value (Δ OFV), and (iii) the parameter estimates associated with the model extension. Alternative metrics (CWRESI, IWRES, and NPDE) also provided valuable information, but with a lower predictive performance of Δ OFV compared to CWRES. This method is a fast and easily automated diagnostic tool for RUV model development/evaluation process; it is already implemented in the software package PsN.

KEY WORDS: conditional weighted residuals; diagnostics; model evaluation; nonlinear mixed effects models; residual error model.

INTRODUCTION

Nonlinear mixed effects (NLME) models are widely used to describe clinical data in drug development for learning about the underlying physiological system, confirming drug effects, simulating different scenarios for dose selection, and decision making (1). These models incorporate mathematical description of structural components (fixed effects) and stochastic components (random effects). Random effects account for multiple sources of variability: inter-individual variability (IIV), inter-occasion variability (IOV), and residual unexplained variability (RUV). RUV incorporates intra-individual variability, inaccuracies in dosing and sampling

history, measurement errors, and model misspecification. It is assumed that the RUV model allows the transformation of a normal, independent, and identically distributed random variable to capture the full complexity of any heteroscedasticity, non-normality, and dependence in the RUV between the model and data. It has been shown that maximum likelihood estimation of model parameters with misspecified RUV model results in biased parameter estimates (2), and conclusions regarding covariate inclusion or parameter uncertainty based on such fit may not be correct (3,4). Thus, selection of an appropriate RUV model is important for maximum likelihood estimation, real-life-like simulations, better utilization of data, and subsequent model-informed decisions.

Extended RUV models have been developed to various RUV characteristics. An autoregressive (AR1) error model has been used to describe serially correlated residuals which may occur with rich data in particular (2,5). One example of this was a case study of glucose tolerance test, where AR1 errors provided a better description of the data (6). A dynamic transform both sides (dTBS) approach account for skewness and scedasticity of the residuals by estimating a

Electronic supplementary material The online version of this article (<https://doi.org/10.1208/s12248-018-0240-7>) contains supplementary material, which is available to authorized users.

¹Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden.

²Department of Pharmacy Practice, Helwan University, Cairo, Egypt.

³To whom correspondence should be addressed. (e-mail: mats.karlsson@farmbio.uu.se)

shape parameter (λ) and a power term (ζ) using a Box-Cox transform to both data and model predictions (7,8). Inter-individual variability IIV in the magnitude of RUV relaxes the assumption that all subjects display a common residual error variance (9). A power error model accounts for residual error scedasticity dependent on model predictions and as such offers an alternative to the commonly used additive plus proportional RUV model. A Student's t-distributed error model allows for symmetric heavy tails in residual distribution and thus introduces outlier robustness in the model (9). Time-varying error magnitude has been observed for example for oral pharmacokinetic profiles with higher error magnitude during the absorption phase (2). A simple implementation of such a RUV extension is through a step function where the error magnitude changes at a certain time (after dose).

Despite the availability of these extended RUV models, they are seldom tested. A factor limiting these extended RUV models from being routinely checked is the absence of a fast, easy-to-use and accurate diagnostic tool, as implementing these extended RUV models one by one can be time-consuming and computationally intensive. This gets even more complicated in simultaneous modeling of multi-dependent variables, where exploring these extended RUV models for each of the dependent variables will result in large number of combinations to be tested. The recommended diagnostic tools for assessing RUV models are limited to plots of residuals *versus* time/model prediction (10,11). Such graphical diagnostic provide guidance only for selection among the standard RUV models: additive, proportional, or combined error model. Extended RUV models are mostly implemented based on subjective decision (e.g., log transformation of both sides), implementation facility (e.g., numerical instability of parameter estimation), or expected features (e.g., autocorrelation with rich frequent sampling schedule). To avoid this limited and case-dependent RUV modeling, here, we investigate if post-processing of common model-based diagnostics can provide additional advantages. We propose a new diagnostic tool, based on standard output such as conditional weighted residuals (CWRES), which scan these extended RUV models for their ability to improve the description of data, without re-estimating the original NLME model.

METHODS

We choose to illustrate the diagnostic procedure using CWRES which are theoretically expected to be normally distributed with mean 0 and variance 1 for a correct NLME model (12). CWRES are computed based on individual's empirical Bayes estimates regardless what estimation method is used (Eq. 1-5):

$$\bar{y}_i = f(\bar{\theta}, \bar{\eta}_i) + h(f(\bar{\theta}, \bar{\eta}_i), \bar{\varepsilon}_i) \tag{1}$$

$$\bar{y}_i \approx f(\bar{\theta}, \bar{\eta}_i) + \left. \frac{df}{d\bar{\eta}_i} \right|_{\bar{\eta}_i=\hat{\eta}_i} * (\eta_i - \hat{\eta}_i) + h(f(\bar{\theta}, \bar{\eta}_i), \bar{\varepsilon}_i) \tag{2}$$

$$E(\bar{y}_i) = f(\bar{\theta}, \hat{\eta}_i) - \left. \frac{df}{d\bar{\eta}_i} \right|_{\bar{\eta}_i=\hat{\eta}_i} * \hat{\eta}_i \tag{3}$$

$$COV(\bar{y}_i) = \left. \frac{df}{d\bar{\eta}_i} \right|_{\bar{\eta}_i=\hat{\eta}_i} * \Omega * \left. \frac{df}{d\bar{\eta}_i} \right|_{\bar{\eta}_i=\hat{\eta}_i} + \text{diag} \left(\left. \frac{dh}{d\bar{\varepsilon}_i} \right|_{\bar{\varepsilon}_i=0} * \Sigma * \left. \frac{dh}{d\bar{\varepsilon}_i} \right|_{\bar{\varepsilon}_i=0} \right) \tag{4}$$

$$CWRES = COV(\bar{y}_i)^{-1/2} * (\bar{y}_i - E(\bar{y}_i)) \tag{5}$$

where \bar{y}_i is the vector of observations for the i th individual, f denotes individual model predictions, $\bar{\theta}$ is the vector of population fixed effects, $\bar{\eta}_i$ is vector of random unexplained individual deviation from the population fixed effects, h is RUV model, $\bar{\varepsilon}_i$ is the vector of residual errors, $E(\bar{y}_i)$ is the expectation of the marginal density of the data given the model, $\hat{\eta}_i$ is vector of empirical Bayes estimates, and $COV(\bar{y}_i)$ is the covariance of the marginal density of the data given the model. Both random effects, $\bar{\eta}_i$ and $\bar{\varepsilon}_i$ are assumed to follow normal distribution with mean 0 and covariance matrix Ω and Σ , respectively. Hence, CWRES are directly linked to the objective function used in FOCE estimation method (12).

CWRES Base Model

As a substitute for assessing extended RUV models on the original data, the extended models were applied to a CWRES distribution. Thus, CWRES calculated from the original NLME model execution were treated as the dependent variable (DV) and modeled first by a linear base model to estimate CWRES distribution's mean and variance as follow:

$$\bar{y}_i = \Theta_1 + \eta_{1i} + \bar{\varepsilon}_{1i} \tag{6}$$

where \bar{y}_i is a vector of CWRES data from individual i , Θ_1 is the population mean of CWRES, η_{1i} is a univariate random variable describing the unexplained deviation of individual i from Θ_1 , and η_{1i} is assumed to follow a normal distribution of a mean 0 with variance ω^2 . $\bar{\varepsilon}_{1i}$ is a univariate random variable describing residual unexplained variability of individual i ; it is assumed to be independent identically normally distributed with a mean 0 and variance σ^2 . When the original NLME model is describing the data adequately, the expected mean of CWRES Θ_1 is 0, the expected value of ω^2 is 0, and the expected value of σ^2 is 1. This base model (Eq. 6) was then extended with the different RUV models (Supplementary material) as follow:

Autocorrelated Errors

The CWRES calculated from the original NLME model were modeled using an autoregressive error model AR1 as

$$Corr(\varepsilon_{1ij}, \varepsilon_{1ik}) = \exp(-\ln(2)/t_{1/2}) * (Time_j - Time_k) \tag{7}$$

where ε_{ij} is the residual error for individual i at time j , ε_{ik} is the residual error for individual i at time point k , and $t_{1/2}$ is half-life governing the duration of the correlation. ΔOFV_{CWRES_AR1} was calculated as the difference between base model objective function value OFV_{CWRES_Base} and AR1 error model objective function value OFV_{CWRES_AR1} .

Dynamic Transform both Sides

The individual predictions (IPREDs) calculated from the original model were used for the dTBS implementation. The dTBS model uses a Box-Cox transformation and as such requires the underlying variable to be positive; this is an issue as CWRES data includes negative observations. Thus, two models were needed to apply dTBS on CWRES data: (1) new base model instead of (Eq. 6) called CWRES dTBS base model and (2) CWRES dTBS model.

In CWRES dTBS base model, CWRES data was first exponentiated (Eq. 8) then a dTBS model (Eq. 9) with both shape parameter λ and power term ζ fixed to zero (log-normal transformation with homoscedastic variance) was fitted to the exponentiated CWRES data to calculate $OFV_{CWRES_dTBS_base}$.

In CWRES dTBS model, a dTBS model (Eq. 10) was fitted to the exponentiated CWRES data with estimating both λ and ζ , to calculate OFV_{CWRES_dTBS} :

$$\bar{y}_i = \exp(\bar{y}_i) \tag{8}$$

$$\ln(\bar{y}_i) = \ln(\widehat{IPRED}) + \bar{\varepsilon}_{1i} * (\widehat{IPRED})^\zeta \quad \text{if } \lambda = 0 \tag{9}$$

$$\frac{\bar{y}_i^\lambda - 1}{\lambda} = \frac{\widehat{IPRED}^\lambda - 1}{\lambda} + \varepsilon_{1i} * (\widehat{IPRED})^\zeta \quad \text{Otherwise} \tag{10}$$

where an estimate of $\lambda=0$ means CWRES data was normally distributed before being exponentiated, $\lambda > 0$ means CWRES data was left skewed before being exponentiated, and $\lambda < 0$ means CWRES data was right skewed before being exponentiated.

ΔOFV_{CWRES_dTBS} was calculated as the difference between dTBS base model objective function value $OFV_{CWRES_dTBS_base}$ and dTBS model objective function value OFV_{CWRES_dTBS} .

Inter-Individual Variability on RUV

The CWRES data outputted from the original model execution was modeled as

$$\bar{y}_i = \Theta_1 + \eta_{1i} + \bar{\varepsilon}_{1i} * \exp(\eta_{2i}) \tag{11}$$

where η_{2i} is random deviation of individual i from $\bar{\varepsilon}_{1i}$, thus allowing different individuals to have different RUV magnitude. ΔOFV_{CWRES_IIV} was calculated as the difference between base model objective function value OFV_{CWRES_Base} and IIV on RUV model objective function value OFV_{CWRES_IIV} .

Power Model

The individual predictions IPRED calculated on normal scale from the original model execution were used to scale individual residual error $\bar{\varepsilon}_i$ where CWRES was modeled as

$$\bar{y}_i = \Theta_1 + \eta_{1i} + \bar{\varepsilon}_{1i} * (\widehat{IPRED})^\zeta \tag{12}$$

where ζ is the power exponent determining the dependence of $\bar{\varepsilon}_i$ on model prediction. IPRED should include only positive values for successful implementation. If for instance IPRED included negative predictions because of log transformation, they were exponentiated back to normal scale. ΔOFV_{CWRES_power} was calculated as the difference between OFV_{CWRES_Base} and power error model objective function value OFV_{CWRES_power} .

t-Distribution Error Model

The Laplacian method with user-defined likelihood had to be used to apply a t-distributed residual error. The conditional likelihood L of CWRES data was defined in the control file for a Laplace base model (Eq. 13) and a t-distributed error model (Eq. 14):

$$L = (1/\sqrt{2\pi\sigma^2}) \exp\left(-\frac{IWRES^2}{2}\right) \tag{13}$$

$$L = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi\sigma^2}} \left(1 + \frac{IWRES^2}{v}\right)^{-\frac{(v+1)}{2}} \tag{14}$$

where σ^2 is the variance of the data, IWRES is the individual weighted residuals, Γ is the gamma function, and v is the degrees of freedom, which is the additionally estimated parameter when extending the RUV model to a t-distributed error. $\Delta OFV_{CWRES_t-dist}$ was calculated as the difference between Laplace base model objective function value $OFV_{CWRES_Laplace_base}$ and t-distribution error model objective function value OFV_{CWRES_t-dist} .

Time-Varying Error

Time, or time after dose, could be used as predictor of residual error magnitude (2), where cutoff time points for the change in variance of ε_i are selected based on data density to cut the data into N equal-size groups. CWRES were modeled as

$$\bar{y}_i = \Theta_1 + \eta_{1i} + \bar{\varepsilon}_{1i} \tag{15}$$

$$\text{If (Time} > X) \quad \bar{y}_i = \Theta_1 + \eta_{1i} + \bar{\varepsilon}_{2i} \tag{16}$$

where ε_{2i} is the residual error after the cutoff time point X . The number of cutoff time points is subjective; here, we used three cutoff time points to cut the data into four equal-size groups, each with a separate error magnitude. ΔOFV_{CWRES_time} was calculated as the difference between OFV_{CWRES_Base} and the time-varying error model objective function value OFV_{CWRES_time} .

Evaluations

The agreement in improvement of fit (ΔOFV) between implementing these extended RUV models on the original NLME model ($\Delta\text{OFV}_{\text{Original}}$; conventional analysis) and implementing them on CWRES data ($\Delta\text{OFV}_{\text{CWRES}}$) was evaluated in both simulated ($n=7$) and real ($n=16$) data examples as described below. For each data set and RUV model considered, there were thus two models evaluated for the original data (the original model and one with RUV model extension) and two models for the CWRES obtained from the original NLME model (the base CWRES model and the extended RUV model). In addition to CWRES, real data exercise was applied to three other diagnostics: conditional weighted residuals with interaction (CWRESI), individual weighted residuals (IWRES) and normalized prediction distribution errors (NPDE). CWRESI considers the interaction between intra- and inter-individual variability as in proportional error model. IWRES are the differences between the observations and IPRED weighted by the expected standard deviation of the residual variability (σ) only.

$$\text{IWRES} = \frac{\bar{y}_i - f(\bar{\theta}, \hat{\eta}_i)}{\sigma} \quad (17)$$

NPDEs are calculated by applying the inverse function of the normal cumulative density function to the decorrelated prediction discrepancies, and so it is normally distributed by construction without any approximations (13).

Simulations

Stochastic simulations and estimation (SSEs) were performed to investigate the ability of CWRES modeling to identify correctly the true extended RUV model in different scenarios, as well as investigating type I error rates of the likelihood ratio test employed, with null hypothesis being the absence of RUV misspecification. The base simulation model was a one-compartment disposition model with first-order absorption, linear elimination, and proportional RUV model. Parameters assigned with inter-individual variability were absorption rate constant, clearance, and volume of distribution, with a correlation between the last two parameters. This base model was extended with the investigated RUV models to produce six additional models. The base model was used to simulate 200 datasets, while each of the extended RUV models was used to simulate 100 datasets; each dataset was estimated with the seven models to calculate mean $\Delta\text{OFV}_{\text{Original}}$ for each extended RUV model. CWRES calculated from the estimated base models and modeled as described previously to calculate mean $\Delta\text{OFV}_{\text{CWRES}}$ for each extended RUV model (Fig. 1). The values of parameters governing each of the extended RUV models when simulating were chosen to produce a $\Delta\text{OFV}_{\text{Original}}$ of up to 100 between the base model and the true extended RUV model.

Real Data Examples

Models varied in complexity, amount of available data, and residual error components (Table I). Four examples

modeled more than one dependent variable simultaneously. Data was log-transformed in nine examples. Two examples already included IIV on RUV model and this was then considered the base RUV model. All examples were developed using the FOCE method, and the interaction option was added in relevant RUV extended models.

Software

NONMEM version 7.3 (ICON Development Solutions, Hanover, MD, USA) (29) was used for the analysis with help of PsN (30), and graphs were generated in R (31).

RESULTS

Simulated Data Examples

CWRES modeling identified the same (correct) RUV model as the conventional analysis as shown in Fig. 2. The highest $\Delta\text{OFV}_{\text{CWRES}}$ signaled the correct RUV extension model used in the simulation for AR1, dTBS, IIV, t-distribution, and time-varying RUV error models, e.g., $\Delta\text{OFV}_{\text{CWRES_AR1}}$ was the highest drop between all CWRES models (red bars) when simulating with AR1 error model. Note that the Power and dTBS models are nested and the results reflected this. Since when simulating with power error model only correction for residual scedasticity was needed, both $\Delta\text{OFV}_{\text{CWRES_power}}$ and $\Delta\text{OFV}_{\text{CWRES_dTBS}}$ were of equal magnitude, -20 and -22 , respectively, so it is clear that dTBS model is not providing any additional advantages over power error model which should be selected as the correct RUV model extension. Type I error rates with CWRES were 1.5, 1, 2, 5.5, 3, and 8.5% for AR1, dTBS, IIV, power, t-distribution, and time-varying RUV models, respectively, taking into account the number of estimated parameters. When simulating with the base model, none of the investigated extension showed an improvement either by CWRES modeling or conventional analysis.

The improvement in goodness-of-fit ΔOFV when applying an extension was in general of similar magnitude whether based on $\Delta\text{OFV}_{\text{CWRES}}$ or $\Delta\text{OFV}_{\text{Original}}$. However, if the method worked perfectly, $\Delta\text{OFV}_{\text{CWRES}}$ would equal exactly $\Delta\text{OFV}_{\text{Original}}$ which is not the case especially for AR1 simulation scenario, where $\Delta\text{OFV}_{\text{CWRES}}$ were -34 , while $\Delta\text{OFV}_{\text{Original}}$ was -100 . The parameters governing these extended RUV models showed a good concordance between their estimates from CWRES and conventional analysis, with a correlation coefficient of 0.93 across all RUV extensions except for dTBS as they are on different scale.

Real Data Examples

At least one of the investigated RUV extensions resulted into a significant $\Delta\text{OFV}_{\text{Original}}$ in all examples except for Daunorubicin and Digoxin PD models. When significant, the improvement was substantial with $\Delta\text{OFV}_{\text{Original}}$ ranging from -2019 for Gastric emptying with autocorrelated error to -4 for r-hFSH PK with IIV on RUV, with a median $\Delta\text{OFV}_{\text{Original}}$ of -71 across models with significant improvement. Similar to the conventional analysis, a significant $\Delta\text{OFV}_{\text{CWRES}}$ were found in all examples except for Daunorubicin and Digoxin

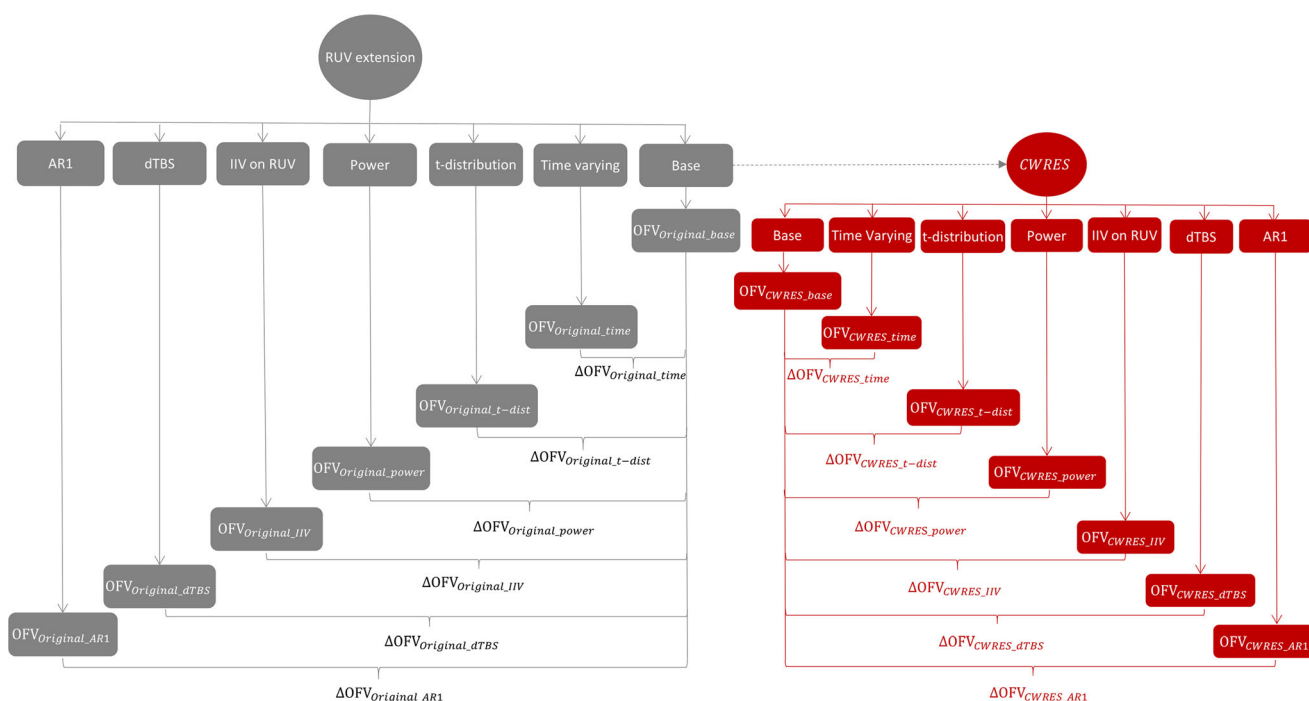


Fig. 1. Schematic presentation of simulation setup. The base model was used to simulate 200 data sets, while each of the extended RUV models was used to simulate 100 datasets, and each dataset was estimated with the seven models to calculate mean $\Delta OFV_{Original}$ for each extended RUV model. CWRES outputted from the base model fitting to the simulated data in each of the different scenarios were treated as dependent variables and modeled with same RUV extensions to calculate mean ΔOFV_{CWRES} for each extended RUV model. Note that different base models were needed for dTBS and t-distribution error models

PD models. The real data examples further supported the good agreement between the true misspecification in error model and what was identified by modeling of CWRES as CWRES modeling identified the most important RUV extension similar to the conventional analysis (Fig. 3). Exceptions were the two models Ethambutol PK and Disufenton sodium,

where the order of the 1st (t-distribution) and 2nd most important extensions (IIV on RUV) were reversed. CWRES modeling identified the same RUV extensions to be significant improvements as conventional analysis except for t-distributed error model with Asenapine PD. The $\Delta OFV_{Original}$ and ΔOFV_{CWRES} displayed a correlation coefficient of 0.88 across

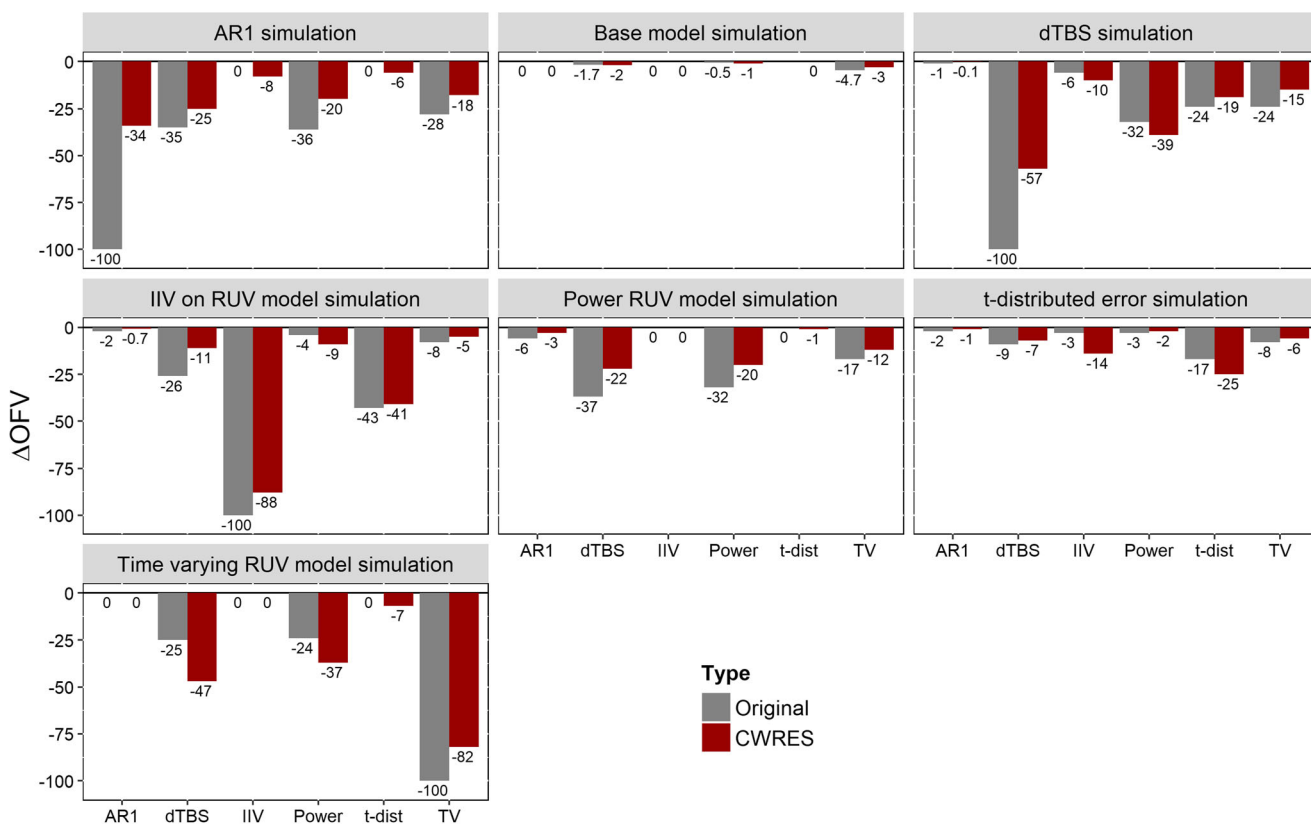
Table I. Summary of Real Data Examples Used for Investigation

Model	Data type	RUV model	Transformation	No. of observations	No. of subjects	No. of THETAs	No. of OMEGAs including covariances	No. of SIGMAS
Asenapine ^a (14)	PD	Additive with IIV	–	7728	1328	16	5	1
Clomethiazole (15)	PK	Additive	Log	2177	772	10	5	1
Daunorubicin ^a (16)	PD	Additive	Log	112	41	7	3	1
Digoxin ^a (17)	PD	Additive	–	787	225	6	2	1
Digoxin ^{ab} (18)	PK/PD	PD: proportional PK: additive	–	941	225	6	3	1
Disufenton sodium ^a (19)	PK	Additive	Log	1196	175	7	3	1
Ethambutol ^a (20)	PK	Combined	Log	1869	189	8	3	1
Gastric emptying ^b (21)	PK/PD	Additive	Log	3385	16	37	17	4
HbA1C ^{ab} (22)	PD	Additive with IIV	Log	8698	412	21	9	1
IGI ^{ab} (23)	PD	Additive	Log	6382	72	26	15	1
Lopinavir ^a (24)	PK	Proportional	–	315	30	9	6	1
Miltefosine ^{a2} (25)	PK	Proportional	–	350	31	7	4	1
Moxonidine ^a (9)	PK	Additive	Log	1021	74	5	6	1
Paclitaxel ^{ac} (26)	PD	Combined	–	530	46	6	3	1
Pefloxacin ^a (27)	PK	Additive	Log	337	74	4	6	1
r-Hfsh (28)	PK	Additive	–	314	60	7	2	1

^a SIGMAS were fixed to 1 and modeled as THETAs (standard deviation)

^b More than one dependent variable

^c Additive component of RUV model was fixed



RUV model extensions

Fig. 2. Plot of mean ΔOFV vs extended RUV models show the agreement between mean $\Delta\text{OFV}_{\text{Original}}$ (gray) and mean $\Delta\text{OFV}_{\text{CWRES}}$ (red) among all seven scenarios of simulations. Abbreviations: AR1 autocorrelated errors, dTBS dynamic transform both sides, IIV inter-individual variability on RUV, Power power error model, t-dist t-distribution error model, TV time-varying error model

all models and a median ratio $\frac{\Delta\text{OFV}_{\text{CWRES}}}{\Delta\text{OFV}_{\text{Original}}}$ of 0.77 among models with significant improvement. It is not surprising that this ratio is below one as when extending the original model, there is always the potential for larger improvement in fit than CWRES as many parameters are re-estimated under the new RUV model.

The evaluation of the other diagnostics CWRES, IWRES, and NPDE with real data examples is shown in Fig. 4, where CWRES outperformed other diagnostics in predicting $\Delta\text{OFV}_{\text{Original}}$. The root mean-squared errors (RMSE) of $\Delta\text{OFV}_{\text{CWRES}}$ were lower than RMSE of other $\Delta\text{OFV}_{\text{diagnostics}}$ and so they are reported as relative to RMSE of $\Delta\text{OFV}_{\text{CWRES}}$ as shown in Table II, where RMSE measures the differences between $\Delta\text{OFV}_{\text{diagnostics}}$ and the actually observed $\Delta\text{OFV}_{\text{Original}}$.

$$\text{RMSE} = \sqrt{E\left((\Delta\text{OFV}_{\text{diagnostics}} - \Delta\text{OFV}_{\text{Original}})^2\right)} \quad (18)$$

DISCUSSION

A new diagnostic tool for post-processing of CWRES was successfully developed and evaluated. CWRES modeling evaluate extended RUV models and assess in a robust and extremely fast way whether extensions are needed to implement. The

method accurately identified the correct type of RUV model misspecifications as well as the expected improvement in fit in terms of ΔOFV and the expected parameters of the extended RUV model, similar to conventional analysis. This method typically does not suffer from local minima problems or other estimation-related issues because CWRES models are simple and quick to run with known expected distribution as shown by the agreement between $\Delta\text{OFV}_{\text{Original}}$ and $\Delta\text{OFV}_{\text{CWRES}}$. In no case was modifications of initial estimates needed for the CWRES models applied here.

Simulation results elucidated the ability of different extensions to describe each other and to produce similar improvement in fit. The t-distributed error model allows the incorporation of outlier robustness into the model while IIV on RUV allows individuals to have different RUV magnitude; thus, both extensions can describe outliers and produce similar model fits. This behavior was shown when simulating with IIV on RUV; t-distributed error model was the 2nd most important RUV extension. This behavior also might explain the inflated $\Delta\text{OFV}_{\text{CWRES_IIV}}$ when t-distributed error model was the most important improvement in both simulated and real data examples, and the reversed order of most important RUV extensions for Ethambutol PK and Disfenton sodium.

Both dTBS and Power error model can account for scedasticity; however, only dTBS can correct for both scedasticity and skewness; thus, on simulating from power error model, both extensions had similar $\Delta\text{OFV}_{\text{CWRES}}$ but

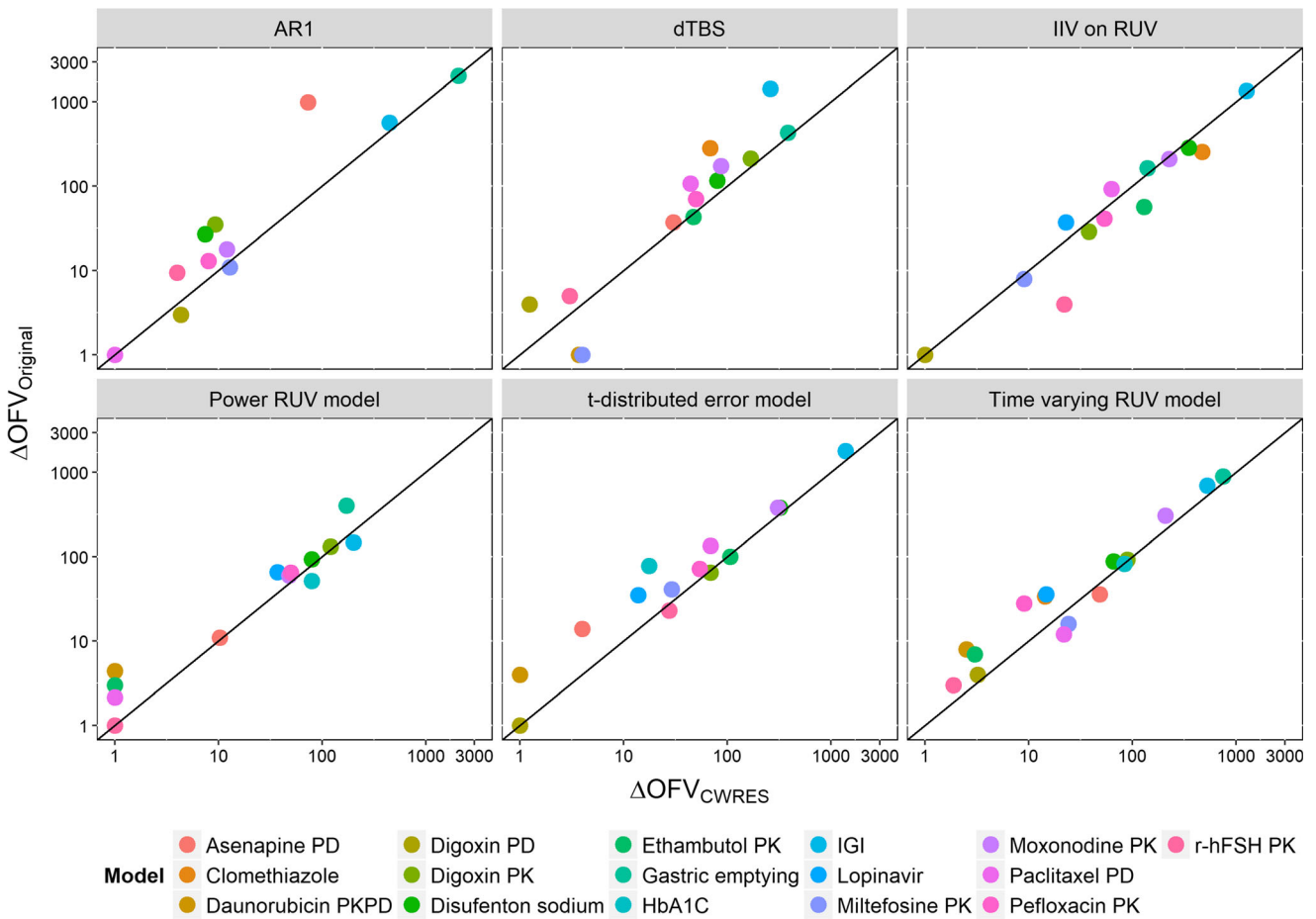


Fig. 3. Plot of absolute $\Delta OFV_{Original}$ vs absolute ΔOFV_{CWRES} among all real data examples for the six extended RUV models; all points around the line of unity showing the agreement between conventional analysis and CWRES modeling

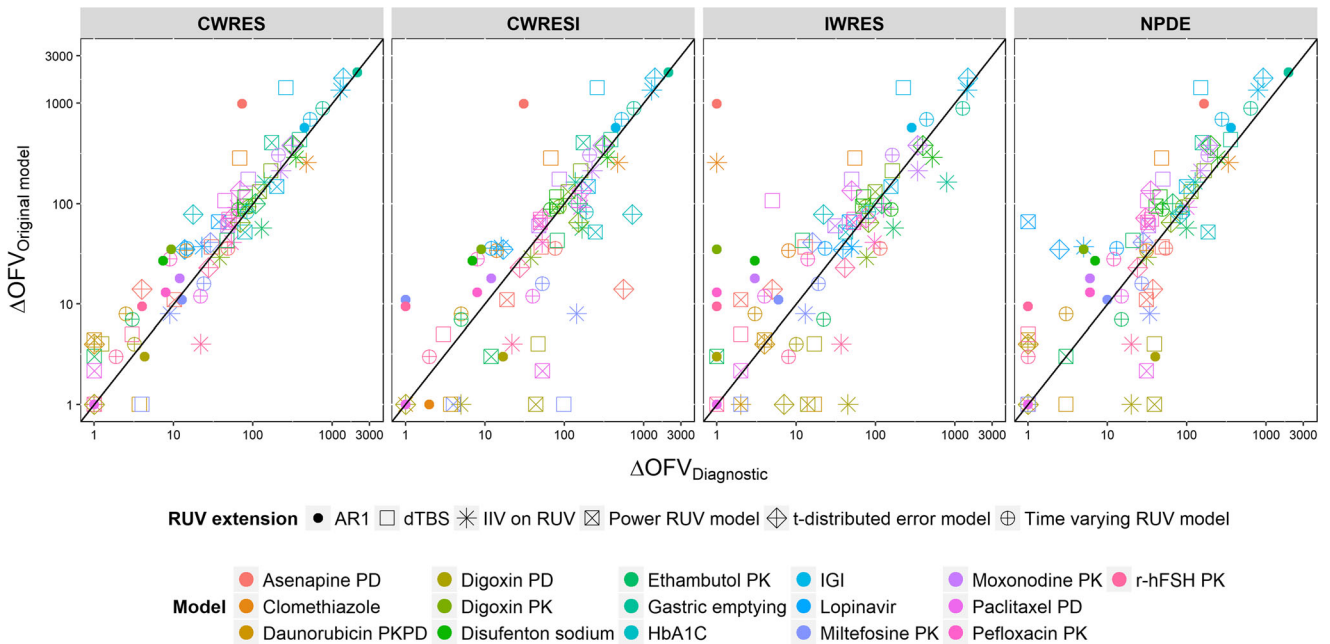


Fig. 4. Plot of absolute $\Delta OFV_{Original}$ vs absolute $\Delta OFV_{Diagnostic}$ for CWRES, CWRESI, IWRES, and NPDE among all real data examples for the six extended RUV models

Table II. The RMSE of $\Delta\text{OFV}_{\text{CWRES}}$ for Different $\Delta\text{OFV}_{\text{Original}}$ Ranges and the Relative RMSE of Other $\Delta\text{OFV}_{\text{diagnostics}}$ to RMSE of $\Delta\text{OFV}_{\text{CWRES}}$

$\Delta\text{OFV}_{\text{Original}}$ range	Number of models	RMSE of $\Delta\text{OFV}_{\text{CWRES}}$	Relative RMSE of $\Delta\text{OFV}_{\text{CWRES}}$	Relative RMSE of $\Delta\text{OFV}_{\text{IWRES}}$	Relative RMSE of $\Delta\text{OFV}_{\text{NPDE}}$
0–10	26	4	8.5	3	3.7
10–100	33	22	6.8	1.7	1.3
100–1000	21	228	1.03	3	1
> 1000	4	621	1	2.8	1.3

not when simulating from dTBS; a profound real data example of this was Pefloxacin where $\Delta\text{OFV}_{\text{Original_power}}(\Delta\text{OFV}_{\text{CWRES_power}})$ was -65 (-50) and $\Delta\text{OFV}_{\text{Original_dTBS}}(\Delta\text{OFV}_{\text{CWRES_dTBS}})$ was -70 (-50) as correction for scedasticity was only in need. On the other hand, a correction for both scedasticity and skewness was needed for Moxonidine PK where $\Delta\text{OFV}_{\text{Original_power}}(\Delta\text{OFV}_{\text{CWRES_power}})$ was -60 (-48) and $\Delta\text{OFV}_{\text{Original_dTBS}}(\Delta\text{OFV}_{\text{CWRES_dTBS}})$ was -243 (-87).

As mentioned in our results, $\Delta\text{OFV}_{\text{Original}}$ is expected to be larger than $\Delta\text{OFV}_{\text{CWRES}}$, as it is more flexible with more parameters to estimate; this difference between the ΔOFVs was most pronounced with the dTBS approach as shown by most results being in the top left triangle in Fig. 3 when correction for scedasticity and skewness resulted in different non-residual error parameter estimates, for example in Disufenton sodium PK: $\Delta\text{OFV}_{\text{Original_dTBS}}(\Delta\text{OFV}_{\text{CWRES_dTBS}})$ was -116 (-80) as implementing dTBS led to a relative change of 126, 12, 30, 10, 6, 16, and 15% in estimates of THETAs' from 1 to 7 respectively.

One interesting aspect is in case of multiple dependent variables, using CWRES modeling identified accurately which dependent variable needed further improvement which decrease the risk of model over parameterization. For example, the Gastric emptying PK model had four dependent variables; by conventional analysis, only the total $\Delta\text{OFV}_{\text{Original_dTBS}}$ (-433) is obtained but this does not provide information about for which dependent variable an extended RUV model is dominating the improvement in fit. On the other hand, by CWRES modeling, four separate $\Delta\text{OFV}_{\text{CWRES_dTBS}}$, one for each dependent variable, are obtained and only the 2nd ($\Delta\text{OFV}_{\text{CWRES_dTBS}} = -207$) and 4th ($\Delta\text{OFV}_{\text{CWRES_dTBS}} = -139$) dependent variables could be identified as important to transform in the original model.

Other common diagnostics CWRESI, IWRES, and NPDE are available for evaluation of model goodness of fit. Similar to CWRES, these diagnostics should be normally distributed when the model adequately describe the data, except for CWRESI because of the interaction between $\bar{\eta}_i$ and $\bar{\varepsilon}_i$. It had been concluded previously that IWRES perform poorly with increasing model non-linearity, leading to biased parameter estimates and misguided model development (12), and that CWRES and NPDE give the best diagnostics in different situations even when there was interaction in the model (32); our results support these conclusions with a further favor of CWRES over NPDE as shown in Fig. 4. Noting that NPDE are sensitive to the number of samples and results shown here were with setting ESAMPLE option to 10000, as lower samples resulted in poor predictions of $\Delta\text{OFV}_{\text{Original}}$.

In conclusion, the principle of model-based diagnostics post-processing for automated model building had been

demonstrated and successfully applied with CWRES modeling, which is a valuable diagnostic tool for RUV model identification during model development/evaluation process, as it provides guidance for the nature and magnitude of potential RUV model misspecification/improvements.

How to Proceed in Practice

CWRES modeling can be easily implemented in analysis software; it is already implemented in PsN (available as “resmod” from version 4.7.0). The procedure requires the original NLME model file and a table with ID, TIME, CWRES, and IPRED data items from the original NLME model execution. CWRES data item can be replaced with other residuals, and DVID data item is needed in case of multiple dependent variables. TIME data item can be replaced by time after dose TAD if available. IPRED data item should be positive on normal scale for successful testing of dTBS and power error models. Options are available to account for multiple occasions and to manually set the number of time cut off points. The extended RUV models are automatically created and run; the output contains $\Delta\text{OFV}_{\text{CWRES}}$ for each extended RUV model together with the parameters of interest for each RUV extension. Since resmod is a fast and easy to perform procedure with no limitations for continuous data, it is a good idea to always consider applying resmod when selecting between-error models for model development. Also, all parameters of the original NLME model should be re-estimated after implementing any of the extended RUV models. As we showed, these extended RUV models address different problems of the residual variability. In our preliminary results, a combination of these extended RUV models sometimes showed a better description of data; however, that was not further investigated. A previously discussed example of this was a combination of dTBS approach with t-distributed error model to address skewed residuals with outliers (7).

ACKNOWLEDGMENTS

The authors would like to thank colleagues in Pharmacometrics research group at Uppsala University for providing their models, data, and associated investigations.

AUTHOR CONTRIBUTIONS

M.M.A.I., R.N., M.C.K., and M.O.K. wrote the manuscript. M.M.A.I., R.N., M.C.K., and M.O.K. designed the research. M.M.A.I. performed the research. M.M.A.I., R.N., M.C.K., and M.O.K. analyzed the data.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

REFERENCES

- Gobburu JV. Pharmacometrics 2020. *J Clin Pharmacol.* 2010;50:151S–7S.
- Karlsson MO, Beal SL, Sheiner LB. Three new residual error models for population PK/PD analyses. *J Pharmacokinet Biopharm.* 1995;23(6):651–72. <https://doi.org/10.1007/BF02353466>.
- Silber HE, Kjellsson MC, Karlsson MO. The impact of misspecification of residual error or correlation structure on the type I error rate for covariate inclusion. *J Pharmacokinet Pharmacodyn.* 2009;36(1):81–99. <https://doi.org/10.1007/s10928-009-9112-1>.
- Long JS, Ervin LH. Using heteroscedasticity consistent standard errors in the linear regression model. *J Am Stat.* 2000;54(3):217–24. <https://doi.org/10.2307/2685594>.
- Chi EM, Reinsel GC. Models for longitudinal data with random effects and AR(1) errors. *J Am Stat Assoc.* 1989;84:452–9.
- Davidian M, Giltinan DM. Some general estimation methods for nonlinear mixed effects models. *J Biopharm Stat.* 1993;3:23–55.
- Dosne A, Bergstrand M, Karlsson MO. A strategy for residual error modeling incorporating scedasticity of variance and distribution shape. *J Pharmacokinet Pharmacodyn.* 2015;43(2):137–51. <https://doi.org/10.1007/s10928-015-9460-y>.
- Box GEP, Cox DR. An analysis of transformations. *J R Stat Soc B.* 1964;26(2):211–52.
- 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(2):207–46.
- Mould DR, Upton RN. Basic concepts in population modeling, simulation, and model-based drug development. Part 2: introduction to pharmacokinetic modeling methods. *CPT Pharmacometrics Syst Pharmacol.* 2013;2:e38.
- Keizer RJ, Karlsson MO, Hooker AC. Modeling and simulation workbench for NONMEM: tutorial on Pirana, PsN, and Xpose. *CPT Pharmacometrics Syst Pharmacol.* 2013;2(6):e50.
- Hooker AC, Staats CE, Karlsson MO. Conditional weighted residuals (CWRES): a model diagnostic for the FOCE method. *Pharm Res.* 2007;24(12):2187–97.
- Comets E, Brendel K, Mentré F. Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R. *Comput Methods Prog Biomed.* 2008;90(2):154–66.
- Friberg L, Greef RD, Kerbusch T, Karlsson MO. Modeling and simulation of the time course of asenapine exposure response and dropout patterns in acute schizophrenia. *Clin Pharmacol Ther.* 2009;86(1):84–91. <https://doi.org/10.1038/clpt.2009.44>.
- Zingmark P, Ekblom M, Odegren T, Ashwood T, Lyden P, Karlsson MO, *et al.* Population pharmacokinetics of clomethiazole and its effect on the natural course of sedation in acute stroke patients. *Br J Clin Pharmacol.* 2003;56(2):173–83. <https://doi.org/10.1046/j.0306-5251.2003.01850.x>.
- Bogason A, Quartino AL, Lafolie P, Masquelier M, Karlsson MO, Paul C, *et al.* Inverse relationship between leukaemic cell burden and plasma concentrations of daunorubicin in patients with acute myeloid leukaemia. *Br J Clin Pharmacol.* 2011;71(4):514–21. <https://doi.org/10.1111/j.1365-2125.2010.03894.x>.
- Henning S, Friberg L, Karlsson MO. Characterizing time to conversion to sinus rhythm under digoxin and placebo in acute atrial fibrillation. *PAGE* 18. Abstr 1504. 2009. www.page-meeting.org/?abstract=1504.
- 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(11):747–55. <https://doi.org/10.1007/s00228-002-0553-3>.
- Jonsson S, Cheng Y, Edenius C, Lees KR, Odegren T, Karlsson MO. Population pharmacokinetic modelling and estimation of dosing strategy for NXY-059, a Nitrone being developed for stroke. *Clin Pharmacokinet.* 2005;44(8):863–78. <https://doi.org/10.2165/00003088-200544080-00007>.
- Jonsson S, Davidse A, Wilkins J, Walt JV, Simonsson US, Karlsson MO, *et al.* Population pharmacokinetics of ethambutol in south African tuberculosis patients. *Antimicrob Agents Chemother.* 2011;55(9):4230–7. <https://doi.org/10.1128/aac.00274-11>.
- Ålskär O, Bagger JI, Røge RM, Knop FK, Karlsson MO, Vilsbøll T, *et al.* Semimechanistic model describing gastric emptying and glucose absorption in healthy subjects and patients with type 2 diabetes. *J Clin Pharmacol.* 2015;56(3):340–8. <https://doi.org/10.1002/jcph.602>.
- Hamrén B, Björk E, Sunzel M, Karlsson MO. 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. <https://doi.org/10.1038/clpt.2008.2>.
- Silber HE, Jauslin PM, Frey N, Gieschke R, Simonsson US, Karlsson MO. An integrated model for glucose and insulin regulation in healthy volunteers and type 2 diabetic patients following intravenous glucose provocations. *J Clin Pharmacol.* 2007;47(9):1159–71. <https://doi.org/10.1177/0091270007304457>.
- El sherbiny D, Ren Y, McIlleron H, Maartens G, Simonsson US. Population pharmacokinetics of lopinavir in combination with rifampicin-based antitubercular treatment in HIV-infected south African children. *Eur J Clin Pharmacol.* 2010;66(10):1017–23. <https://doi.org/10.1007/s00228-010-0847-9>.
- Dorlo TP, Van Thiel PPAM, Huitema AD, Keizer RJ, Vries HJ, Beijnen JH, *et al.* Pharmacokinetics of miltefosine in old world cutaneous leishmaniasis patients. *Antimicrob Agents Chemother.* 2008;52(8):2855–60. <https://doi.org/10.1128/aac.00014-08>.
- Friberg LE, Henningson A, Maas H, Nguyen L, Karlsson MO. Model of chemotherapy-induced myelosuppression with parameter consistency across drugs. *J Clin Oncol.* 2002;20(24):4713–21. <https://doi.org/10.1200/jco.2002.02.140>.
- Wahlby U, Thomson AH, Milligan PA, Karlsson MO. Models for time-varying covariates in population pharmacokinetic-pharmacodynamic analysis. *Br J Clin Pharmacol.* 2004;58(4):367–77. <https://doi.org/10.1111/j.1365-2125.2004.02170.x>.
- Karlsson MO, Wade JR, Loumaye E, Munafo A. The population pharmacokinetics of recombinant- and urinary-human follicle stimulating hormone in women. *Br J Clin Pharmacol.* 2002;45(1):13–20. <https://doi.org/10.1046/j.1365-2125.1998.00644.x>.
- Beal S, Sheiner LB, Boeckmann A, Bauer RJ. NONMEM User's Guides. Icon Development Solutions, Ellicott City; 1989–2010.
- Lindbom L, Pihlgren P, Jonsson EN. PsN-toolkit—a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Comput Methods Prog Biomed.* 2005;79(3):241–57. <https://doi.org/10.1016/j.cmpb.2005.04.005>.
- Team RC. R: a language and environment for statistical computing. Vienna, Austria; 2014. Available from: <http://www.R-project.org>.
- Nyberg J, Bauer RJ, Hooker AC. Investigations of the weighted residuals in NONMEM7 *PAGE* 19. Abstr 1883. 2010. www.page-meeting.org/?abstract=1883.