

**TUTORIAL**

# Approaches to handling missing or “problematic” pharmacology data: Pharmacokinetics

Donald J. Irby<sup>1</sup> | Mustafa E. Ibrahim<sup>2</sup> | Anees M. Dauki<sup>1</sup> | Mohamed A. Badawi<sup>1</sup> |  
Sílvia M. Illamola<sup>3</sup> | Mingqing Chen<sup>1</sup> | Yuhuan Wang<sup>4</sup> | Xiaoxi Liu<sup>4</sup> | Mitch A. Phelps<sup>1</sup> |  
Diane R. Mould<sup>5,1</sup>

<sup>1</sup>Division of Pharmaceutics and Pharmacology, College of Pharmacy, The Ohio State University, Columbus, OH, USA

<sup>2</sup>Department of Pharmaceutical Sciences, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

<sup>3</sup>Department of Experimental and Clinical Pharmacology, College of Pharmacy, University of Minnesota, Minneapolis, MN, USA

<sup>4</sup>Division of Clinical Pharmacology, Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT, USA

<sup>5</sup>Projections Research Inc, Phoenixville, PA, USA

**Correspondence**

Mitch A. Phelps, Division of Pharmaceutics and Pharmacology, College of Pharmacy, The Ohio State University, 506 Riffe Hall, 496 W. 12th Ave., Columbus, OH 43210, USA.  
Email: phelps.32@osu.edu

Diane R. Mould, Projections Research Inc., 535 Springview Lane, Phoenixville, PA 19460, USA.  
Email: drmould@pri-home.net

**Funding information**

No funding was received for this work.

**Abstract**

Missing or erroneous information is a common problem in the analysis of pharmacokinetic (PK) data. This may present as missing or inaccurate dose level or dose time, drug concentrations below the analytical limit of quantification, missing sample times, or missing or incorrect covariate information. Several methods to handle problematic data have been evaluated, although no single, broad set of recommendations for commonly occurring errors has been published. In this tutorial, we review the existing literature and present the results of our simulation studies that evaluated common methods to handle known data errors to bridge the remaining gaps and expand on the existing knowledge. This tutorial is intended for any scientist analyzing a PK data set with missing or apparently erroneous data. The approaches described herein may also be useful for the analysis of nonclinical PK data.

**OVERVIEW**

Data from clinical trials are frequently incomplete, particularly data sets collected during large, late-phase trials or during routine clinical patient care or follow-up visits. Portions of data may be missing or inaccurate as a result of factors such as study site noncompliance, patient

noncompliance, inappropriate sample handling, data entry errors, and analytical problems. How “problematic” data are handled can impact its interpretation, especially when data used for population pharmacokinetic (PPK) modeling contains missing or erroneous data. Before beginning an analysis, pharmacometricians often spend a large portion of time dealing with problematic data.

Donald J. Irby, Mustafa E. Ibrahim, Anees M. Dauki, and Mohamed A. Badawi contributed equally.

Senior authors Mitch A. Phelps and Diane R. Mould contributed equally.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics

During data cleaning (data quality assurance [QA]), the first step is to identify missing or problematic data. Concentration-time data and dosing records are often the primary concern, but other issues, such as missing or questionable covariate data, must also be considered. Once issues/discrepancies are identified, the next challenge is to evaluate frequency of occurrence of each type of problem and the associated reason to establish appropriate methods for handling these erroneous data. Prior studies have addressed handling of specific types of problematic data, although no set of broad recommendations spanning the various types of problematic data have been previously presented.

Through reviews of published methods, simulations of data sets with known errors, and evaluations using different methods for handling these errors, this tutorial aims to provide guidance for dealing with problematic clinical (and some nonclinical) concentration versus time, dosing, and covariate data.

This tutorial is intended to be used by scientists analyzing pharmacokinetic (PK) data with either missing data or where apparently questionable or erroneous data are present. Although QA and data quality control (QC) are essential to successful modeling, this tutorial assumes the data set has already undergone appropriate QC or was assembled from locked, clean data. Basic assessments include exploratory data analysis by plotting and summarizing existing data (e.g., summary statistics) to ensure all data are within expected ranges relative to the subject population. Communication with the clinical research staff, analytical laboratory scientists, and/or sponsor is important to attempt to explain any missing or apparently problematic data. For additional information on pharmacology data QA and QC, we refer the reader to previously published reviews.<sup>1,2</sup>

## METHODS

### Simulated data sets

In addition to reviewing published studies evaluating methods for handling missing or erroneous data, simulation and estimation was used to fill in gaps or to expand on published information. A single 400-subject data set was simulated using SimCyp<sup>3</sup> to ensure appropriate ranges of weight, age, creatinine clearance (CrCL) and sex. This covariate data set was randomly sampled without replacement to generate 100 data sets of 200, 100, 50, or 25 unique subjects with the widest possible range of covariate values to evaluate the impact of missing or erroneous concentration data, data that are below limit of quantification (BLQ), and the impact of missing or incorrect covariate data on the ability to accurately and precisely estimate PK parameter values. A total of 100 replicates of concentration-time data were simulated with each of the data sets using multiple PK models, including

intravenous (i.v.) bolus and oral dosing, one-compartment and two-compartment structural models, various scenarios of clearance (CL) and volume of distribution (V), and various covariate effect levels (covariate factor [FAC]) for continuous (weight [WT]) or categorical (sex [SEX]) covariate effects on CL. For each scenario, three sizes of the covariate effect (weak, 0.3 and 0.8; moderate, 0.5 and 0.7; or strong influence, 0.75 and 0.5; for continuous and categorical covariates, respectively) on CL were evaluated. The effect of WT on CL was modeled with a power function as in Equation (1):

$$CL = \theta_{CL} \cdot \left( \frac{WT}{70} \right)^{FAC} \quad (1)$$

where  $\theta_{CL}$  represents the typical value of CL for a given value of WT normalized to 70 (kg). The effect of SEX on CL was modeled according to Equation (2):

$$CL = \theta_{CL} \cdot FAC^{SEX} \quad (2)$$

where SEX was coded as 0 for females and 1 for males. Various percentages of data were removed or adjusted to simulate missing or erroneous data, respectively, for each evaluation. Table 1 provides a summary of the study subject characteristics. Additional details on the simulated data sets are in the Supplementary Material.

### General model-fitting approaches

Each simulated data set with either BLQ or missing or erroneous covariate data was evaluated using NONMEM (version 7.3; Icon PLC, Dublin, Ireland) with the first-order conditional estimation method, implementing the M3 method, as appropriate (see the Concentration data BLQ section), to estimate model parameters. Model output was automatically tabulated and summarized graphically using an R script. Additional details on the data simulation and parameter estimation processes, including NONMEM code, are provided in the Supplementary Material. Overall flow diagrams of the simulation and estimation approaches are shown in Figure S1.

**TABLE 1** Summary of 400-patient data set

Covariate	Median (minimum, maximum)	
	Females, <i>N</i> = 200	Males, <i>N</i> = 200
WT	67 (40, 117)	79 (53, 120)
AGE	41 (21, 60)	43 (21, 65)
SECR	0.6 (0.4, 1.3)	0.8 (0.5, 1.4)
CrCL	123 (41, 315)	127 (74, 234)

Abbreviations: AGE, age (years); WT, baseline body weight (kg); CrCL, creatinine clearance (ml/min); SECR, serum creatinine (mg/dl).

## Analysis of simulated missing data scenarios and data-handling methods

To evaluate the impact of various factors on the accuracy and precision of estimated PK parameters, multiple output variables between the experimental (i.e., 5%–50% BLQ or missing/erroneous data) and control (i.e., 0% BLQ or 0% missing/erroneous data) conditions were compared. Output parameters can be found in the Supplementary Material Output Tables sections. Model performance in all scenarios was evaluated by comparing parameter estimates, associated relative standard errors (RSEs), proportion of successful covariance steps, and measures of bias (relative mean error [RME]) and precision (root mean square error [RMSE]). Details on these metrics for evaluating the handling methods applied to the various missing or erroneous data scenarios are provided in the Supplementary Material.

## CONCENTRATION VERSUS TIME DATA

PPK models should be developed from data where all dosing times, concentration measurements, and sampling times are known. Unfortunately, that is rarely the case. Two commonly encountered types of problematic PK data are errors in concentration measurements and errors in recorded PK sampling times. Causes for erroneous concentrations may include problems with the bioanalytical assay (e.g., assay variability, presence of interfering endogenous compounds), differences between assay laboratories and/or changes to the analytical technique over time, sampling anomalies (e.g., collecting a PK sample from an i.v. line used to deliver the drug dose, poor or variable sample stability, mislabeling samples), or errors in recorded data. Cause for errors in sampling times may include failure to capture actual sample time or recording an incorrect sample time.

### Analytical assay accuracy and precision

The analytical method used to quantify a given compound and/or metabolites is an essential element of PK studies. All analytical measurements have associated error, and Graves et al.<sup>4</sup> investigated the impact of this error on PK parameters estimated from several methods for three different ratios of absorption rate constant to elimination rate constant ( $K_a/K_e$ ). Although assay error is often a minor component of residual unexplained variability (RUV), the authors found that random assay error may alter PK curve shape leading to false inclusion of additional compartments. For the lowest  $K_a/K_e$  ( $K_a/K_e = 2$ ), all parameter estimation methods except the nonlinear mixed-effects approach underestimated  $K_e$  and overestimated  $K_a$  and  $V$ . These biases increased when assay

error increased up to 30%. Even at the allowable  $\pm 20\%$  at the lower limit of quantification (LLOQ) level according to bioanalytical guidelines,<sup>5,6</sup> Graves and colleagues found that at 20% “assay error,” PK parameters are already overestimated. A review of the bioanalytical validation report before starting data analysis will signal the potential of obtaining biased PK estimates because of the accuracy and precision of the analytical assay used (e.g., overestimated PK parameters may be obtained if the bioanalytical method used has acceptance criteria close to 20% at the LLOQ level).

### Assay selectivity–endogenous interference

When concentration data are confounded by interfering endogenous compounds (including the same analyte in cases where the drug itself is an endogenous compound, such as a glucocorticoid or growth factors), baseline correction can be applied.<sup>7</sup> However, this methodology ignores the analytical assay variability, and postdose differences do not consider measurement errors associated with endogenous baseline. Other methods have been proposed to model endogenous compound concentrations.<sup>8,9</sup> With linear kinetics, when the rate of endogenous drug production is constant, without feedback control of production (endogenous drug is at steady state), the effect of the intrinsic drug represents a constant increment to exogenous drug in the sampled compartment (Supplementary Code 1).

In some cases, however, there may be fluctuations of the endogenous values over time attributed to various feedback or control systems. Circadian rhythms are evident in many physiological and endocrine systems, leading to concentrations that fluctuate over time. Circadian rhythms can be represented using trigonometric functions that oscillate over 24 h (Equation 3),

$$E_{\text{placebo}} = \text{amplitude} \times \cos\left((T - \text{phase}) \times \frac{2\pi}{24}\right) \quad (3)$$

where amplitude is the maximum extent of the oscillation from the point of equilibrium,  $\cos$  is the cosine function,  $T$  is time, and phase is the amount of time between successive oscillations. Circadian patterns may be evident, particularly if placebo data are collected over a period of time. However, if placebo data are collected at limited times (e.g., daily, monthly), and sampling only occurs at times that reflect the same part of the curve, a potential circadian rhythm may be hidden. If unrecognized, circadian patterns can obscure significant concentration-effect relationships. Rose and colleagues<sup>10</sup> aimed to characterize the PPK of a recombinant follicle-stimulating hormone (FSH) taking into account endogenous FSH levels. In this case, the change in endogenous FSH amount over time in the central compartment was described as a turnover model with a zero-order production rate (Equation 4):

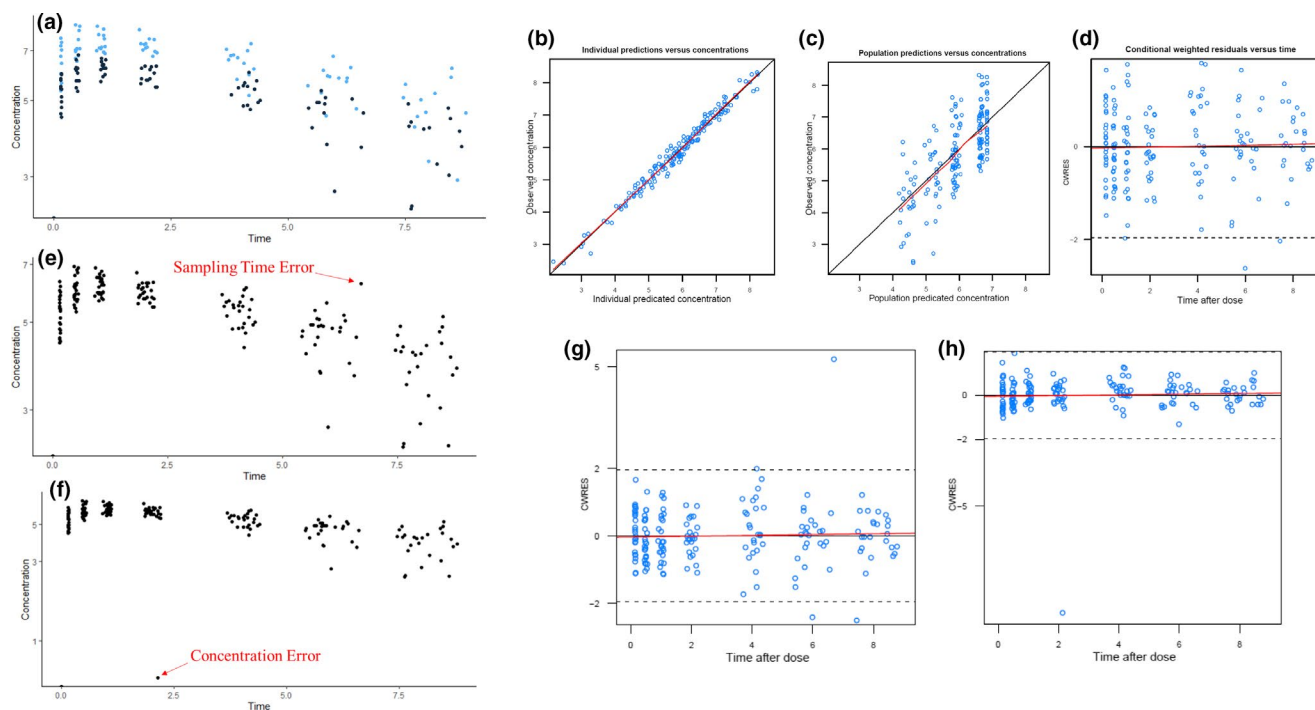
$$\frac{d\text{FSH}_{\text{en}}(t)}{dt} = K_{\text{in}} - K \times \text{FSH}_{\text{en}}(t) \quad (4)$$

where  $\text{FSH}_{\text{en}}$  is the endogenous FSH level,  $K_{\text{in}}$  is a zero-order production rate, and  $K$  is the elimination rate constant. The authors demonstrated that total FSH concentration is affected, and endogenous FSH levels change over time as a consequence of feedback from ovarian hormones in subjects with endogenous FSH levels. If not accounted for, the resulting PK parameter estimates are biased, leading to erroneous characterization of drug disposition. Therefore, whenever initial conditions of the system are nonzero, regardless of the amount of the endogenous compound in circulation, the initial amount or concentration must be accounted for by the model so that its circulating concentration is considered during PK analysis. However, adjustments for endogenous substance and any time dependent variations that it exhibits would depend on the overall ability to measure the endogenous substance and the impact on exogenous substance administered.

## Interassay bias

Ideally, concentration data used for PK analysis should be obtained from the same laboratory using the same validated analytical assay, or when multiple analytical labs are used, those

assays should be cross-validated. However, when data originates from multiple assays, the interchangeability of results from those different laboratories or assays should be evaluated. A graphical visualization of dose-normalized concentration (for drugs with linear PK) versus time data may reveal analytical discrepancies. Figure 1a shows an example where results from assays 1 and 2 are not in agreement. Notably, although we can clearly see the bias in the concentration-time plots in this example, there is no obvious indication of bias in the residual or diagnostic plots for this analysis (Figure 1b–d). Furthermore, assay bias is more often subtle and less apparent. Assessing “assay” as a covariate (e.g., on CL) is therefore recommended and would also be a strategy for estimating the degree of difference or bias between the assays. If plots and/or covariate analysis suggest disagreement between assays, a scaling factor that eliminates method differences can be introduced (Supplementary Code 2). However, this scenario is undesirable since the pharmacometrician may not know which assay should be adjusted. In addition, even if the assay results are interchangeable, assay precisions may be different. This potential variability between assays should be taken into account in the analysis, and separate residual errors should be used by introducing an independent indicator variable (NONMEM Guide 2011<sup>11</sup>). This variable (ASAY), which indicates the assay methodology for each concentration, is defined with a value of 0 for assay 1 and 1 for assay 2. Estimating different residual variabilities associated with each



**FIGURE 1** Example quality control plots of a data with different types of errors. Assay bias: (a) concentration versus time scatterplot (blue, assay 1; black, assay 2), (b) individual predicted versus observed, (c) population predicted versus observed, and (d) conditional weighted residuals (CWRES) versus time for a merged set of concentration data obtained from two different assays. Sampling time error: (e) concentration versus time scatterplot and (g) CWRES versus time for a data set where one sample time (2.1 h) was incorrectly labeled (6.7 h). Concentration error: (f) concentration versus time scatterplot and (h) CWRES versus time for a data set where one concentration (5.3 µg/ml) was incorrectly recorded (0.53 µg/ml)

assay is demonstrated in Supplementary Code 3. The equation resolves to Equation (5):

$$Y = F + F \times \text{EPS}(1) \quad (5)$$

when  $\text{ASAY} = 0$ , and Equation (6):

$$Y = F + F \times \text{EPS}(2) \quad (6)$$

when  $\text{ASAY} = 1$ , where  $F$  is the prediction value that is evaluated by the ADVAN model and  $Y$  is the user-defined value relating the differences between  $F$  and the observed data. The estimate of  $\text{EPS}(1)$  reflects RUV associated with assay 1 and  $\text{EPS}(2)$  reflects RUV associated with assay 2. This strategy may improve model predictions by reducing bias in parameter estimates.<sup>12</sup>

## Sample contamination and interfering components

Aside from analytical methodology, errors occurring before and after samples are processed and analyzed sometimes account for significant error in concentration measurements.<sup>13</sup> Sample integrity may be compromised in several ways including sample “contamination” with something that could interfere with drug measurement, sample storage or transportation under incorrect conditions, or mislabeling. For example, blood samples are often collected from a venous catheter where contamination with i.v. fluid may occur if sampling procedures are not followed correctly, which causes sample dilution and lowering of actual drug concentrations. Similarly, when a blood sample is collected from the i.v. line used to deliver a drug, if the line is not flushed properly, the sample is contaminated with the drug in the i.v. line, and measured drug concentrations will be erroneously high.

## Drug stability

Drug stability during sample processing and storage is an important consideration. Proper handling of samples after collection, including processing and storage within a specific time period and transportation at appropriate conditions (e.g., temperature, light) will minimize erroneous results, especially when dealing with chemically or metabolically unstable drugs. This is particularly important in clinical centers where samples are collected then sent to centralized laboratories for analysis.

## Sampling time errors

One of the most common causes of postanalytical error is inaccurate data entry. PK data available for analysis may be

susceptible to measurement error (ME) in sampling times. Specifically, the ME associated with the sampling time variable in PK modeling belongs to the Berkson error type (i.e., the variability in the true sampling time is larger than that of the scheduled time).<sup>14</sup> This type of error is more commonly present in data obtained during routine clinical practice and less so in PK data collected during a well-controlled clinical trial.

Actual sampling times may be missing or recorded incorrectly, and nominal protocol times are used instead. Studies using trough concentrations of drugs with a short half-life for PK modeling provide a good example of blood drawn at times different than scheduled times. Santalo et al.<sup>15</sup> reported vancomycin trough concentrations are often measured early, producing results that do not reflect actual troughs and may subsequently lead to inappropriate dosing. Similarly, as discussed in the next paragraph, dose times that are incorrectly recorded result in an overall shift of the PK curve.

Sampling times are assumed to be correct for most PK clinical pharmacology analyses. However, this may be a flawed assumption. The impact of recording errors for sampling times has been explored and reported previously.<sup>16–18</sup> Based on a sensitivity analysis, Karlsson et al.<sup>17</sup> demonstrated that random shifts of sampling times do not cause significant changes in parameter estimates aside from increasing residual error. These results are in accordance with Wang and Davidian.<sup>18</sup> Recently, Choi et al.<sup>16</sup> conducted simulation studies to evaluate the effects of using nominal time instead of actual sampling time on model parameters. The authors found the most important determinant of bias attributed to Berkson error in time is the degree of the curvature (nonlinearity) of the concentration-time profile. Although the magnitude of ME is another important factor, its effect may not become notable unless the drug exhibits nonlinear PK. Thus, the data should be reviewed first to evaluate the impact of sample time errors. The smaller the curvature around sampling times, regardless of the magnitude of ME, the lower the expected bias in PK parameter estimates. Conversely, when substantial nonlinearity is expected, use of actual sampling times becomes more important. The authors also found that bias decreases systematically as sample number within a dose interval increases (more dense sampling design).<sup>17,18</sup> In this case, it may be assumed that the number of sample time errors falls related to the number of correct sample times with dense sampling.

Errors occurring before or after the assay run are often identified during standard QC of a data set when looking at individual or overall concentration-time plots. They are also sometimes identified when reviewing summary statistics of concentration-time data. Errors in concentration measurements can sometimes be difficult to distinguish from errors in sample time, and in some cases, it may not be clear that erroneous data are present. For example, observing high

concentrations at a given timepoint in one patient relative to others could be the result of incorrectly recorded sampling times (e.g., 6.7 h instead of 2.1 h; Figure 1e) or may be explained by incorrect recording of a concentration (e.g., 0.53  $\mu\text{g/ml}$  instead of 5.3  $\mu\text{g/ml}$ ; Figure 1f). These types of errors typically show up as outlier data points and may be identified in initial modeling through diagnostic plots, such as predicted versus observed concentrations or conditional weighted residual values (CWRES; Figure 1g,h), although sometimes errors are more subtle. In fact, sampling time errors may not be possible to identify without direct knowledge that a sampling time error occurred (e.g., a note made by the clinical site indicating the recorded time was an approximation of the actual sample time). Even in this case, all erroneous concentration or time data identified or suspected, should be removed from the analysis. If time needs to be scaled (e.g., samples from a particular center appear questionable), the variable TSCALE or XSCALE may be introduced in the \$PK block<sup>19</sup> (Supplementary Code 4).

In summary, both measured drug concentrations and recorded sampling times are key components of PK studies. Any interference or deviation from the true condition should be taken into account in the analysis and adjusted, if possible.

### Summary of recommendations related to errors in concentration versus time data

- Before initiating data analysis, reviewing the bioanalytical report will signal the potential of obtaining biased PK estimates attributed to the accuracy and precision of the analytical assay used.
- If concentration data comes from different sites and/or assays, verify the PK is not different between sites or assays.
- When a cross-validation is unavailable, graphical visualization of PK data (e.g., dose-normalized concentrations versus time) will help to identify agreement between results.
- If there is poor agreement between assays, a scaling factor that adjusts for assay differences should be introduced.
- If the drug to be modeled is also an endogenous compound, baseline endogenous concentrations should be accounted for. If the endogenous compound is known or believed to be time invariant, endogenous concentrations are modeled as a constant increment to exogenous drug. If it is known to vary over time, appropriate functions can be introduced to the model.
- If erroneous concentration or time data are identified or suspected, those data are replaced with correct data, if available, or removed from the analysis.
- When nominal times are recorded instead of actual times, the modeler should check preclinical or other clinical data to understand curvature in the PK profile, which determines the potential for bias in parameter estimates.

### Concentration data BLQ

BLQ data are commonly encountered during PK data analysis. In 2001, Beal described seven methods for handling BLQ observations (labeled M1–M7), several of which have since been widely adopted as standards of practice in pharmacometric modeling.<sup>20</sup> Despite widespread acceptance of the M3 method as generating the least biased results, it increases the time a model takes to converge and often inflates residual error unless BLQ observations are excluded from the computation of residual and weighted residuals (e.g., setting MDVRES = 1 [missing dependent variable for residual]). Furthermore, less information is available about when it is critical to use M3 (i.e., what percentage of BLQ observations requires this approach to avoid biased model parameters), the impact of both numbers of samples per subject or overall number of subjects in the data set, the impact of BLQ observations on parameter estimation for compounds that exhibit nonlinear PK, and the impact of BLQ observations on structural model selection (e.g., two-compartment structural model versus one-compartment structural model). In addition, the influence of BLQ data on structural model features, such as the relative magnitudes of CL and volume, has not been fully explored.

Several publications have examined various approaches to handling BLQ values, including deleting all BLQ observations (M1), the 'YLO' approach (M2), setting all BLQ observations to 0 (M7), setting all BLQ observations to half the LLOQ (LLOQ/2) (M5), setting the first BLQ observation in a subject to LLOQ/2 and removing the rest (M6), and likelihood-based use of M3 or M4 methods; these published results are summarized in Table 2.<sup>21–26</sup> Differences are evident between these published studies regarding the point at which ignoring/deleting BLQ observations was found to cause bias and impact parameter precision, which can be partly attributed to methodological differences, such as differing percentages of BLQ tested, use of the M5 versus M6 method for incorporating LLOQ/2, and approaches to simulating data sets and statistical methods used to evaluate the impact of the various methods employed for handling BLQ observations. These dissimilarities can prevent generalizable interpretability. We conducted a comprehensive, consistent analysis of the main BLQ data-handling methods (i.e., M1, M3, M5, and M6) to summarize current knowledge and address apparent discrepancies between previously published studies.

### Results of BLQ simulation and parameter estimation: M3 versus other handling methods

Across multiple scenarios (Supplementary Material: Handling of BLQ Data section), common BLQ data-handling methods were compared. In the simplest case, a densely

**TABLE 2** Summary of published literature evaluating methods for handling concentration data BLQ

Title	Methods	Conclusions
Hing et al. <sup>21</sup> NONMEM V	One-compartment model Rat study with one sample per animal BLQ: 10%–50% Methods tested: M1 and M7 and four substitution methods	M1 and M7 led to biased CL and IIV estimates Loss of precision for all methods occurred at BLQ >25% No impact of the number of animals used at each sampling point
Duval et al. <sup>26</sup> NONMEM VI	Two-compartment model Data sets based on (a) the ratio of the AUC of the distribution phase to the total AUC and (b) the ratio of the half-life of the distribution phase to the half-life of the elimination phase BLQ: 5%–50% Methods tested: M1 and M6	A bias on CL of >20% was observed with M1 at BLQ $\geq$ 20% No major trends were observed for $V_c$ and Q between M1 and M6 substitution IIV on CL is improved with M6, whereas the loss of information on IIV was observed for all other parameters, regardless of method
Keizer et al. <sup>22</sup> NONMEM VI	i.v. one-compartment model i.v. two-compartment model Oral one-compartment model BLQ: 10%, 20%, 40% Methods tested: M1, M6, and M3	<i>IV one-compartment model:</i> no bias was observed for any of the methods when the BLQ was <20%; considerable bias occurs with M1 and M6 at 40%; IIV was comparable at 10% and 20%, except for M3, which showed higher variation <i>IV two-compartment model:</i> bias in fixed parameters was observed after M1 and M6 at 10% BLQ, except for the estimation of V and IIV in V <i>One-compartment oral model:</i> at <20% BLQ, all methods except M1 provided reasonable and broadly similar performance for both fixed effects and IIV
Xu et al. <sup>23</sup> NONMEM VI	One-compartment model Two-compartment model BLQ: 1%, 2.5%, 7%, 10% Methods tested: M1 and M3	<i>One-compartment model:</i> the impact of ignoring BLQ <10% was minimal <i>Two-compartment model:</i> when the BLQ was <5%, M1 did not create bias in the fixed-effect parameters, whereas more pronounced bias in the estimates IIV was observed. The greatest impact was on $V_p$
Ahn et al. <sup>25</sup> NONMEM VI	Two-compartment model with first-order absorption BLQ: 10%–40% Methods tested: M1, M2, M3, and M4	M3 and M4 produced similar results without log transformation Parameter estimates were biased with M1, especially when the BLQ was 40% Clearance was more negatively biased as %BLQ increased. $V_p$ and Q were more positively biased The most accurate and precise estimates were obtained with M3
Bergstrand et al. <sup>24</sup> NONMEM VI	Model A: one-compartment model, transit compartments, BLQ in absorption phase Model B: two-compartment model BLQ: 10%–30% Methods tested: M1, M2, and M3	Model A: CL, $V_c$ , and IIV on CL and $V_c$ were not biased by presence of BLQ samples and similar for each method. $K_{at}$ , mean transit time and number of transit compartments were biased with M1 M3 generated the best performance Model B: M1 led to substantial bias in CL, Q, and $V_p$ . The M3 method was the least biased

Abbreviations: AUC, area under the curve; BLQ, below the limit of quantification; CL, clearance; IIV, interindividual variability; i.v., intravenous; Q, intercompartmental clearance; V, volume of distribution;  $V_c$ , volume of central compartment;  $V_p$ , volume of peripheral compartment.

sampled (five samples/subject), i.v. bolus one-compartment model with CL = 20 L/h and V = 70 L, the M3 method did not provide any added benefit for accurately estimating CL and its associated between-subject variability (BSV) for CL (BSVCL) compared with M1 across all data sets (5% - 50% BLQ). However, between 30% and 50% BLQ observations, M3 achieved less-biased estimates of V and BSV for V (BSVV). When V remained unchanged (V = 70 L) and CL was reduced (decreased from CL = 20 to CL = 10 L/h), the M3 method resulted in less-biased BSVV and BSVCL estimates at 30% or more BLQ. No differences were detected for estimates of CL and V between the M3 and M1 methods

between 5% and up to 30% BLQ. At more than 30% BLQ, the M3 produced significantly less-biased results compared with M1.

The M5 method produced biased results at scenarios 10%–50% BLQ. With M6, bias on CL and BSVCL was similar to M1 and M3 up to 20% BLQ, after which significant bias was observed. Similar trends were observed for V. A previous evaluation of the M6 method for a one-compartment model showed considerable bias and imprecision on CL at 40% BLQ, whereas no systematic bias was observed at 10% and 20%.<sup>22</sup> Of note, 30% BLQ was not assessed in this study. An additional study reported unpredictable results when

using M6 for BLQ observations in the elimination phase and M5 for BLQ observations in the absorption phase, sometimes inflating or sometimes reducing the bias.<sup>24</sup> Nonetheless, the M3 method was determined to be the best overall method. The size of the data set did not appear to influence the results, as similar trends were observed for all data set sizes of 25 to 400 subjects.

After simulations of an i.v. bolus one-compartment model with sparse sampling (two samples/subject), bias was observed for V between 20%–50% BLQ for M1, whereas bias on CL was similar between M1 and M3 for all BLQ percentages. However, in smaller data sets (i.e.,  $n = 25$  subjects), bias on CL was apparent between 20%–50% BLQ with M1. These results are in agreement with previous assessments of one-compartment models.<sup>21–23</sup> Collectively, these data indicate for a simple i.v., one-compartment model, either the M1 or the M3 methods are approximately equivalent with moderate or large data sets and when up to 30% of the observations are BLQ. However, the M3 method is superior when the portion of BLQ observations is >30% or with small data sets. Furthermore, use of the M6 method generated biased parameter estimates of CL or BSVCL at BLQ percentages as low as 20% and as low as 10% for M5, suggesting these data-handling methods should not be used for data sets with moderate to large percentages of BLQ.

The one-compartment extravascular dosing scenario was similar to the i.v. dosing scenario. Accuracy of CL estimates was similar between the M3 and M1 up to 30% BLQ. At 50% BLQ, however, M3 produced less-biased CL estimates. Not surprisingly, the time at which the BLQ observations occurred was important. When the percent of BLQ in the absorption phase was 5%, M1 generated similar results to M3. When the percentage of BLQ in the absorption phase was  $\geq 10\%$ , the use of M3 was superior to M1. Using RMSE as the primary criteria for comparison, M3 is recommended for less-biased estimates of V, BSVV, and BSVCL, regardless of overall percentage of BLQ. Using normalized RMSE (nRMSE) criteria (RMSE values normalized to 0% missing; see the Supplementary Material: Assessing Performance of Missing Data Handling Methods section), when the percent of BLQ was 5% or greater in the absorption phase and the overall BLQ percentage was  $\geq 20\%$ , M3 generated the least-biased results of the other evaluated methods. Previous oral one-compartment studies suggest that excluding observations produces bias on absorption parameters (e.g., mean transit time, number of transit compartments) at <30% BLQ.<sup>24</sup> Similar to our findings, Bergstrand and colleagues also observed accuracy and precision of CL estimates were comparable between M3 and M1, up to 30%.<sup>24</sup>

In an i.v. bolus two-compartment model with  $CL = 70$  L/h, bias in CL and V were similar between M1 and M3 at BLQ percentages up to 30%. However, estimates for

peripheral compartments (volume [V2] and intercompartmental clearance [Q]) were more sensitive to missing data. Using RMSE, M3 yielded less bias in these parameters at  $\geq 5\%$  BLQ. Using nRMSE, M3 was superior for V2 and Q when the BLQ percentage was 30% or more. BSVCL was less biased when using M3 when the BLQ was 20% or more. When simulating slower CL values ( $CL = 20$  L/h), the results were similar except for 50% BLQ, in which the M3 method generated less-biased estimates for CL than the other methods evaluated. The size of the data set had no influence on these results. Previous evaluations of two-compartment models have observed bias with M1 and M6 with BLQ percentages as low as 5% or 10%.<sup>22–26</sup> This is likely attributed to the loss of information on peripheral compartments resulting from omitting concentrations in the terminal phase, particularly in individuals with small peripheral volumes who would have declining concentrations below the limit of detection earlier than others.

Finally, nonlinear PK models were evaluated. M3 was superior at reducing bias on the maximum rate of metabolism ( $V_{max}$ ) and the Michaelis-Menten rate constant ( $K_m$ ) at all BLQ percentages. Volume estimates were stable throughout all percentages between M1 and M3, whereas M5 and M6 resulted in biased parameters at  $\geq 5\%$ . These results suggest that M3 should be used for achieving less-biased estimates of  $V_{max}$ ,  $K_m$ , and BSV for  $V_{max}$  and BSV for V at BLQ 5% or more.

In summary, the selection of an appropriate method to handle BLQ data is an important consideration in PK model development. Based on previous studies and our current analysis, when developing a simple i.v. bolus one-compartment model with rich sampling, the M3 method should be used in cases with approximately  $\geq 30\%$  BLQ observations. Otherwise, excluding observations when the BLQ percentage is <30% may not excessively bias CL and V estimates, although the M3 method should still be tested. The M3 method should be implemented with smaller data sets (50 subjects or fewer) with sparse sampling at approximately  $\geq 20\%$  BLQ observations. In an oral one-compartment model, the percentage of BLQ in the absorption phase is a determinant of using M3 versus excluding observations. If the percentage of BLQ in the absorption phase is >5%, M3 should be implemented for less-biased absorption parameters. Nonetheless, bias on CL is stable through 30% BLQ. When using a two-compartment model, CL and V tend to have similar bias up to 30% BLQ, regardless of the method used. However, ignoring all BLQ observations creates bias in V2 and Q parameters, and the M3 method would therefore be warranted for BLQ percentages greater than 5%. Finally, the M3 method is necessary for less-biased estimates of  $V_{max}$  and  $K_m$  in nonlinear models when BLQ values are present. Most studies report bias when replacing all BLQ observations with LLOQ/2.



If this method is implemented, replacing the first observation per subject with LLOQ/2 and discarding the remaining observations generates less bias than replacing all remaining BLQ observations with LLOQ/2. Finally, alternative methods for handling BLQ data can also be evaluated. For example, fixing the additive component to the variance of a uniform distribution of  $[0, \text{LLOQ}]$  in an additive-plus-proportional error model can potentially shorten run times and stabilize parameter estimation.<sup>27</sup>

Model misspecification is an additional concern with BLQ observations. A previous investigation of one-compartment i.v. model misspecification reported steadily increasing type I error rates (lower objective function value [OFV] for two-compartment models compared with one-compartment models at the  $\alpha$  0.05 and 0.01 levels) using M1 as the BLQ percentage increased. For example, when approximately 50% of the data was BLQ, 96% of 500 simulations had lower OFV for two-compartment models. When the M2 method was implemented, this percentage significantly decreased to nominal levels, leading to recommendations of M2 implementation when the BLQ > 10%.<sup>28</sup> For nonlinear models, 25% of the simulations had lower OFV when modeled as linear elimination at 10% BLQ. At 30% BLQ, approximately 50% of the simulations would be incorrectly classified as linear elimination. Similar to one-compartment versus two-compartment models, these percentages decreased to nominal levels after M3 implementation. Taken together, model misspecification and parameter estimate bias warrant the use of likelihood-based approaches to handling BLQ in most modeling scenarios.

### Summary of recommendations for handling BLQ data

- When the percentage of BLQ data is low (5% or less), all methods (M1, M3, M5, M6) should perform similarly, regardless of the model used and regardless of sampling density.
- If the percentage of BLQ data exceeds 5%, M5 should not be used.
- M6 generates biased CL, V, BSVCL, and BSVV at BLQ percentages greater than 20% for one-compartment models with rich sampling.
- In i.v. bolus one-compartment models, with 30% or less BLQ observations, the M3 method does not appear to provide a major advantage over M1 when rich sampling data are available.
- For i.v. bolus two-compartment models with rich sampling, the M3 method should be implemented at 5% or more BLQ for less biased estimates of V2 and Q. CL and V tend to have similar bias up to 30% BLQ between M3 and M1. M6 generates biased results.

- For drugs that exhibit nonlinear PK, the M3 method should be implemented at 5% or more BLQ.
- A decision tree for choosing the best BLQ handling method is shown in Figure 2.

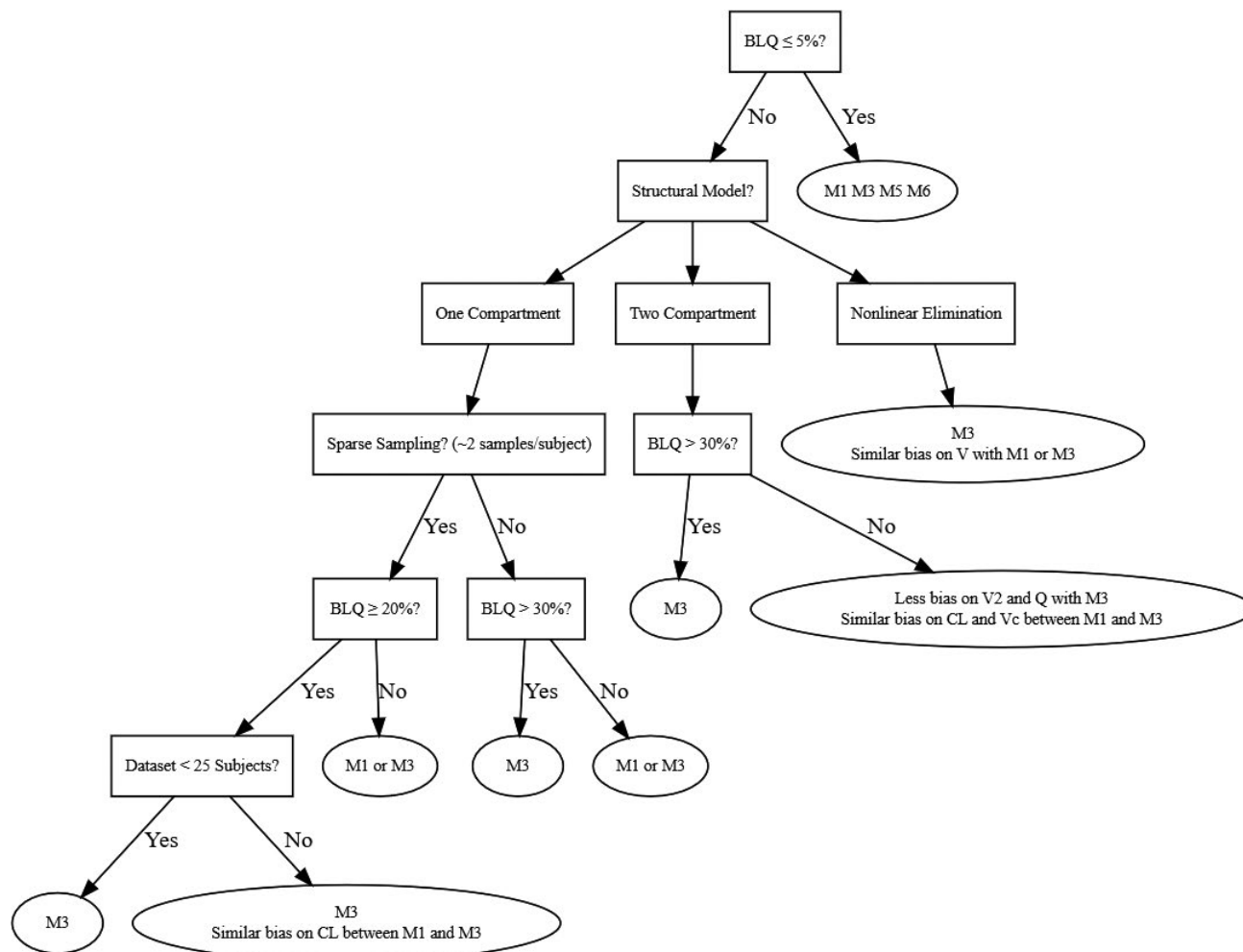
## DOSE DATA

### Missing dose amounts or dose times

Missing or erroneous dose information is commonly encountered during PK data analysis. During a clinical PK study, some dosing records, including administered dose, dose time, or duration, may be missing or recorded inaccurately. Often, dose and dose time are collected only at infrequent clinical visits. Common reasons for erroneous dosing information include incomplete or erroneous data capture, data loss during transition from the clinic into medical records or case report form, and so on. These errors are generally minimized during a study, although some study designs are inherently more prone to these types of problems. For example, in trials where study subjects self-report the times and/or doses administered, the rate of missing or erroneous data is relatively high.<sup>29</sup> Similarly, studies conducted within a general clinical setting, as opposed to a specialized research unit, tend to have higher rates of missed or misreported dosing data. In other cases, subjects may have started on study drug before enrollment in a particular trial.

Furthermore, identification of missing or erroneous dose records may be challenging and will depend on the data structure, amount of missing or erroneous data, PK properties of the drug, and so on. When possible, merging smaller studies with rich sampling into larger studies with sparse sampling can help to identify missing or erroneous data (concentration vs. time and dosing data) in the larger study. Comparing model results from densely sampled studies with parameters estimated from suspect data can also be insightful.

The choice of method for handling missing dose information can influence the quality of the PK data analysis. Ignoring missing records is never acceptable because observed drug concentrations are a direct result of dose administration (whether or not that dose is recorded) combined with drug distribution and elimination. If missing dose information is ignored, the analysis will inappropriately associate observed concentrations to distribution and/or clearance processes. Nedelman previously demonstrated the utility and flexibility of NONMEM in assessing the reliability of dose data, providing an example of how one might assess accuracy of observed or assumed dose times, the influence these assumptions may have on parameter estimates, and the stability of the model.<sup>30</sup> The next several paragraphs highlight approaches, some simple and others more complex, that can be applied to address missing or unreliable dose data.



**FIGURE 2** Decision tree for handling of concentration data below the limit of quantification (BLQ). CL, clearance; Q, intercompartmental clearance; V, volume of distribution;  $V_c$ , volume of central compartment;  $V_2$ , peripheral volume

One method for treating missing dosing records, sometimes referred as the omitting method (OM), is to exclude the “influenced” PK observations from analysis. This method tries to mask the problem by only analyzing the intact subset of data. Although not a preferred approach, this method may still be feasible if the remaining data carry sufficient information about the PK behavior under study. Empirical examples include when only a few dosing records are missing while most records are intact within a given study subject or when a dosing record is missing during later stages of a multidose study after steady-state concentrations are achieved, provided the duration of missing dose information is sufficiently short to not impact steady-state concentrations. Studies of self-administered monoclonal antibodies often fall into this group, as patients administer most doses, but may miss a few doses. Under these conditions, it may be acceptable to exclude the “influenced” observations from analysis. However, it is sometimes difficult to precisely determine which portions of the data were “influenced” and the extent of influence. There are many situations when this method might be

used with caution. For example, when PK data are only available after a single dosing event or when only sparse PK data are available and prior to when steady-state PK is achieved.

Several other methods for handling missing or erroneous dosing information have been reported.<sup>31–37</sup> The prescribed dose method (PDM) assumes all prescribed doses were administered per prescription/study design.<sup>31</sup> With this assumption, one simply fills in the missing or erroneous records and proceeds with analysis as if dosing data were intact. This is the approach often taken in larger, late-phase studies, where dose level and time are intentionally not captured, and consequently they are assumed (e.g., using ADDL in NONMEM). Notably, in these cases the data are not considered missing since they were intentionally not captured. While straightforward, the PDM can be inaccurate, which inherently introduces error into the analysis. Capturing the dose information prior to a PK sample collection can prevent inaccuracies. In some large clinical trials, patient compliance data are collected (pill counts), which can be used to adjust the administered amount over intervals of time (a comprehensive

summary of compliance [or adherence], its measurement, modeling adherence, and the impact of improper assumptions about adherence in the modeling and simulation processes was previously published by Kenna et al.<sup>38</sup>). Several handling methods exist where the shortcomings of OM and PDM can be minimized by estimating the missing or erroneous dosing information based on observed residual drug amounts. These methods are discussed in the subsequent paragraphs.

The concentration-time method (CTM), which is similar to PDM, also assumes per protocol dose amount and intervals.<sup>32</sup> However, CTM specifically addresses the issue with incorrect shifting of dosing time on individual levels. The individual time shift is parameterized and used to calculate the residual drug amount. For example, for a drug with repetitive oral administrations and empirical one-compartmental PK model, the residual amount CT is expressed as Equation (7):

$$CT = \frac{\text{Dose} \times ka}{V \times (ka - k)} \times \left( e^{-k \times (DT+TP)} \times \frac{1 - e^{-ND \times k \times \text{tau}}}{1 - e^{-k \times \text{tau}}} - e^{-ka \times (DT+TP)} \times \frac{1 - e^{-ND \times ka \times \text{tau}}}{1 - e^{-ka \times \text{tau}}} \right) \quad (7)$$

where Dose is the dose amount, V is the volume of distribution, ka is the absorption rate constant, k is the elimination rate, ND is the number of doses, tau is the dosing interval, TP is the time passed dose, and DT is incorrect dosing time shift. DT is parameterized the same way as other PK parameters in the form of  $DT_{\text{pop}} \times e^{\eta_{DT}}$ . The residual amount is then accounted when fitting current PK observations by incorporating it to the observed concentrations (Equation 8):

$$Y = (F + CT) \times \varepsilon_{\text{prop}} + \varepsilon_{\text{add}} \quad (8)$$

where Y is the observed concentration, F is the predicted concentration, CT is the residual amount correction from missing dosing history,  $\varepsilon_{\text{prop}}$  is the proportional error, and  $\varepsilon_{\text{add}}$  is the additive error.

The extrapolation-subtraction method (ESM) uses a “subtraction curve” estimated from the terminal-elimination phase of current data.<sup>31</sup> The method adjusts the influenced data by subtracting the residual amount over time.

The missing dose method (MDM) does not amend the data set<sup>31</sup> though, it parameterizes the residual drug amount as an individual PK parameter (A0; Equation 9):

$$A0 = A0_{\text{pop}} \times e^{\eta_{A0}} \quad (9)$$

where A0 is the individual residual drug amount,  $A0_{\text{pop}}$  is the population residual amount, and  $\eta_{A0}$  is the individual variability. All drug concentrations after the missing records are treated as results from both A0 and available dosing events. A modification of MDM, the missing dose mixture method (MDMM), combines MDM and PDM via a mixture model approach. In

the original article by Soy et al.<sup>31</sup>, the authors presented the MDM and MDMM methods and compared their performance with PDM and ESM in handling missing dosing histories. The MDM method was concluded to be superior overall in the demonstrated examples. The method was later provided with an implementation by Gupta et al.<sup>33</sup> (Equation 10):

$$C = (\text{SDF} + C0 \times e^{-k_e \times t}) \times \varepsilon \quad (10)$$

where C is the observed concentration, SDF denotes the concentration from last (correct) dose (i.e., concentration without correction, similar to F from the CTM method), C0 is the residual A0 in the form of concentration,  $k_e$  is the elimination rate, t is time after dose, and  $\varepsilon$  is residual error. The MDM method was reportedly applied in practical analysis by Friberg et al.<sup>34</sup>

CM methods<sup>37</sup> (CM1 and CM2) provide two different angles to attack the issues from erroneous dosing information (classified as “noncompliance history”). CM1 treats the non-compliance dosing history as an additional source of RUV and offers two submethods to label records that may carry it. One submethod (M3) put an interindividual variability (IIV) term  $\eta_{\sigma}$  ( $\eta_{\sigma}$ ) on RUV and estimates the individual  $\eta_{i,\sigma}$ . Please note that the Gibiansky, et al.,<sup>37</sup> publication uses M1 through M6 as labels for 6 different submethods for handling erroneous dose data. Here we retain the same M1–M6 labeling as presented by Gibiansky, et al., though we caution the reader to not confuse these submethods for handling erroneous dose data with the M1–M7 methods for handling BLQ data, as described in the *Concentration data BLQ* section of this tutorial. The model with proportional RUV is introduced as follows (Equation 11):

$$C_{ij} = C_{ij} + \sigma_{ij} \times C_{ij} \times \exp(\eta_{i,\sigma}) \quad (11)$$

where  $C_{ij}$  is the jth observed concentration in individual i;  $C_{ij}$  is the jth model predicted value in individual i;  $\sigma_{ij}$  is the residual random variable; and  $\eta_{i,\sigma}$  is the individual value of random effect on RUV, which is assumed to be normally distributed with mean zero and standard deviation  $\omega$ . Records with post hoc estimated  $>0$  are labeled as potential noncompliant records and can be excluded from analysis to reduce estimation bias (M5).

The other submethod of CM1 (M4) models the RUV with a mixture model structure, assuming noncompliant dosing history leads to concentration records with RUV following a different distribution than do normal records. The model with proportional RUV is introduced as follows (Equation 12):

$$C_{ij} = C_{ij} + \sigma_{ij} \times C_{ij} \quad (12)$$

where  $C_{ij}$  is the jth observed concentration in individual i,  $C_{ij}$  is the jth model predicted value in individual i, and  $\sigma_{ij}$  is the

residual random variable.  $\sigma_{ij}$  follows a normal distribution with mean zero, but a variance to be estimated individually for each group of patients (i.e., compliance vs. noncompliance). The mixture model structure is needed to implement this method. Records associated with the more variable RUVs are labeled as potential noncompliant records and can be excluded from analysis to reduce estimation bias (M6).

Both combinations of M3/M5 and M4/M6 were reported to reduce the estimation bias from noncompliant dosing history, with similar performance.

The CM2 method works on the dosing records instead of concentration records. First, one has to know which dosing records can be attributed as noncompliant. This can come from knowledge of the data condition or be deduced by applying either M3 or M4 from CM1. Once this is done, the noncompliant dosing records can be included in analysis, but with a flag to label them in the data set during model run. The labeled dosing records are implemented with a relative bioavailability term to distinguish them from compliant dosing. The concentration at time  $t$  is then formulated as Equation (13):

$$C(t) = C(F_{NC}) + C \quad (13)$$

where  $C(F_{NC})$  is the concentration from noncompliant dosing,  $F_{NC}$  is the relative bioavailability to be estimated, and  $C$  is the concentration with normal dosing records. It is worth noting that to use the CM2 method, concentration records associated with noncompliant dosing need to be excluded from the analysis.

Recently, Wang et al. published another implementation of the MDM method and expanded the methodology (MDM2) to handle different scenarios of missing dosing information.<sup>35</sup> First,  $A_0$  was parameterized as in MDM to represent the residual drug amount from missing dosing histories. Then, a mixture model structure is applied to put patients into different groups so that each group of patients will be identified with their specific  $A_{0pop}$  values and  $\eta_{A_0}$  distributions. This approach addresses potential issues when the distribution of  $A_0$  cannot be described with a single log-normal distribution. The authors also published a new method, called the compartment initialization method (CIM), which addressed the residual amount using a different approach. Rather than estimating the individual residual amount,  $A_0$ , the CIM method “back calculates”  $A_0$  using the most recent concentration record and current individual estimate of volume  $V$ . Each individual  $A_0$  is then used to initialize the corresponding compartment if a missing dosing record precedes current observations. The CIM method demonstrated equal performance to MDM for handling missing dosing histories, although CIM outperformed all other tested methods in other missing dosing record scenarios.

Besides all aforementioned methods, there was another publication that looked into the uncertain dosing information

from a Bayesian perspective.<sup>36</sup> Mu et al. examined the potential benefits of adapting a hierarchical Bayesian approach when part of the dosing history is uncertain. The study adopted a simplistic but representative scenario: multiple doses were given at known and equal intervals, with two doses (the first and the middle doses in the more realistic setting) at the known level, but all others at uncertain levels. For each of the uncertain doses, independently, they have certain probabilities of being totally missed, taken as the correct dose, or taken as twice the correct dose. With a simple one-compartment PK model with bolus dosing, the model predicted concentration is expressed in a recursive form as the following (Equation 14):

$$f_{ijk} = \frac{D + r_{ik} \times D}{V_{ij}} + f_{ij,k-1} \times \exp\left(-\frac{CL_{ij}}{V_{ij}} \times (t_k - t_{k-1})\right) \quad (14)$$

$$k = 1, 2, \dots; f_{ij0} = 0$$

where  $f_{ijk}$  is the observed concentration for  $i$ th individual at  $j$ th occasion and  $k$ th timepoint;  $D$  is the dose given at certain dosing records (value 0 otherwise);  $CL_{ij}$ ,  $V_{ij}$  are the clearance and volume of distribution of  $i$ th individual at  $j$ th occasion, respectively;  $t_k$  is the time after dose at the  $k$ th timepoint; and  $r_{ik}$  is the compliance factor, which serves as a switch to specify the amount of drug for an uncertain dose, 0, 1, or 2 times the nominal dose.

Incorporation of the compliance factor  $r_{ik}$  was shown to improve estimation bias on some parameters, whereas others were shown to be insensitive to this technique. It was speculated that this may be attributed to the specific sampling design that was used. It is worth noting that to find the appropriate model for  $r_{ik}$  to approximate distribution of the simulated uncertain dose levels, for example, 0:1:2 with 60%:35%:5%, the authors had to go through a process of trial and error. Eventually a function of two binomial distributions was found to provide reasonable estimates of compliance factors. However, because of the relative simplicity of tested scenarios, speculated sampling dependent results, and the use of trial and error, it is difficult to predict the generalizability of this method in other scenarios.

In summary, when dosing records are missing, all methods to replace them will inevitably adopt one or more assumptions. The PDM method and parts of the CTM method (regarding the prescribed dose amount and interval) apply direct and consistent assumptions based on the clinical protocol. It is important to recognize that those assumptions may not be accurate for all instances of the missing data. Alternative methods, including ESM, MDM, MDMM, MDM2, CM1/CM2, and CIM, also allow for estimation and adjustment of the missing dosing information based on the observed PK data. Each method has its strengths and weaknesses, depending on the specific data situation. Next we summarize our recommendations for implementing these methods based on the specific missing data scenarios or assumptions.

## Summary of recommendations for handling missing or erroneous dose data

- PDM assumes dose times and/or dose amounts do not deviate from those specified in the study and/or that deviations in times and/or amounts are insignificant. PDM only works well when these assumptions are generally robust.
- CTM applies when the dosing amount is known and only dosing time is missing.
- All other methods for handling missing dose information assume the missing data deviate from the study design and that these deviations may significantly influence the PK analysis.
- When PK observations influenced by missing dose data constitute only a small portion of all available PK data for an individual, OM is likely to yield acceptable PK parameter estimates.
- When some of the dose level information is missing within a full dosing history, use of either the MDM or CIM method will provide the best opportunity for yielding accurate PK parameter estimates. When dosing records are missing more randomly throughout the data, CIM and the expanded MDM (MDM2) methods are more appropriate.

## COVARIATE DATA

As with other data types, missing, erroneous, or outlier covariate data can negatively impact PK data analysis. Data QA strategies and exploratory data analysis used for concentration versus time data, such as descriptive statistics and generation/review of various plots (scatterplots of covariates vs. subject or time, histograms of covariates, etc.) can be useful for covariate data as well,<sup>1</sup> and these should be employed prior to modeling and simulation efforts. In this section, published strategies for addressing covariate data that have been confirmed to be missing or are questionable are reviewed. The impact of missing or erroneous covariate data is also evaluated using simulation and estimation to systematically compare the three most common strategies for handling missing covariate data: the complete case (CC; elimination of the entire patient from the data set), imputing to a reference value (IRV), and joint (JM; continuous) or mixture modeling (MM; categorical) to estimate the covariate value. The results of these analyses are discussed next.

### Missing covariates

Missing covariate data are usually obvious, although they may sometimes be difficult to identify with zero (0) values that may represent true zeros versus missing covariate data,

such as can occur with numerical assignments for patient sex or other categorical covariates, making assignment of a coded value for a missing covariate important. After covariate data have been confirmed to be missing and strategies to locate the missing data have been exhausted, appropriate methods for addressing the missing covariate data should be employed.

Several methods for handling missing covariate data in PPK analysis have been proposed in the literature. A population model of topotecan PK used the JM method to impute missing values for body weight (~20% missing) using observed values for body surface area, sex, and CrCL.<sup>39</sup> This approach was successful at producing unbiased estimates of weight for utilization in the covariate model, as demonstrated by a linear regression of the observed versus estimated weights from the joint model (the data were uniformly distributed about the line of unity, and the regression coefficient was 0.88). Utilization of the JM method in this study allowed for an unbiased assessment of the effect of body weight on topotecan PK in the published data set, which led to a 5% reduction in interpatient variability on CL. One of the most comprehensive studies published recently evaluated six methods for handling missing data for the categorical covariate sex, with up to 50% missing data. These methods included CC, single imputation to the mode value (SI<sub>mode</sub>), single imputation using weight to predict the missing value (SI<sub>WT</sub>), multiple imputation using both weight and estimated steady-state concentration (MI), prediction of the missing covariate using a maximum likelihood mixture model developed from the individuals without missing covariate data (MOD), and a version of the MOD mixture model where the proportion of missing males and females was estimated as a fixed effect parameter in the model (EST).<sup>40</sup> This study also evaluated these missing data handling methods in the contexts of data missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). The MI, MOD, and EST methods produced unbiased and precise estimates under the MCAR and MAR scenarios, although only the EST method performed well with the MNAR scenario. Although the SI<sub>mode</sub> and SI<sub>WT</sub> methods were already known to produce more bias than the other methods evaluated, they along with the CC method remain among the most common methods employed in pharmacometric analysis. A subsequent study by the same group also evaluated the impact of  $\eta$  shrinkage on bias and precision for the MI method. This study found that when shrinkage is high (~40%), the estimates remain unbiased although they become less precise.<sup>41</sup> About the same time, a separate group evaluated the CC method, estimation of a separate parameter for the effect of the missing covariate (EXTRA), and three mixture models fixing the likelihood of missing binary categorical data (sex) to the observed

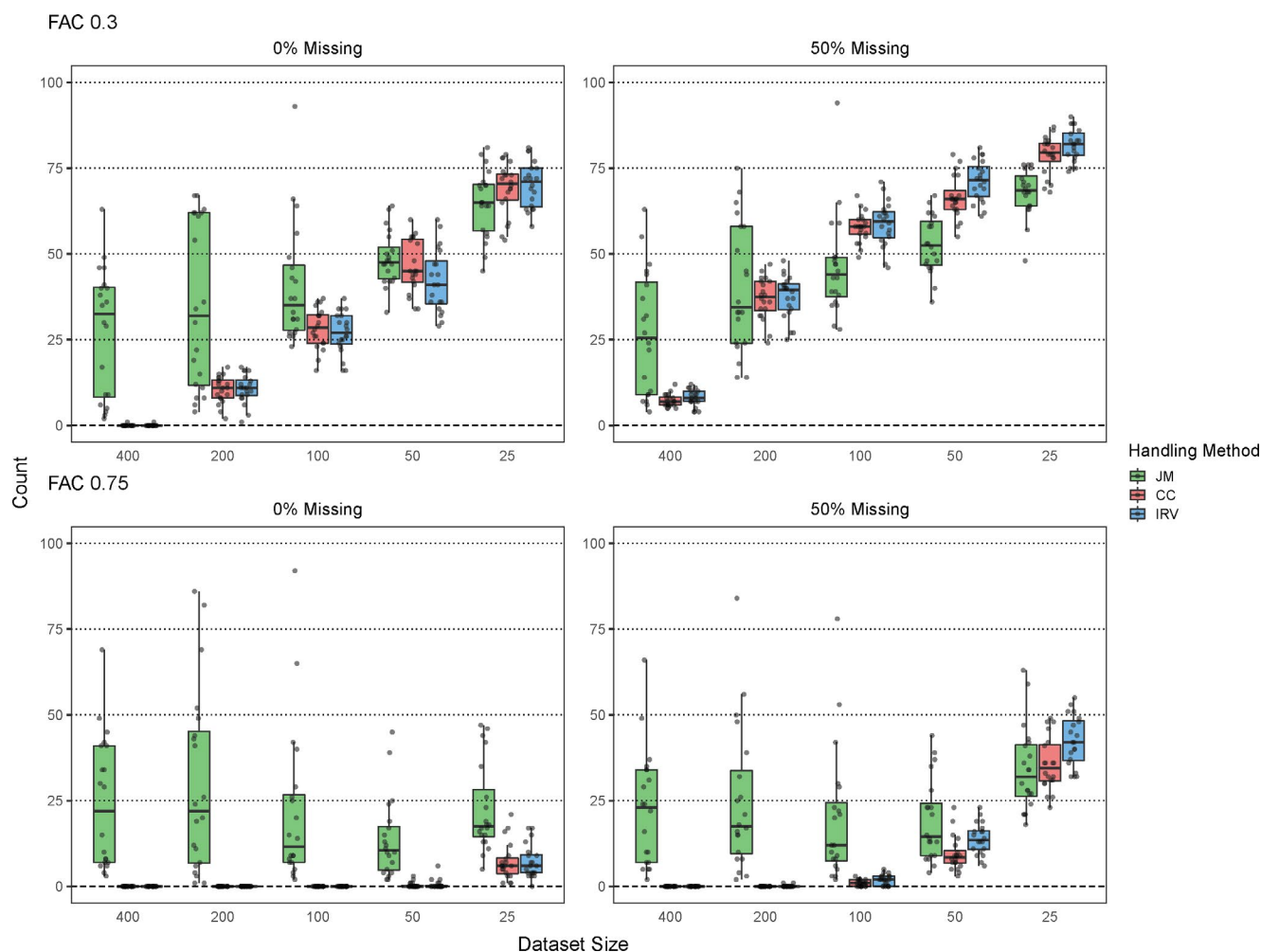
binary proportion, estimating the proportion using only PK data, and estimating the proportion using a mixture of observed proportion and PK data.<sup>42</sup> These authors concluded that all methods performed similarly compared with each other with both low and high levels of missingness (10% or 50%) and with data having balanced and unbalanced levels of the binary categories (1:1 and 3:1). The CC and EXTRA methods were actually less biased than the mixture methods for the MNAR data.

When individual patient data are partially missing, the last observation carried forward (LOCF) or last observation carried backward (LOCB) approaches are commonly used. The default behavior in NONMEM is LOCB for time-varying covariates. Although this method performs reasonably well, in general it is somewhat less intuitive than LOCF, and it can be problematic in patient populations where the covariate values are changing rapidly (e.g., body weight in young children or infants). The LOCB approach in NONMEM can also result in different

simulation profiles as compared with the LOCF method, as demonstrated by Upton and Mould.<sup>43</sup> This study showed that simulations from NONMEM were dependent on the value of TIME in the row immediately prior to the record where the covariate value changed. The authors showed that a “dummy” row inserted in the data set with the same TIME as the record where the covariate value changes is a useful strategy to implement LOCF behavior in NONMEM.

## Simulations to address gaps in the literature

Detailed methods for the covariate simulations along with summaries of the results can be found in the Supplementary Material. Here we describe the key findings of the simulations when applying CC, IRV, and JM methods for WT on CL and CC and the MM method for SEX on CL across four scenarios for CL and V within data sets of various sizes,



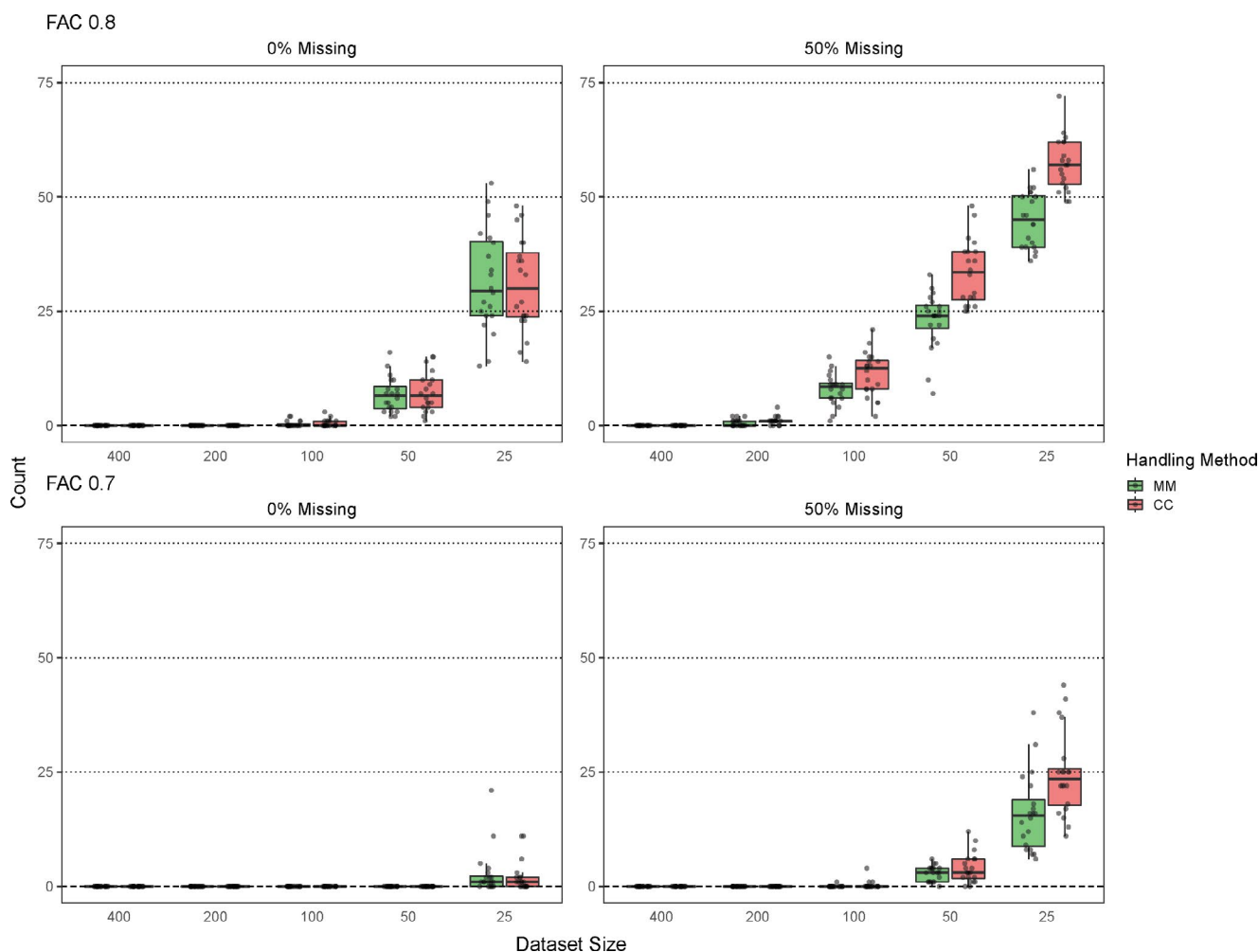
**FIGURE 3** Comparing the number of runs that failed to identify the covariate effect between handling methods. These data are summarized from all simulations involving a single intravenous or extravascular dose with WT as the covariate on CL. CC, complete case; CL, clearance; FAC, covariate effect size; JM, joint modeling; IRV, imputation to a reference value; WT, weight

various strengths of covariate effect, and various levels of missing data.

Many of the trends observed during simulation were as expected. In general, the population parameters for CL and V were always estimated well, irrespective of the amount of missing covariate data. Overall, the smaller the data set, and the smaller the effect of the covariate, the higher the RSEs, RMEs, and RMSEs of the covariate effect parameter, FAC, and the lower the likelihood that the covariate effect would be found to be significant. Weak trends were also observed with other estimated parameters being less accurate or less precise with smaller data sets and with greater amounts of missing covariate data, as expected, although these effects were very small compared with those affecting the FAC parameter. Accordingly, the FAC estimate (including its related RSEs, RMEs, and RMSEs) was the model parameter most affected by missing data. In some scenarios different handling methods performed equally, whereas in other scenarios there was an advantage

with one or two of the three methods. The key factors influencing the accuracy and precision of the estimated parameters, the ability of the model to detect the covariate effect (i.e., significance of the likelihood ratio test with  $p$  value of 0.05 based on change in OFV), and the ability of the model to converge with a successful covariance step across the various scenarios are discussed in the following paragraphs. A summary of recommendations for the preferred methods of handling missing covariate data under various circumstances based on the literature review and the results of the simulation exercises are also provided.

In most scenarios, the JM and MM methods slightly outperformed the IRV or CC methods in correctly identifying a significant FAC effect (Figures 3 and 4, Tables S1–S6, and Figures S2–S5). In this regard, the relative impact of these preferred handling methods was more pronounced in the smaller data sets (200 subjects or less), with increasing amounts of missing data (20%–50%) and for the weaker covariate effects (power of 0.5 or less for WT, difference in

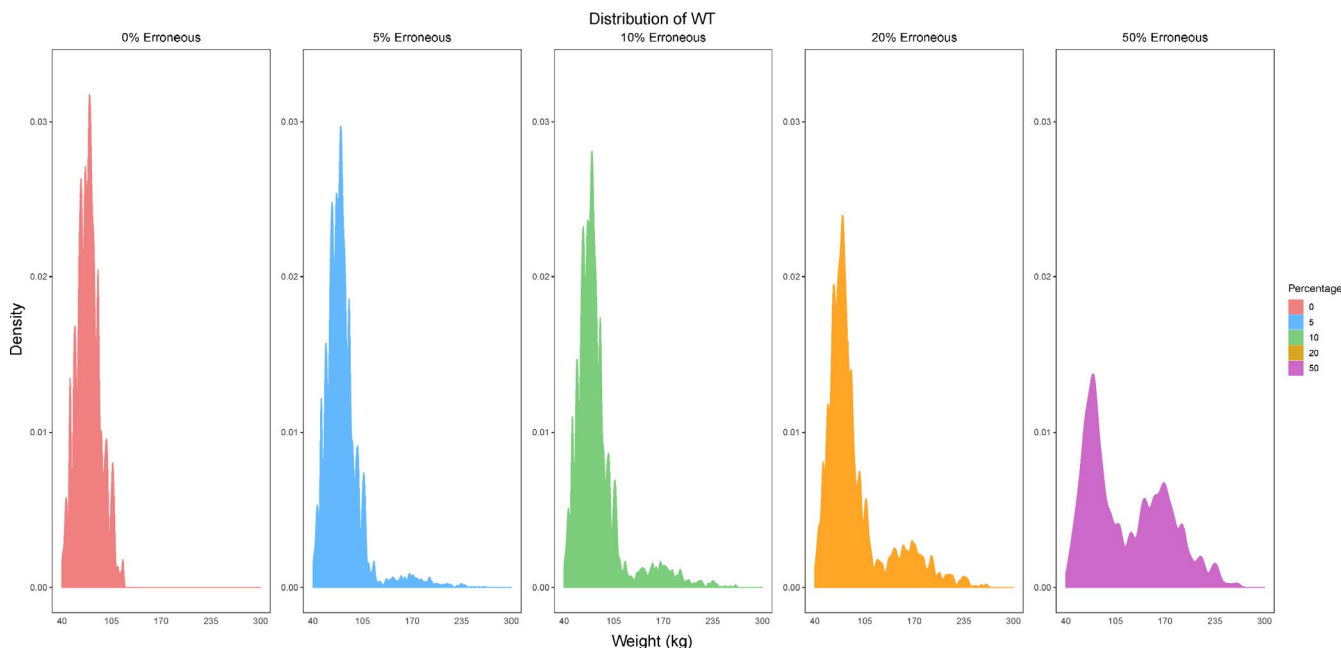


**FIGURE 4** Comparing the number of runs that failed to identify the covariate effect between handling methods. These data are summarized from all simulations involving a single intravenous or extravascular dose with SEX as the covariate on CL. CC, complete case; CL, clearance; FAC, covariate effect size; MM, mixture modeling; WT, weight

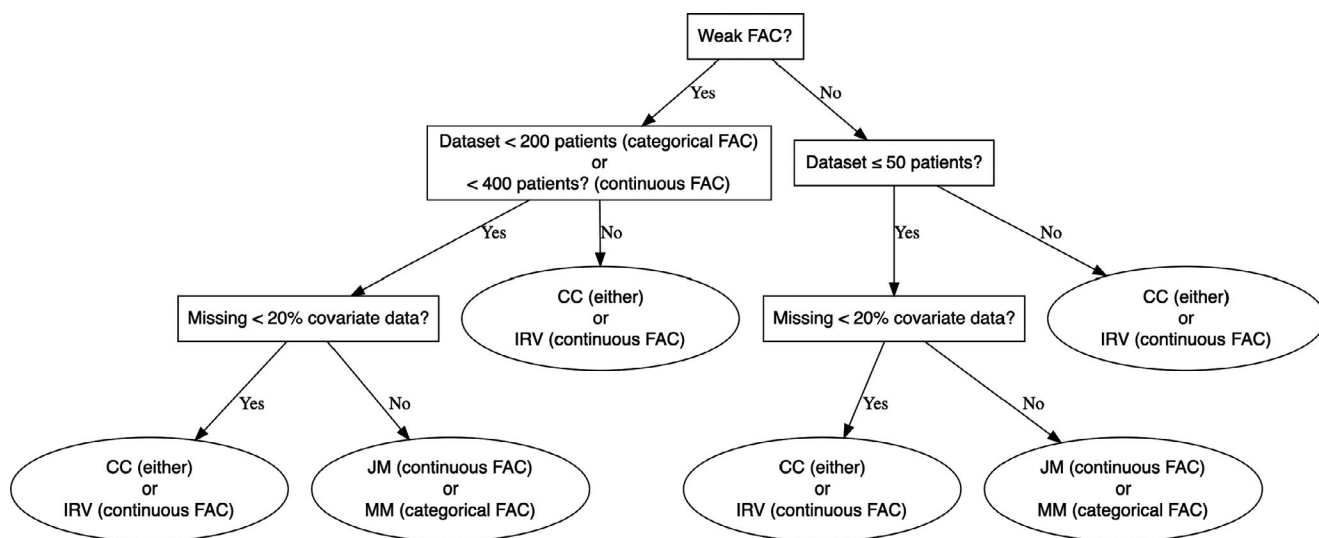
population value of 30% or less for SEX). The JM, IRV, and MM methods also slightly outperformed the CC method in terms of the RMSE of BSVCL, BSVV, and BSVKA with increasing amounts of missing covariate data (Tables S1–S4 and Figures S6–S9). However, there was not a clear preference in handling method with respect to any of the other estimated model parameters.

In some instances, the IRV and/or CC methods were preferable to the JM or MM methods. For example, the JM and MM methods failed to complete the covariance step more frequently than the other methods, especially in the scenarios

involving fewer patients (200 subjects or fewer), sparser sampling (4 vs. 7 timepoints), weaker covariate effects, and with additional parameter estimates (full OMEGA block, etc.; Tables S1–S4). Similarly, the use of the JM or MM methods were often associated with longer run times, although the increases in run times were modest for all of our tested scenarios. In cases where the data are of higher quality (larger data sets, 20% or less missing data) the CC or IRV methods would be preferable to either the JM or MM methods as there is minimal impact to correctly identifying a significant covariate effect.



**FIGURE 5** Diagnostic density plots demonstrating the manifestation of erroneous covariate data (units switched from kg to lb) in the distribution of WT. The results are stratified by the percentage of erroneous covariate data. WT, weight



**FIGURE 6** Decision tree for handling missing covariate data. CC, complete case; FAC, covariate effect size; IRV, imputing to a reference value; JM, joint modeling MM, mixture modeling



## Erroneous covariate data—the units switch

Erroneous covariate data may also be encountered within a data set, and it may not be completely obvious that the erroneous data are present. An example of erroneous data that are encountered periodically is the mixing of units within a single covariate, such as weight (kg vs. lb). Because the ranges for weight in kg and weight in lb can overlap within a typical data set, it may not be obvious that some data are reported in lb, for example, when the expected unit is kg (Figure 5).

To evaluate the impact of unit switching on covariate analysis, we explored the impact of switching kg units to lb in 5%, 10%, 20%, and 50% of the population across the five different data set sizes of 400, 200, 100, 50, and 25 patients. On average, even a 5% switch within the 400-patient data sets increased the absolute value of the RME FAC to roughly 40% (Table S5 and S6 and Figures S10 and S11). As expected, increasing amounts of unit switching (10%, 20%, 50%) caused accuracy to further decline (Tables S5 and S6 and Figures S10 and S11). The difference between the estimated and true FAC level was approximately the same between the high and low FAC levels (i.e., FAC = 0.75 and FAC = 0.3, respectively) when the units were switched (Tables S5 and S6 and Figures S10 and S11). A unit switch in weight from kg to lb significantly decreased the ability to identify WT as a covariate for both the low and high FAC levels (FAC = 0.3 and FAC = 0.75).

## Summary of recommendations for handling missing/erroneous covariate data

- The JM and MM methods typically performed equally well or better than the CC or IRV methods for identifying the covariate effect for WT and SEX, respectively. Therefore, JM or MM methods would generally be a safe and conservative approach in all scenarios for identifying even weak continuous or categorical covariate effects, respectively.
- For large data sets (approximately 200 or more subjects), and with relatively minor amounts of missing covariate data (up to approximately 20% of the subjects have a missing covariate), all of the handling methods tested worked approximately equally in terms of identifying the covariate effect and estimating model parameter values accurately and precisely. Therefore, since the CC and IRV methods are generally easier to apply, these might be preferred over the JM or MM approaches in these scenarios.
- Although the JM and MM approaches still outperformed the other methods for detecting the covariate effect, a higher failure rate for the covariance step was observed with these methods compared with the CC and IRV methods.

- Caution should be exercised when there are apparently erroneous covariate data present within a data set, and the use of diagnostic plots such as histograms and/or density plots may be helpful to detect a units switch for continuous covariates. A decision tree for selecting the best handling method for missing or erroneous covariate data is displayed in Figure 6.

## ACKNOWLEDGMENTS

The authors thank the expert reviewers along with all of the authors of the previous work that we referenced. These contributions helped to refine this tutorial.

## CONFLICT OF INTEREST

The authors declared no competing interests for this work.

## REFERENCES

1. Bonate PL, Strougo A, Desai A, et al. Guidelines for the quality control of population pharmacokinetic-pharmacodynamic analyses: an industry perspective. *AAPS J.* 2012;14(4):749-758.
2. Data Quality Management in Clinical Research. [https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical\\_conduct/data\\_quality\\_management-2015\\_05\\_15.pdf](https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical_conduct/data_quality_management-2015_05_15.pdf). Accessed April 2, 2021.
3. Jamei M, Marciniak S, Feng K, Barnett A, Tucker G, Rostami-Hodjegan A. The Simcyp population-based ADME simulator. *Expert Opin Drug Metab Toxicol.* 2009;5(2):211-223.
4. Graves DA, Locke CS Jr, Muir KT, Miller RP. The influence of assay variability on pharmacokinetic parameter estimation. *J Pharmacokinet Biopharm.* 1989;17(5):571-592.
5. Guideline on bioanalytical method validation. In: Agency EM ed.; 2012:1–23. [https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-bioanalytical-method-validation\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-bioanalytical-method-validation_en.pdf). Accessed April 2, 2021.
6. Bioanalytical method validation guidance for industry. In: Services USDoHaH ed; 2018. <https://www.fda.gov/media/70858/download>. Accessed April 2, 2021.
7. Mordenti J, Cavagnaro JA, Green JD. Design of biological equivalence programs for therapeutic biotechnology products in clinical development: a perspective. *Pharm Res.* 1996;13(10):1427-1437.
8. Gabrielsson JWD. *Pharmacokinetic and Pharmacodynamics Data Analysis: Concepts and Applications*. 4th ed. Stockholm: Swedish Pharmaceutical Press; 2000.
9. Schindel F. Consideration of endogenous backgrounds in pharmacokinetic analyses: a simulation study. *Eur J Clin Pharmacol.* 2000;56(9–10):685-688.
10. Rose TH, Röshammar D, Erichsen L, Grundemar L, Ottesen JT. Characterisation of population pharmacokinetics and endogenous follicle-stimulating hormone (FSH) levels after multiple dosing of a recombinant human FSH (FE 999049) in healthy women. *Drugs R&D.* 2016;16(2):165-172.
11. ICON Development Solutions. *NONMEM Users Guide. Part V Introductory Guide*. Ellicott City, MD: ICON Development Solutions; 2011.
12. Duong JK, Veal GJ, Nath CE, et al. Population pharmacokinetics of carboplatin, etoposide and melphalan in children: a re-evaluation

- of paediatric dosing formulas for carboplatin in patients with normal or mild impairment of renal function. *Br J Clin Pharmacol*. 2019;85(1):136-146.
13. Lippi G, Chance JJ, Church S, et al. Preanalytical quality improvement: from dream to reality. *Clin Chem Lab Med*. 2011;49(7):1113-1126.
  14. Carroll RJ, Carroll RJ. *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC; 2006.
  15. Santalo O, Baig U, Poulakos M, Brown D. Early vancomycin concentrations and the applications of a pharmacokinetic extrapolation method to recognize sub-therapeutic outcomes. *Pharmacy (Basel)*. 2016;4(4):1-9.
  16. Choi L, Crainiceanu CM, Caffo BS. Practical recommendations for population PK studies with sampling time errors. *Eur J Clin Pharmacol*. 2013;69(12):2055-2064.
  17. 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):207-246.
  18. Wang N, Davidian M. A note on covariate measurement error in nonlinear mixed effects models. *Biometrika*. 1996;83(4):801-812.
  19. ICON Development Solutions. *NONMEM Users Guide. Part VIII Help Guide*. Ellicott City, MD: ICON Development Solutions; 2006.
  20. Beal SL. Ways to fit a PK model with some data below the quantification limit. *J Pharmacokinet Pharmacodyn*. 2001;28(5):481-504.
  21. Hing JP, Woolfrey SG, Greenslade D, Wright PM. Analysis of toxicokinetic data using NONMEM: impact of quantification limit and replacement strategies for censored data. *J Pharmacokinet Pharmacodyn*. 2001;28(5):465-479.
  22. Keizer RJ, Jansen RS, Rosing H, et al. Incorporation of concentration data below the limit of quantification in population pharmacokinetic analyses. *Pharmacol Res Perspect*. 2015;3(2):e00131.
  23. Xu XS, Dunne A, Kimko H, Nandy P, Vermeulen A. Impact of low percentage of data below the quantification limit on parameter estimates of pharmacokinetic models. *J Pharmacokinet Pharmacodyn*. 2011;38(4):423-432.
  24. Bergstrand M, Karlsson MO. Handling data below the limit of quantification in mixed effect models. *AAPS J*. 2009;11(2):371-380.
  25. Ahn JE, Karlsson MO, Dunne A, Ludden TM. Likelihood based approaches to handling data below the quantification limit using NONMEM VI. *J Pharmacokinet Pharmacodyn*. 2008;35(4):401-421.
  26. Duval V, Karlsson MO. Impact of omission or replacement of data below the limit of quantification on parameter estimates in a two-compartment model. *Pharm Res*. 2002;19(12):1835-1840.
  27. Hu C, Randazzo B, Sharma A, et al. Improvement in latent variable indirect response modeling of multiple categorical clinical endpoints: application to modeling of guselkumab treatment effects in psoriatic patients. *J Pharmacokinet Pharmacodyn*. 2017;44:437-448.
  28. Byon W, Fletcher CV, Brundage RC. Impact of censoring data below an arbitrary quantification limit on structural model misspecification. *J Pharmacokinet Pharmacodyn*. 2008;35(1):101-116.
  29. Wang W, Husan F, Chow SC. The impact of patient compliance on drug concentration profile in multiple doses. *Stat Med*. 1996;15(6):659-669.
  30. Nedelman JR. On some “disadvantages” of the population approach. *AAPS J*. 2005;7(2):E374-E382.
  31. Soy D, Beal SL, Sheiner LB. Population one-compartment pharmacokinetic analysis with missing dosage data. *Clin Pharmacol Ther*. 2004;76(5):441-451.
  32. Kumar VV, Duffull SB. Methods to account for inaccuracies in the dosing history when performing population pharmacokinetic analysis. *Pharm Res*. 2008;25(12):2740-2749.
  33. Gupta P, Hutmacher MM, Frame B, Miller R. An alternative method for population pharmacokinetic data analysis under non-compliance. *J Pharmacokinet Pharmacodyn*. 2008;35(2):219-233.
  34. Friberg LE, Isbister GK, Hackett LP, Duffull SB. The population pharmacokinetics of citalopram after deliberate self-poisoning: a Bayesian approach. *J Pharmacokinet Pharmacodyn*. 2005;32(3-4):571-605.
  35. Wang Y, Liu X. Handling missing dosing history in population pharmacokinetic modeling: an extension to MDM method. *CPT Pharmacometrics Syst Pharmacol*. 2019;8(1):39-49.
  36. Mu S, Ludden TM. Estimation of population pharmacokinetic parameters in the presence of non-compliance. *J Pharmacokinet Pharmacodyn*. 2003;30(1):53-81.
  37. Gibiansky L, Gibiansky E, Cosson V, et al. Methods to detect non-compliance and reduce its impact on population PK parameter estimates. *J Pharmacokinet Pharmacodyn*. 2014;41(3):279-289.
  38. Kenna LA, Labbe L, Barrett JS, Pfister M. Modeling and simulation of adherence: approaches and applications in therapeutics. *AAPS J*. 2005;7(2):E390-E407.
  39. Mould DR, Holford NH, Schellens JH, et al. Population pharmacokinetic and adverse event analysis of toptotecan in patients with solid tumors. *Clin Pharmacol Ther*. 2002;71(5):334-348.
  40. Johansson AM, Karlsson MO. Comparison of methods for handling missing covariate data. *AAPS J*. 2013;15(4):1232-1241.
  41. Johansson AM, Karlsson MO. Multiple imputation of missing covariates in NONMEM and evaluation of the method's sensitivity to  $\eta$ -shrinkage. *AAPS J*. 2013;15(4):1035-1042.
  42. Keizer RJ, Zandvliet AS, Beijnen JH, Schellens JHM, Huitema ADR. Performance of methods for handling missing categorical covariate data in population pharmacokinetic analyses. *AAPS J*. 2012;14(3):601-611.
  43. Upton RN, Mould DR. Investigations of the representation of covariate changes in NONMEM. Paper presented at: Meeting of the Population Approach Group in Australia & New Zealand (PAGANZ 2013); February 13-15, 2013; Brisbane, Australia.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Irby DJ, Ibrahim ME, Dauki AM, et al. Approaches to handling missing or “problematic” pharmacology data: Pharmacokinetics. *CPT Pharmacometrics Syst. Pharmacol*. 2021;10:291–308. <https://doi.org/10.1002/psp4.12611>