Research Article

Performance of Methods for Handling Missing Categorical Covariate Data in Population Pharmacokinetic Analyses

Ron J. Keizer,[1,2,4,5] **Anthe S. Zandvliet,**[1,2] **Jos H. Beijnen,**[1,2,3] **Jan H. M. Schellens,**[2,3] **and Alwin D. R. Huitema**[1,2]

Abstract. In population pharmacokinetic analyses, missing categorical data are often encountered. We evaluated several methods of performing covariate analyses with partially missing categorical covariate data. Missing data methods consisted of discarding data (DROP), additional effect parameter for the group with missing data (EXTRA), and mixture methods in which the mixing probability was fixed to the observed fraction of categories (MIX_{obs}), based on the likelihood of the concentration data (MIX_{conc}), or combined likelihood of observed covariate data and concentration data (MIX_{joint}). Simulations were implemented to study bias and imprecision of the methods in datasets with equal-sized and unbalanced category ratios for a binary covariate as well as datasets with non-random missingness (MNAR). Additionally, the performance and feasibility of implementation was assessed in two real datasets. At either low (10%) or high (50%) levels of missingness, all methods performed similarly well. Performance was similar for situations with unbalanced datasets (3:1 covariate distribution) and balanced datasets. In the MNAR scenario, the MIX methods showed a higher bias in the estimation of *CL* and covariate effect than EXTRA. All methods could be applied to real datasets, except DROP. All methods perform similarly at the studied levels of missingness, but the DROP and EXTRA methods provided less bias than the mixture methods in the case of MNAR. However, EXTRA was associated with inflated type I error rates of covariate selection, while DROP handled data inefficiently.

KEY WORDS: categorical data; covariates; missing data; pharmacokinetics; pharmacometrics.

INTRODUCTION

In population pharmacokinetic (PopPK) and or pharmacodynamic modeling analyses, the researcher is often confronted with missing data due to left censoring (*e.g.*, data below the lower limit of quantification), data not being measured or recorded (*e.g.*, genotype data), or data missing for other reasons. Handling of missing covariate data in PopPK analyses in an appropriate way, *i.e.*, one that does not induce bias or imprecision in model parameter estimation is important, and several methods are available to the modeler. However, most established methods focus on missing *continuous* data (1,2). Briefly, these consist of

[1] Department of Pharmacy & Pharmacology, The Netherlands Cancer Institute/Slotervaart Hospital, Amsterdam, the Netherlands.
[2] Division of Clinical Pharmacology, Department of Medical Oncology, The Netherlands Cancer Institute, Amsterdam, the Netherlands.
[3] Division of Drug Toxicology, Section of Biomedical Analysis, Department of Pharmaceutical Sciences, Faculty of Science, Utrecht University, Utrecht, the Netherlands.
[4] Pharmacometrics Research Group, Department of Pharmaceutical Biosciences, Uppsala University, Box 59175124, Uppsala, Sweden.
[5] To whom correspondence should be addressed. (e-mail:ron.keizer@farmbio.uu.se)

case deletion, single and multiple imputation methods, and maximum likelihood methods.

In several PopPK analyses performed by our group, missing data of the categorical type were encountered, which were handled in various ways (3–7). In this article, we evaluated the performance of several methods for handling missing categorical data in nonlinear mixed-effects analyses. This was done both by simulations and by retrospective analysis of some of the cited studies. We have investigated the impact of these methods on the estimation of parameters, type I error of covariate inclusion (*i.e.*, the probability of falsely demonstrating a significant covariate relationship), and the feasibility of implementing the methods.

The nature of missing data can be separated into several classes with somewhat confusing terms: data can be missing at random (MAR), missing completely at random (MCAR), or missing not at random (MNAR) (1,8). The situation of MCAR occurs when the missingness does not depend on any (observed or unobserved) data or model parameters. The situation of MNAR, in which the missingness pattern is (partly) dependent on the missing data itself, occurs, *e.g.*, in the situation of data below the lower limit of quantification. The MAR situation occurs when the observed missingness pattern depends on the observed data, but not on the missing data. Both situations of random missingness and non-random missingness were investigated in this study.

For application in PopPK analyses performed using nonlinear mixed-effects modeling with NONMEM, we

identified five methods for handling missing categorical data, which are presented in Table I: discarding subjects with missing data (DROP), estimation of an additional covariate effect for the missing group (EXTRA), and several methods using mixture models (MIX). A mixture model is a model that explicitly assumes that the population consists of two or more subpopulations. Each of these subpopulations can be defined as having its own model with associated population parameter values (9). With two subpopulations, as in the analyses presented here, some fraction p of the population has one set of typical values, while the remaining fraction $(1-p)$ has another set. The fraction(s) of subjects in each category can either be set to the fraction observed in the non-missing subjects (MIX$_{obs}$), estimated based on the maximization of the likelihood of PK concentration data alone (MIX$_{conc}$) or estimated based on the likelihood of both PK concentration data and observed categories for non-missing subjects (MIX$_{joint}$).

Using simulations, we first evaluated the performance of these methods in terms of bias and imprecision in the estimation of PK parameters and covariate effect(s) for a representative PK example with a non-time-varying covariate, as well as a scenario for a time-varying covariate, and a sparse sampling design. We also included situations with non-ignorable missingness, *i.e.*, the percentage of missing depended on the value of the covariate. In statistical literature, several methods have been presented to cope with non-ignorable missing data, most of them using maximum likelihood approaches to handle these data (10,11). Subsequently, we have evaluated the missing data methods using two real datasets, mainly to evaluate the feasibility of implementation.

METHODS

Missing Data Mechanism

The "missing data mechanism," as described in the "INTRODUCTION," can be more formally described by (12)

$$\pi_i = Pr(R_i = 1 | y_1, x_i, z_i) \tag{1}$$

Table I. Methods for Dealing with Missing Categorical Data

Method	Description
DROP	Case deletion method. Data from the patient of which the covariate data are missing and are not used in the estimation of the model parameters
EXTRA	Estimate separate parameter for the effect of missing covariate
MIX$_{obs}$	Use of a mixture model to appoint the category for each subject. p_{mix} fixed to the observed population fraction
MIX$_{conc}$	Use of a mixture model. p_{mix} estimated, with likelihood defined by the observed PK data alone
MIX$_{joint}$	Use of a mixture model. p_{mix} estimated, with likelihood defined by both observed categories for each subject, and PK data

p_{mix} mixing probability

with π_i describing the probability of missingness in data for covariate z_i (missing indicated by R_i), y_i a vector describing the observed data, and x_i a vector of other covariates. In the situation of MCAR, π_i does not depend on any of the vectors y, x, or z. In the situation of MAR, π_i can depend on y_i and/or x_i, but not on z_i. If π_i depends on z_i, the missing data mechanism is said to be MNAR.

Simulation Study

A virtual compound was studied which was administered orally and exhibited first-order absorption (ka=0.5 h^{-1}, F=1) to the central compartment (V=50 L) and first-order distribution to a peripheral compartment (Q=5 Lh^{-1} V_{per}= 100 L). Elimination was linear from the central compartment (CL=10 Lh^{-1}). CL, V, and *ka* were assigned an inter-individual variability of 25% using a log-normal distribution, with 50% correlation between individual CL and V. The proportional residual error was set at 20%, and no additive error was included. Sample points were at t=0.5, 1, 1.5, 2, 4, 6, 8, 12, 16, and 24 h after a 100-mg oral dose.

The simulated covariate effect was incorporated as an effect of the binary covariate sex (SEX) on clearance, with varying fractions of male (M, SEX=0) *versus* female subjects (F, SEX=1), and two levels of missingness (10% and 50%). The covariate effect of sex was purely hypothetical, *i.e.*, there was no physiological rationale behind the choice of covariate name. The design of the various datasets is shown schematically in Fig. 1, with black fractions indicating males and white indicating females. Grayed-out areas represent censored parts of the dataset. To be able to assess type I error rates, another binary covariate, DUM, having no relation to SEX and no effect on any PK parameter was simulated in a similar ratio as the SEX covariate. The DUM variable was set to 0 and 1 for the appropriate fractions of patients (see below). Datasets were simulated containing 100 patients each. For each sub-analysis, 100 datasets were simulated and re-estimated. With respect to the sex ratios, several dataset designs were investigated.

(a) "Balanced"—in which the ratio M/F was 1:1
(b) "M+"—in which the M/F ratio was 3:1
(c) "F+"—in which the M/F ratio was 1:3
(d) "NonRand"—in which the overall ratio was 1:1 and 50% of the dataset was missing, but the probability of being missing (p_{mix}) was three times higher for males than for females. Thus, in the non-missing part of the dataset, the M/F ratio was 1:3, but in the missing part of the dataset the ratio was 3:1.
(e) "Sparse sampling"—which used a covariate scenario design similar to "Balanced," however with only four PK samples taken the first 24 h and additional trough samples taken after 1 and 2 weeks of treatment (100 mg, q.d.). The simulations were performed for both a scenario of MCAR and a scenario of MNAR. In the MNAR scenario, the missingness depended on the value of the covariate (when COM=1, the probability of missingness was three times higher). Baseline probabilities (for COM=0) of being missing were 10%, 20%, and 30% (so 30%, 60%, and 90% for occasions where COM=1).
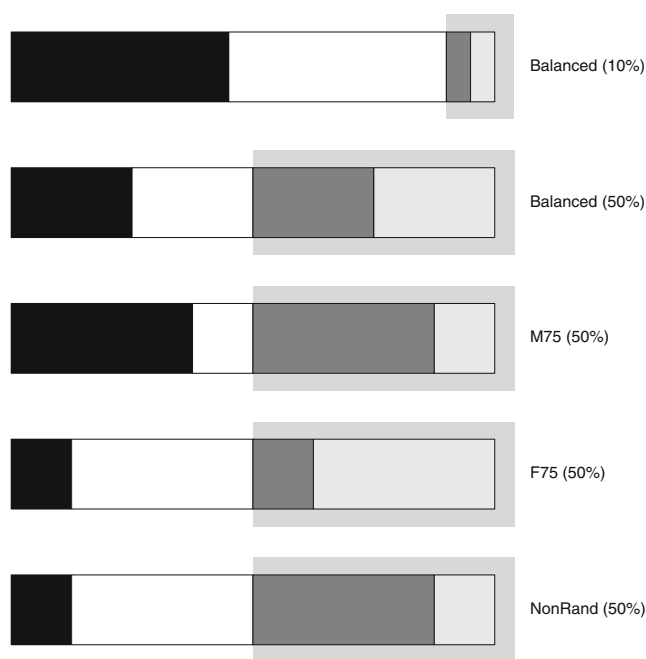
Fig. 1. Dataset designs used in the simulation analysis. *Black* indicates the fraction of male subjects; *white* indicates females. *Left part* shows the observed covariate data and *right part* (*grayed out*) shows the missing part. Missingness percentage in observed population given between *brackets*. *NonRand* probability of missing is 25% for females and 75% for males

These scenarios, with a moderate covariate effect magnitude on CL (θ_{male}=1.25), a moderate magnitude of between-subject variability (25%), and a moderate magnitude of residual error (20%), were considered "moderately informative" to assess the impact of SEX on CL. Additionally, a collection of "highly informative" datasets was simulated as well in which the covariate effect magnitude on CL was increased (θ_{male}=2 instead of 1.25), or the magnitude of between-subject variability was lower (ω= 10% instead of 25%), or the magnitude of residual error was lower ($\varepsilon_{\text{prop}}$=10% instead of 20%), as well as one dataset with higher IIV (ω=50% instead of 25% for both CL and V). The analysis was also repeated based on datasets simulated with no effect of the covariate on any PK parameter ("non-informative"), with similar PK parameters to the "moderately informative" scenario, except for the covariate effect. Besides the scenarios above, a scenario with a *time-varying* covariate was studied. The same virtual drug was used, but now the covariate ("COM", use of co-medication that increased CL by 1.25 times) was allowed to change randomly between occasions (0/1).

Models and Estimation

In general, the simulated data were reanalyzed using the "true" basic structural PK model using all missing data methods. For the time-varying scenario, re-estimations were only performed for DROP and EXTRA since MIX approaches (as currently implemented in NONMEM) cannot be applied to analyze time-varying covariate data. For the sparse sampling scenario, data were analyzed using both the true model and the simpler one-compartment PK model (using DROP and EXTRA methods).

In general, the covariate was implemented as

$$\text{CL} = \theta_{\text{Base}} \cdot \theta_{\text{COV}}^{Ri} \cdot e^{\eta} \qquad (2)$$

with R_i being the covariate of interest, *e.g.*, SEX or COM, θ_{COV} the parameter estimate for the covariate effect, and η signifying the inter-individual variability in CL. For the EXTRA method, where covariate was missing, the following equation was used:

$$\text{CL} = \theta_{\text{Base}} \cdot \theta_{\text{MGRP}} \cdot e^{\eta} \qquad (3)$$

with θ_{MGRP} estimated as the coefficient for the "missing group." For the mixture methods, a probability of belonging to either group of the covariate was defined as:

$$\begin{aligned} Pr(R_i = 1) &= \theta_{\text{R}} \\ Pr(R_i = 0) &= 1 - \theta_{\text{R}} \end{aligned} \qquad (4)$$

The fraction(s) of subjects in each category (θ_{R}) were either set to the fraction observed in the non-missing subjects (MIX$_{\text{obs}}$) or estimated based on the maximization of the likelihood of PK concentration data alone (MIX$_{\text{conc}}$). In the MIX$_{\text{joint}}$ approach, the covariate data were added as dependent variables to the dataset and jointly analyzed in NONMEM based on a joint maximum likelihood of the nonlinear mixed-effects analysis of concentration data and the likelihood of observing the covariate data. For missing covariates of R_i, the maximum likelihood estimate for R_i in MIX approaches was used in the covariate effect equation (Eq. 2). In the Electronic Supplementary Material, NONMEM code snippets of these implementations are supplied.

Analysis of Output

The imprecision and the bias of parameters were evaluated using the relative root mean squared error (RMSE) across all 100 runs, *e.g.*, for CL.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(\frac{\text{CL}_{est} - \text{CL}_{nom}}{\text{CL}_{nom}} \right)^2}{n}}$$

with CL_{est} and CL_{nom} being the estimated and true CL, respectively. A difference in OFV of >3.84, correlating with a significance level of $p<0.05$, was considered a significant covariate effect. The following output measures were recorded and evaluated:

- RMSE of CL, θ_{male}, and p_{mix} (MIX methods only)
- RMSE for other PK parameters
- Type I/II error rate of covariate effect

Type I error rates were calculated as the percentage or runs that identified DUM as a significant covariate on CL. Type II error rates, *i.e.*, the probability of falsely rejecting a true covariate relationship, were calculated as the percentages of runs that did not identify SEX as a significant covariate on CL. For the calculations of type I/ II errors, simulations were repeated 500 times instead of 100 times to allow a more precise calculation of these parameters.

Real Dataset 1: Efavirenz

A PK dataset of efavirenz was analyzed, containing data from 172 outpatients, including 315 plasma concentrations at a single time point, and 40 full PK curves (13). A linear two-compartment oral model, with first-order absorption, was fit to these data. In the previously described analysis, only two significant covariates were identified: belonging to the Asian race (19.8% missing, $n=34$) and having a total bilirubin level of >1.5 the upper level of normal (11.0% missing, $n=19$), which both significantly increased the bioavailability. The covariate equation that was implemented was

$$F = F_{base} \cdot \theta_1^{\text{Asian}} \cdot \theta_2^{\text{Bili}}$$

with Asian and Bili being binary operators. The original (continuous) laboratory values for Bili were not available for analysis. Since only data after oral administration of efavirenz were available, the absolute bioavailability could not be estimated and F_{base} was set to 1 (a logit transformation of F was not used here). In the original analysis, the EXTRA method was used for handling the missing data; thus, besides the effect sizes of the covariates, two additional effect sizes were estimated for subpopulations where data were missing.

Real Dataset 2: Nevirapine

A dataset of nevirapine was reanalyzed, which consisted of data from 173 patients with nevirapine plasma concentrations measured at 757 time points, including full curves from 13 patients (4). A linear one-compartment model with elimination from the central compartment was fit to the data, which included inter-individual and inter-occasion variability in CL, but not in the other parameters. Three covariates were identified as significant: weight (WT), hepatitis C (HepC) comorbidity, and elevated aspartate aminotransferase (ASAT) >1.5 times the upper limit of normal. From only six patients, data on ASAT were missing (3.5%); the EXTRA method was applied to handle this. For all patients, WT was available. The following equation was used as the covariate model:

$$\text{CL} = (\text{CL}_{base} + (\text{WT} - 70) \cdot \theta_1) \cdot \theta_2^{\text{HepC}} \cdot \theta_3^{\text{ASAT}}$$

with WT being a patient's weight in kilograms and θ_1 the effect of weight on CL normalized to 70-kg individuals. θ_2 and θ_3 are the effects on clearance of hepatitis C comorbidity and elevated ASAT, which are incorporated into the equation by the binary operators HepC and ASAT. Again, the original (continuous) laboratory values were not available for analysis.

Software

NONMEM 7.1 (ICON Development Solutions, Ellicott City MD, USA) (9) was used for the simulation and estimation. The FOCEI method was used throughout the analysis, except for the estimation of $\text{MIX}_{\text{joint}}$, for which the Laplacian conditional estimation method was used. Pirana was used as the modeling environment (14), and the statistical software R (version 2.10.1 or later;

www.r-project.org) was used to perform descriptive statistics and to create plots.

RESULTS

Simulation Study

Figure 2 shows the distribution of bias for CL obtained with each missing data method at the two levels of missingness, while Table II shows the RMSE for CL and the covariate effect θ_{male} in the balanced datasets with random missingness (MCAR). Table II shows that, as expected, RMSE for the estimation of CL (Fig. 2), V, and the covariate effect magnitude θ_{male} (Fig. 3) were equal between all methods at both 10% and 50% of missingness.

In Fig. 4, the distribution of estimates for p_{mix} is shown for the MIX methods, again for the balanced datasets with MCAR data. This shows that at both levels of missingness, the $\text{MIX}_{\text{joint}}$ method performed well at estimating p_{mix} and provided a less biased estimate of the true mixing probability than when using the observed fraction. Remarkably, the distribution of estimated MIX_{conc} was very wide and almost uniform, showing that this method was unable to correctly estimate p_{mix}. Additionally, in many runs, the estimation of p_{mix} ran into a boundary (0 or 1), thereby hindering model convergence. We attempted to apply a logit transformation to constrain the p_{mix} between 0 and 1 instead of applying parameter boundaries. Although this did resolve the boundary problems, the percentage of successful runs did not improve and the logit-transformed value for p_{mix} shot up to very high or very low values. For the mixture methods, no correlation was observed between p_{mix} and θ_{male}.

The simulations studying the situation of the covariate having no effect on PK parameters showed similar performance to the situation in which there was an actual effect present: again, all methods performed well at both 10% and 50% of missingness (results not shown). Especially for MIX_{conc}, the overestimation of effect size was also associated with high type I error rates compared to $\text{MIX}_{\text{joint}}$, which is shown in Table III. It was observed that at both 10% and 50%, the type I error using the EXTRA methods was higher than when using the mixture methods or the DROP method. Lowest type I errors were obtained using the MIX_{obs} method at both missingness levels. Type II errors (false negatives) were about equal for all methods. The $\text{MIX}_{\text{joint}}$ produced slightly higher type II errors than the other mixture methods. In the "sparse sampling scenario," type I and II errors were similar between methods and close to nominal value, and were not increased for EXTRA.

In line with expectations, when models were fitted to the "informative" datasets, which were simulated with a higher effect magnitude and lower magnitudes of intra-individual variability and residual error, estimation of covariate effects was much more precise and comparable between methods. Estimation of p_{mix} was also comparable between all mixture methods. The results of the analyses of the "informative" datasets are not shown in this paper as the differences between methods were very low or non-existent and probably also not commonly observed in actual PK analyses. Results from the scenario which analyzed simulated data with a higher percentage of IIV (50% in both CL and V) were also
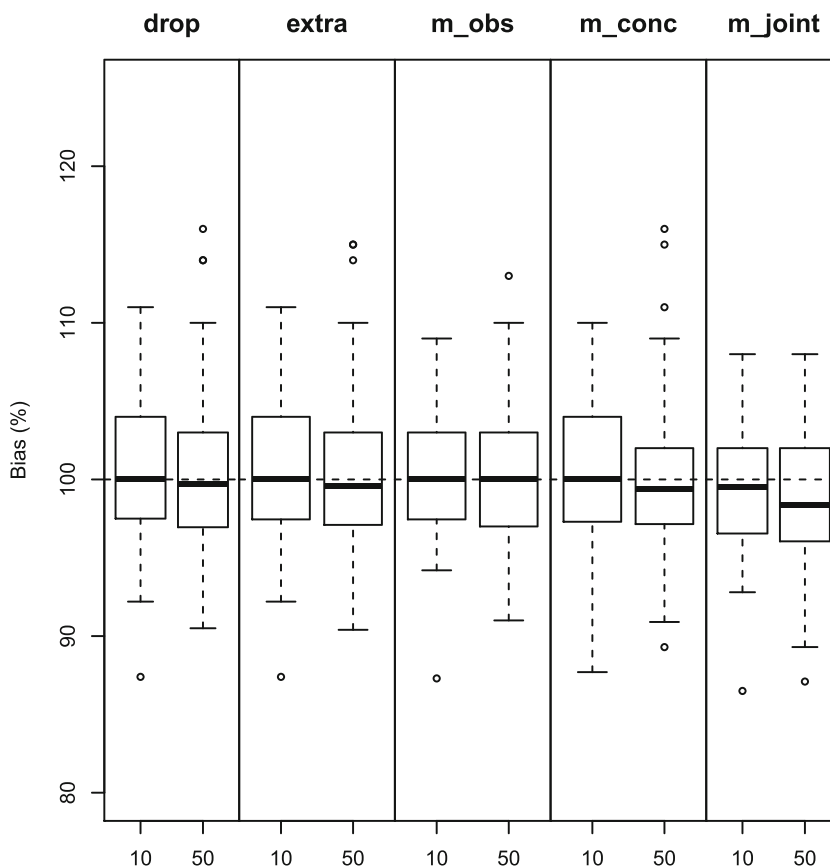
Fig. 2. Distribution of bias in CL, balanced dataset. *Numbers in the bottom* indicate RMSE

evaluated, which were very similar to those obtained for the "moderately informative" scenario and are therefore not presented.

In the unbalanced datasets "F+" and "M+," the estimation of CL and the covariate effect magnitude were highly similar to the results obtained with the balanced (50/50) dataset using all missing data methods (not shown). The estimation of mixing probabilities in the MIX methods showed the lowest bias and imprecision at all levels when using MIX$_{joint}$ (Electronic supplementary material (ESM) Fig. S-1), similar to our experience with the balanced dataset.

In the situation of non-random missingness with 50% missing covariate data, low bias and imprecision in the estimation of CL and θ_{male} were observed for the DROP and EXTRA methods, showing RMSE values similar to those obtained with the balanced dataset. However, using the MIX$_{obs}$ method, a considerable positive bias in the estimation of CL was observed (mean bias, +10%; RMSE=8%), while bias was much lower using MIX$_{conc}$ or MIX$_{joint}$, which is shown in Fig. 5. Furthermore, p_{mix} could only be estimated without bias using the MIX$_{joint}$ method and was overpredicted using MIX$_{conc}$ (not shown). These data suggest that if the estimate for p_{mix} using the MIX$_{conc}$ method is markedly different from p_{mix} established from the observed data, this might be an indication of the presence of non-random missingness (MNAR) in the dataset. When re-estimating models based on datasets containing no covariate effect, the estimation of CL was unbiased with all methods, although MIX$_{conc}$ and, to a lesser extent, MIX$_{joint}$, overestimated the parameter value.

The performance of the methods in the *sparse sampling* scenarios showed similar performance for all methods regarding bias and imprecision in parameter estimation, with good to moderate estimation for CL, but a much higher bias/imprecision for the covariate effect magnitude. Most

Table II. RMSE Values Obtained for Missing Data Methods for Both the Typical PK Scenario and the Sparse Sampling Scenario

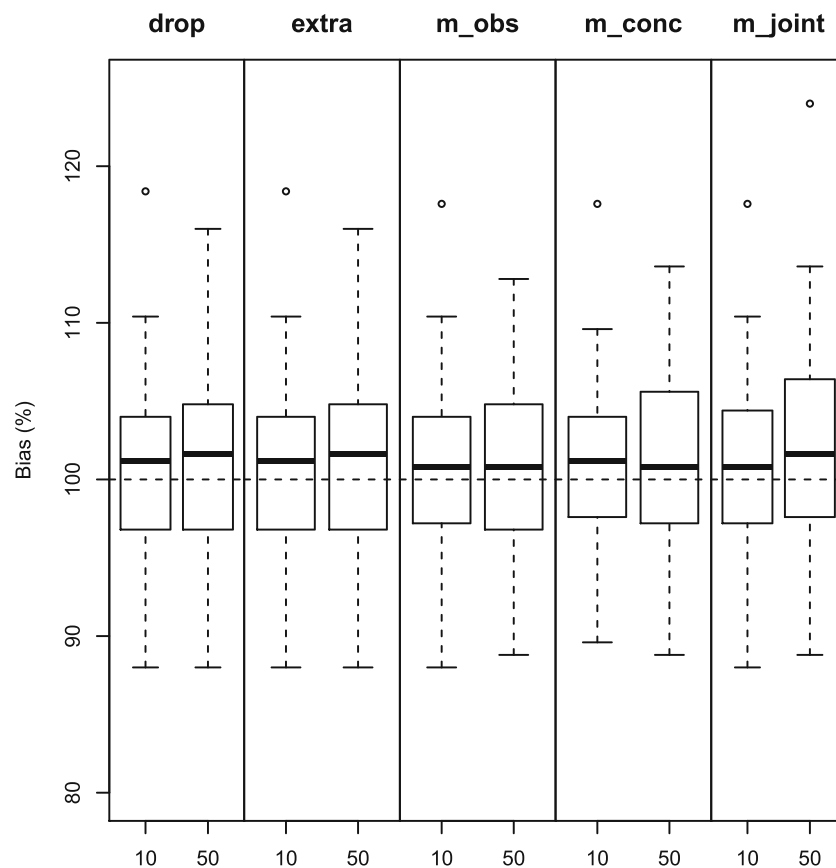| Error	Missingness (%)	DROP (%)	EXTRA (%)	MIX$_{obs}$ (%)	MIX$_{conc}$ (%)	MIX$_{joint}$ (%)
CL	10	4	4	4	4	4
	50	5	5	4	5	4
θ_{male}	10	6	6	6	6	6
	50	7	7	7	7	7
CL (sparse)	10	23	23	22	22	27
	50	34	27	22	25	28
θ_{male} (sparse)	10	48	229	265	259	204
	50	505	207	356	304	270

Fig. 3. Distribution of bias in covariate effect, balanced dataset. *Numbers in the bottom
indicate RMSE*

interestingly, at the higher level of missingness, the ability to identify the true structural model (two-compartment *versus* one-compartment) was different between DROP and the other methods. For DROP, the ability to identify the correct model from the simulated data was about similar for the 10% missingness scenario, but in the 50% missingness scenario, for 14% of the re-estimations, the simpler one-compartmental model would have been selected (using a cutoff of $p < 0.01$ (dOFV > 6.63 in model selection) as the structural model (ESM Fig. S-2 compares DROP and EXTRA methods).

For the simulated scenario of a *time-varying* covariate, the performance of both DROP and EXTRA methods was good, as can be seen from Fig. 6. For the scenario in which data were either MCAR or MNAR, bias and imprecision were not relevantly increased for either method.

Real Dataset 1: Efavirenz

Parameter estimates for CL were highly similar between all missing data methods; however, *V* was considerably different between estimates using the mixture methods and EXTRA methods (Table IV). Additionally, estimates for *ka* were different between methods. However, similar estimates were obtained for θ_{Asian} when using any of the mixture methods, but were estimated to be about 30% higher than when using the EXTRA method. It was noted that model estimation was not supported from the remaining data when using the DROP method as the parameters for the structural

model established for this compound (two-compartment) could not be estimated. Estimation of a one-compartmental model resulted in the phenomenon of flip-flop PK, which could not be resolved easily, and the DROP method was therefore considered invalid for this specific analysis. The mixture methods were also able to estimate the effect magnitude with a lower imprecision than the EXTRA method (lower relative standard errors, RSE). The difference between methods in the estimation of θ_{Bili} was less pronounced, with slightly higher estimates for the mixture method compared to EXTRA. Mixing probabilities estimated with the MIX_{joint} method for both covariates were close to those observed and could be estimated with reasonable accuracy (~30%). In contrast, the MIX_{conc} method estimated the p_{mix} close to 0% (resulting in an imputation of 0 for all subjects for this category) and was therefore fixed to this percentage to allow the estimation of uncertainty for the other parameter estimates.

Real Dataset 2: Nevirapine

Table V shows that the results obtained for all model parameters were highly similar between missing data methods, including estimates for CL and θ_{ASAT+}. This was expected as only a low number of data was missing. However, due to this low level of missingness, use of the MIX_{conc} was shown to be not feasible: estimation of p_{mix} was not possible since the estimation ran into the 100% boundary, and p_{mix} was therefore fixed.
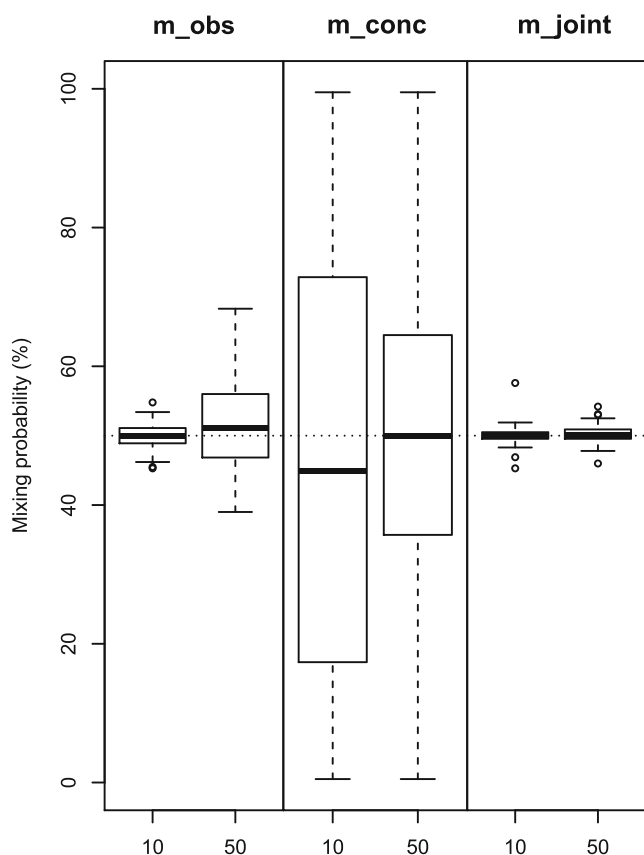
Fig. 4. Estimated p_{mix}, balanced dataset

DISCUSSION

In this article, we investigated several methods of handling missing covariate data in population PK analyses, implemented in NONMEM. The simulation studies and modeling analyses of real datasets described here showed the advantages and drawbacks of the use of various methods for handling missing categorical covariate data. Adequate assessment of covariate relationships is important since a biased estimation of effect size may also lead to falsely establishing significance of covariates or erroneous rejection of true covariate relationships. A frequently used approach for handling missing data is single or multiple imputations (MI) (15), but these techniques were not evaluated in this article. MI has been applied in modeling with NONMEM, *e.g.*, in (16), but this technique requires additional simulation steps and cannot be deployed in a single NONMEM estimation. Furthermore, since imputation requires additional computation, estimates obtained by these methods may be considered less efficient than those obtained from likelihood-based methods (15). Therefore, we chose not to include imputation methods in this analysis and focused on methods that can be applied more conveniently from within NONMEM. The current analysis was restricted to binary covariates. While some of our results and conclusions may be extrapolated to categorical covariate data of more than two categories (*e.g.*, metabolizer status), the performance of the missing data methods should ideally be investigated as a new case.

In the simulation study presented here, it was shown that all five methods provided adequate performance (bias/imprecision) at either low (10%) or high (50%) level of missingness. DROP is the simplest method presented here, but since in the DROP method data are discarded, information not only on the covariate effect but also on other PK parameters is discarded, thereby qualifying on theoretical grounds as an inferior method compared to the other methods. This issue was demonstrated in the analysis of sparse data for the two-compartmental PK model. In 14% of the cases, the simpler one-compartmental model was chosen over the true model, *i.e.*, resulting in lack of significance due to the lack of data. EXTRA is the next simplest method to implement and performs rather well in the estimation of covariate effects and PK parameters. A disadvantage of using the EXTRA method is that it showed higher type I errors. This is probably due to the fact that, in principle, a model misspecification is introduced in the model and model misspecifications are usually associated with higher type I error rates. This could be circumvented by adjusting to the actual significance level by performing a simulation and re-estimation analysis for the covariate analysis (17), although it is impractical to do so for every covariate analysis. A considerable advantage of the DROP and EXTRA methods over the MIX methods is that they are relatively simple to implement and can be applied in analyses were data from more than one covariate are missing

Table III. Type I/II Error Rates for Covariate Inclusion in Simulation Study, Balanced Dataset, for Both the Typical PK Scenario and the Sparse Sampling Scenario

Error	Missingness (%)	DROP (%)	EXTRA (%)	MIX$_{obs}$ (%)	MIX$_{conc}$ (%)	MIX$_{joint}$ (%)
Type I	0[a]	6	6	6	6	6
	10	8	19	8	10	10
	50	4	15	4	8	9
Type I, sparse sampling	0[a]	8	8	8	8	8
	10	7	6	7	8	13
	50	6	8	6	13	7
Type II	0	2	2	2	2	2
	10	2	1	1	1	3
	50	9	8	5	5	11
Type II, sparse sampling	0	59	59	59	59	59
	10	65	63	63	57	71
	50	75	77	75	64	75

[a] Since missingness was 0%, the missing data methods were not actually applied for the 0% scenario, so the type I/II errors indicate the "gold standard"
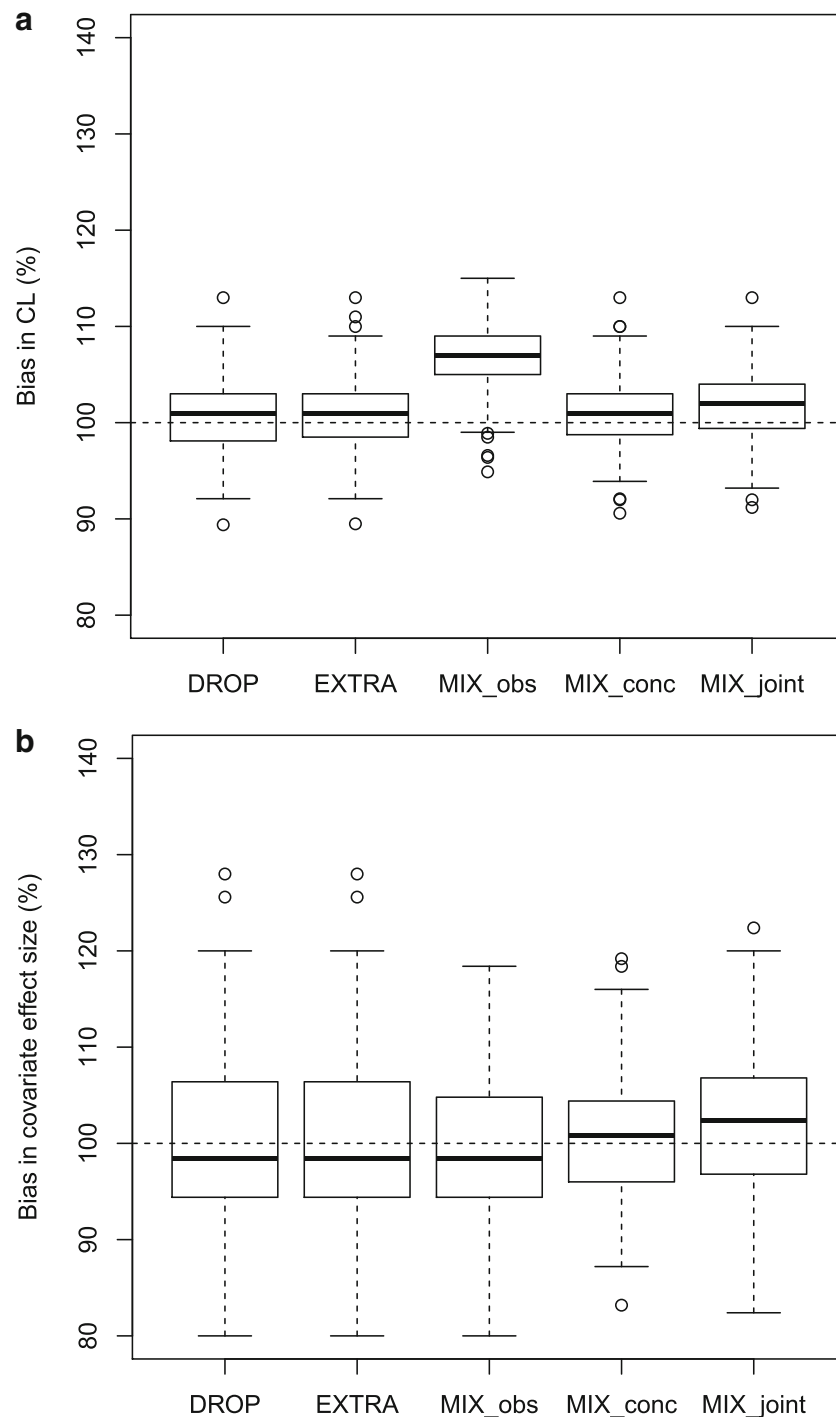
a

Bias in CL (%)

b

Bias in covariate effect size (%)

DROP   EXTRA   MIX_obs   MIX_conc   MIX_joint

Fig. 5. "NonRand" dataset: bias in CL (*left*) and covariate effect size (*right*)

as well as in automated covariate modeling procedures. For a scenario studying a sparse sampling design, performance was very similar across methods.

MIX_{obs}, the simplest mixture method for dealing with missing covariate data, often provides adequate results and also showed the lowest type I/II error of covariate inclusion of all methods. However, when non-random missingness (MNAR) is present in the data, the use of the observed category distribution, as is done in the MIX_{obs} method, induced bias in the estimation of the

parameter correlating with the covariate (CL in this analysis) while estimation of the covariate effect magnitude itself was unaffected (Fig. 6). In this situation, all other methods resulted in estimates of CL and covariate effect size that were only slightly biased. Forcing the mixture model into a different mixture than present in the actual overall population causes the failure of MIX_{obs} in the situation of MNAR data. In our example, this could only be "corrected" by estimation of a higher value for CL. The other methods showed very low bias in the
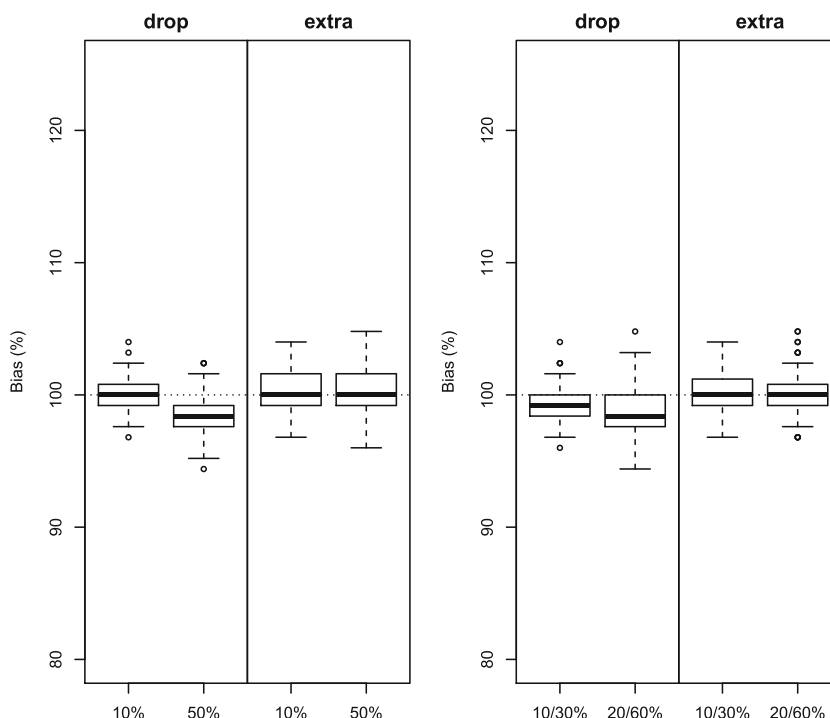
Fig. 6. Distribution of bias in covariate effect for the datasets with time-varying covariates for MCAR (*left*) and MNAR (*right*) scenarios

MNAR scenario. It may be that at more extreme scenarios of MNAR (for even higher probabilities of missingness in one category or at even higher percentages of missingness), bias and imprecision will increase.

Both in the simulations and in the analysis of real data, MIX$_{\text{conc}}$ performed poorly, regularly running into the 0 or 1 boundary for p_{mix}, which was (logically) especially the case when the amount of missing data was low. As expected, analysis of datasets with larger covariate effects ($\theta_{\text{male}}=2$)

showed better performance, equaling the performance of MIX$_{\text{joint}}$.

The analysis of the efavirenz dataset showed that the MIX methods also allow the handling of more than one covariate with missing data. However, the implementation of mixture models quickly becomes more complex at higher numbers of covariates with missing data, requiring x^n subpopulations in the mixture model for n covariates with missing data and x categories per covariate as well as a

Table IV. Parameter Estimates for Missing Data Methods: *Efavirenz* Dataset and Model

Parameter	Unit	DROP[a]	EXTRA	MIX$_{\text{obs}}$	MIX$_{\text{conc}}$	MIX$_{\text{joint}}$
CL	L h^{-1}	–	11.9 (4%)	12.1 (4%)	12.1 (4%)	11.6 (4%)
V	L	–	308 (19%)	149 (19%)	149 (19%)	144 (27%)
Q	L h^{-1}	–	30 (30%)	40.4 (12%)	40.4 (12%)	40.6 (11%)
V_2	L	–	244 (45%)	332 (30%)	333 (30%)	332 (31%)
ka	h^{-1}	–	0.591 (19%)	0.268 (27%)	0.267 (29%)	0.277 (50%)
θ_{Asian}	–	–	1.57 (26%)	2.04 (15%)	2.03 (14%)	2.04 (16%)
$\theta_{\text{Asian,miss}}$	–	–	1.13 (17%)	–	–	–
$\theta_{\text{bili+}}$	–	–	1.67 (17%)	1.71 (16%)	1.75 (16%)	1.71 (15%)
$\theta_{\text{bili+,miss}}$	–	–	0.922 (17%)	–	–	–
$p_{\text{mix,Asian}}$	%	–	–	7.25[b]	8.55 (70%)	7.36 (28%)
$p_{\text{mix,bili}}$	%	–	–	7.97[b]	0.1[c]	6.97 (29%)
ω_{CL}	%	–	47.0 (9%)	45.2 (9)	45 (9%)	44.9 (9%)
ω_{ka}	%	–	–	68.9 (21)	69 (21%)	55.4 (18%)
$\varepsilon_{\text{prop}}$	%	–	23.9 (15%)	24 (15%)	24 (15%)	24.2 (16%)
ε_{add}	mg L^{-1}	–	0.312 (21%)	0.312 (21%)	0.312 (21%)	0.310 (24%)

Values in parentheses are the RSEs, calculated as $(\omega/\theta)\times100\%$
[a] Estimation of both one- and two-comparmental models was not supported by the data
[b] Fixed
[c] Estimated close to boundary, therefore fixed

Table V. Parameter Estimates for Missing Data Methods: *Nevirapine* Dataset and Model

Parameter	Unit	DROP	EXTRA	MIX$_{obs}$	MIX$_{conc}$	MIX$_{joint}$
CL	L h^{-1}	3.28 (3%)	3.28 (3%)	3.26 (2%)	3.28 (3%)	3.3 (2%)
V	L	132 (10%)	131 (9%)	131 (9%)	131 (9%)	130 (10%)
ka	h^{-1}	1.94 (14%)	2.13 (14%)	2.13 (14%)	2.13 (14%)	2.22 (15%)
θ_{WT}	L h^{-1} kg^{-1}	0.0175 (39%)	0.0167 (40%)	0.0168 (39%)	0.0169 (40%)	0.0176 (39%)
θ_{HepC}	–	0.727 (11%)	0.725 (11%)	0.731 (10%)	0.731 (10%)	0.717 (11)
θ_{ASAT+}	–	0.882 (9%)	0.884 (10%)	0.885 (9%)	0.864 (7%)	0.888 (10%)
$\theta_{ASAT,miss}$	–	–	0.82 (8%)	–	–	–
p_{mix}	%	–	–	8.98[a]	100[b]	9.06 (25%)
ω_{CL}	%	24.8 (9%)	24.3 (9%)	24.6 (9%)	24.4 (9%)	24.3 (9%)
κ_{CL}	%	12.3 (18%)	12.4 (19%)	12.4 (19%)	12.4 (19%)	11.6 (23%)

Values in parentheses represent RSE, calculated as $(\omega/\theta) \times 100\%$
[a] Fixed
[b] p_{mix} estimated close to boundary, therefore fixed

multitude of additional lines of code to cope with the estimation of the covariate effects. We did not evaluate the estimation of mixture models with more than four subpopulations (=two covariates), but it is likely that for the distinction of higher numbers of subpopulations, estimation of the mixing probabilities will become unstable if the dataset does not contain sufficient information. To complicate matters, correlations between mixture probabilities may also be present in the data, *e.g.*, a larger portion of women having increased bilirubin levels than men. Therefore, the use of mixture models is limited to simpler covariate analyses.

In the simulation study, it was noted that non-random missingness induced a difference in the estimated mixing probabilities between the MIX$_{obs}$ and MIX$_{conc}$ methods. Since such a difference was also found in the analysis of efavirenz for $p_{mix,bili}$, it is tempting to conclude that non-random missingness in that covariate was also present in the dataset of efavirenz. This may of course be possible, but the amount of missingness (11%) and the fraction of patients with elevated bilirubin levels (observed, 8%) were likely too low to substantiate this conclusion. Convergence problems with the DROP method for this problem could not easily be solved. Possibly, with more exhaustive efforts, the problem may have been solved, *e.g.*, using different estimation methods or changes in model structure, but our goal was to judge the general performance in a retrospective analysis without performing additional model construction.

Due to the very low amount of missing data on ASAT (3.5%) in the nevirapine dataset, no relevant differences were observed between parameter estimates obtained with the different methods. The low amount of missing ASAT data explains why p_{mix} could not be estimated in the MIX$_{conc}$ method: due to the low amount of information in the remaining concentration data on ASAT (n=6 patients), this was not sufficient to discriminate between patients with increased and normal ASAT. Using MIX$_{joint}$, p_{mix} could be estimated adequately due to the extra information on ASAT supplied to the model, and the estimate expectedly did not differ much from the observed p_{obs}.

A drawback of implementing the MIX$_{joint}$ method is that the Laplacian estimation method in NONMEM has to be used, which is known to be more prone to convergence problems during estimation. We did not, however, experience any problems during the simulation study or with the real dataset. Another hurdle in the implementation of MIX$_{joint}$ is that the dataset needs to be updated to include the data on the studied covariate as a dependent variable. In addition to the approaches presented here, it is possible to incorporate the observed fraction of patients as a prior on the estimated covariate effect. In terms of dependence on the use of observed covariate data for the estimation of mixing probability, this approach can be placed somewhere between MIX$_{obs}$ and MIX$_{conc}$, depending on the weight given to the priors. Such an approach is likely to perform similar to MIX$_{joint}$ and would also circumvent possible problems associated with using the Laplacian estimation method.

Concluding, we showed that at low to intermediate percentages of missingness, all evaluated methods perform similarly with respect to bias and imprecision in parameter estimates. In the situation of non-random missingness, DROP and EXTRA methods showed less bias than the mixture methods. Since it is usually not known whether data are missing randomly or not, and due to its ease of implementation, the DROP or EXTRA method can be considered appropriate methods for handling missing categorical data. If the data are rich enough to allow appropriate model construction and parameter estimation, the DROP method is advised since the EXTRA method may increase the type I error for covariate selection. As shown, the DROP method is, however, data inefficient. Since the simulation scenarios presented in this article are inherently arbitrary, if the problem is relevantly different from the scenarios presented here, it is advised to implement a simulation study to evaluate the performance of missing data methods.

ACKNOWLEDGMENTS

REFERENCES

1. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods. 2002;7(2):147–77.
2. Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol. 2008;60:549–76.
3. Bosch TM, Huitema ADR, Doodeman VD, et al. Pharmacogenetic screening of CYP3A and ABCB1 in relation to population pharmacokinetics of docetaxel. Clin Cancer Res. 2006;12 (19):5786–93.
4. de Maat MM, Huitema AD, Mulder JW, Meenhorst PL, van Gorp EC, Beijnen JH. Population pharmacokinetics of nevirapine in an unselected cohort of HIV-1-infected individuals. Br J Clin Pharmacol. 2002;54(4):378–85.
5. Joerger M, Huitema ADR, van den Bongard D, Schellens JHM, Beijnen JH. Quantitative effect of gender, age, liver function, and body size on the population pharmacokinetics of Paclitaxel in patients with solid tumors. Clin Cancer Res. 2006;12(7 Pt 1):2150–7.
6. Joerger M, Huitema AD, van den Bongard HJ, et al. Determinants of the elimination of methotrexate and 7-hydroxy-methotrexate following high-dose infusional therapy to cancer patients. Br J Clin Pharmacol. 2006;62(1):71–80.
7. Zandvliet AS, Huitema ADR, Copalu W, et al. CYP2C9 and CYP2C19 polymorphic forms are related to increased indisulam exposure and higher risk of severe hematologic toxicity. Clin Cancer Res. 2007;13(10):2970–6.
8. Rubin DB. Inference and missing data. Biometrika. 1976;63 (3):581–92.
9. Beal SL, Sheiner LB. NONMEM users guides. Ellicott City, MD: Icon Development Solutions; 1989.
10. Stubbendick AL, Ibrahim JG. Maximum likelihood methods for nonignorable missing responses and covariates in random effects models. Biometrics. 2003;59(4):1140–50.
11. Ibrahim JG, Chen M-H, Lipsitz SR, Herring AH. Missing-data methods for generalized linear models. J Am Stat Assoc. 2005;100:332–46.
12. Parzen M, Lipsitz SR, Ibrahim JG, Lipshultz S. A weighted estimating equation for linear regression with missing covariate data. Stat Med. 2002;21(16):2421–36.
13. Kappelhoff BS, Huitema ADR, Yalvaç Z, et al. Population pharmacokinetics of efavirenz in an unselected cohort of HIV-1-infected individuals. Clin Pharmacokinet. 2005;44(8):849–61.
14. Keizer RJ, van Benten M, Beijnen JH, Schellens JHM, Huitema ADR. Piraña and PCluster: a modeling environment and cluster infrastructure for NONMEM. Comput Meth Prog Bio. 2011;101 (1):72–9.
15. Schafer JL. Multiple imputation: a primer. Stat Methods Med Res. 1999;8(1):3–15.
16. Tannenbaum SJ, Holford NHG, Lee H, Peck CC, Mould DR. Simulation of correlated continuous and categorical variables using a single multivariate distribution. J Pharmacokinet Pharmacodyn. 2006;33(6):773–94.
17. Wählby U, Jonsson EN, Karlsson MO. Assessment of actual significance levels for covariate effects in NONMEM. J Pharmacokinet Pharmacodyn. 2001;28(3):231–52.