Received: 25 November 2020 | Revised: 2 March 2021 | Accepted: 3 March 2021

DOI: 10.1002/psp4.12616

PERSPECTIVE

Evaluation of bias in weighted residual calculations when handling below the limit of quantification data using Beal's M3 method

Concentration data below the limit of quantification (BLQ) are common in population pharmacokinetic (PK) analyses, and one method used to accommodate these during nonlinear mixed effects modeling is the M3 method. A recent community discussion questioned potential bias in weighted residual plots when M3 is applied, and a simulation study was conducted to evaluate this bias. Weighted-residual bias in subjects with BLQ data was found to be small and probably ignorable in both intense and sparse sampling designs.

Weighted residuals, both traditional weighted residuals (WRES) and conditional weighted residuals (CWRES), are common metrics to graphically evaluate model acceptability in population analyses.[1] They represent the difference between the observed concentration and the prediction under the model, which are then weighted to standardize and decorrelate the residuals. WRES and CWRES are commonly plotted against TIME and population predictions and are expected to be randomly scattered around zero with the bulk of the data points within two standard deviation units.

Limited by the lower limit of quantification (LLOQ) of analytical techniques, it is not uncommon to have concentrations reported as below the LLOQ (BLQ) in PK studies. Although various approaches have been proposed to accommodate BLQ data,[2–4] Beal's M3 method currently appears to be most common. It integrates the likelihood function over the interval [-∞; LLOQ] and maximizes the likelihood of the concentration being BLQ with respect to model parameters. However, by default, the M3 method suppresses the computation of the entire set of weighted residuals for any subject with at least one BLQ observation.

In a 2010 NONMEM Users Network Archive thread,[5] it was suggested that this was a bug in the NONMEM software (ICON plc Development Solutions). Tom Ludden provided an historical perspective that Stuart Beal intentionally excluded the calculation of weighted residuals for each subject with BLQ data due to a concern that all weighted residuals for that subject might be biased. A particularly lucid explanation was contributed by Matt Hutmacher, acknowledging that "residuals do not provide great diagnostic value unfortunately for data sets with censored data. BQL observations influence the fit through the censored likelihood, but these observations are not represented in the residual diagnostic plots."

Although the concern for bias is real, recent NONMEM functionality, MDVRES (missing dependent variable for residual calculation),[6] allows one to easily obtain previously suppressed weighted residuals for concentrations above the LLOQ in subjects that have at least one concentration reported as BLQ. While using the M3 method, assigning MDVRES = 1 to censored BLQ data excludes the residuals from being calculated for these observations while allowing residuals to be computed for observations above LLOQ. Our recent question raised in the forum[7] motivated us to conduct a simulation study to investigate the extent to which weighted residual calculations in subjects having some BLQ data might be biased when using the M3 method together with MDVRES. It was not our intent to evaluate bias in decisions made based on plots using weighted residuals.

Simulations were performed assuming a one-compartment PK model with a depot compartment using mrgsolve package version 0.10.1.[8] Between-subject variability (BSV) was assumed to be log-normally distributed. Parameter values (and BSV) used for this simulation were the following: clearance (CL) 8.0 L/h/70 kg (20%), volume of distribution (V) 25.0 L/70 kg (25%), and absorption rate constant 1.5 h^{-1} (30%). The residual error model (RUV) used in the simulation was a proportional error model with a coefficient of variation (CV) of 15%. Normally distributed body weights with a mean of 70 kg and a standard deviation of 10 kg were used for scaling CL (allometric exponent 3/4) and V (allometric exponent 1).

Each in silico subject received a dose of 100 mg. Sampling strategies were chosen to assure that most simulated concentrations would be neither above nor below

LLOQ. Two scenarios were considered. In the first scenario (scenario 1), single doses in a rich sampling scheme were assumed with concentrations simulated at 0.5, 1, 1.5, 2, 3, 4, 6, 9, and 12 h after dosing. In a second scenario (scenario 2), a sparse sampling scheme was assumed with dosing at steady state as might be seen in a clinical outpatient study. The administration of 100 mg doses every 12 h was simulated with 60% of subjects providing three samples and 40% of subjects providing four samples. Samples were randomly collected at approximately 0.5, 2, 6, and/or 12 h. A total of 1000 subjects were simulated for each scenario. From the simulated data set for each scenario, reduced data sets were created that included observations labeled as BLQ. An LLOQ was chosen as 0.1 mg/L to produce data sets in which approximately 50% of subjects had at least one BLQ observation. The reduced data set for scenario 1 included 48% of subjects with a BLQ observation at 12 h. In addition, 12% of those subjects also had BLQ data at 9 h, and 0.2% also demonstrated BLQ data at 6 h. For scenario 2, 40% of subjects demonstrated a BLQ observation at the 12-h trough collection.

Analysis was performed with the first-order conditional estimation with interaction algorithm using ADVAN2/TRANS2 subroutine in NONMEM 7.5 (ICON Development Solutions). The full data set containing no BLQ data in each scenario was analyzed to provide our least biased estimates of CWRES and WRES in the standard manner. The two data subsets that include BLQ observations were analyzed with the M3 method, and MDVRES functionality was used to allow the calculation of the weighted residuals for subjects with BLQ observations. The results from these scenario-paired analyses were nearly identical with the difference in all parameter values being <3%. WRES and CWRES from these four analyses were tabulated. From these four tables of weighted residuals, subsets were constructed that contained weighted residuals from only subjects who had at least one BLQ observation as these are the residuals suspected of bias. The resulting weighted residual subsets, heretofore called SUBSET and M3, were compared to evaluate bias for both scenario 1 and scenario 2.

Bias was evaluated at the population level, and because the original concern was that residuals for the entire individual could be biased, we also evaluated bias at the individual level. Bias was evaluated at the population level by plotting the paired CWRES (and WRES) for M3 versus SUBSET data. The averages and 95% confidence intervals (CIs) of the CWRES (WRES) without regard to individual identification numbers for the SUBSET and M3 data sets in both scenarios were also calculated. Finally, the pairwise deviations between the SUBSET CWRES (WRES) and the M3 CWRES (WRES) data were calculated. The mean and 95% CIs of the paired deviations were computed to determine the overlap in CIs.

To explore bias at the individual level, the set of weighted residuals for each individual was averaged and plotted as a histogram. Ideally, if there is no bias for an individual, the observations are expected to be randomly scattered about the predicted curve, providing an average weighted residual that is close to zero. If persistent bias existed in the predictions for an individual using the M3 method, the average of the weighted residuals could be large in either a positive or negative direction. For example, in scenario 1 with eight residuals calculated in most individuals, we considered a weighted residual sum that exceeds ±8 standard deviation units, for an average exceeding ±1 unit, as an indication that substantial bias might exist for that subject. The means and 95% CIs of these individual average weighted residuals were computed and compared between the SUBSET and M3 data. Finally, the pairwise deviations between the means of the individual average weighted residuals calculated with the SUBSET and the M3 data sets were plotted as a histogram for each scenario. The mean and 95% CIs of the pairwise deviations were also calculated.

For bias evaluation at the population level, the CWRES calculated with the M3 method align well with the SUBSET CWRES in both scenarios (Figure 1, upper panels). Although the WRES calculated with M3 method align well with the SUBSET WRES in scenario 1, the alignment in scenario 2 exhibits a distinct pattern (Figure 1, lower right panel). Although small, the very lowest WRES tend to be upwardly biased, whereas higher WRES tend to be downwardly biased. This pattern is largely absent in scenario 1, although the very highest WRES tend to be slightly downwardly biased (Figure 1, bottom left). The means of CWRES (WRES) in the SUBSET and M3 data sets are comparable and the 95% CIs overlap (Table 1, top).

For bias evaluation at the individual level, the 95% CIs of the means of individual average weighted residuals between the SUBSET and M3 data sets overlap in both scenarios, suggesting a difference that is not significant (Table 1, bottom). Histograms of the individual average weighted residuals also indicate that the distributions of individual average weighted residuals are comparable between SUBSET and M3 (Figures S1 and S2). In addition, histograms of individual average weighted residuals for both scenarios demonstrate that most of them are within the range of ±1 standard deviation units with few individuals outside the range, although it is clear that the distribution of weighted residuals is wider for scenario 2 than scenario 1. The 95% CIs of the pairwise deviations do not include zero, and this deviation bias can readily be seen in the histograms (Figure S3). However, the mean deviations are quite small at less than 0.1 standard deviation units. These small deviations may simply be because of different amounts of data available for fitting between the two data

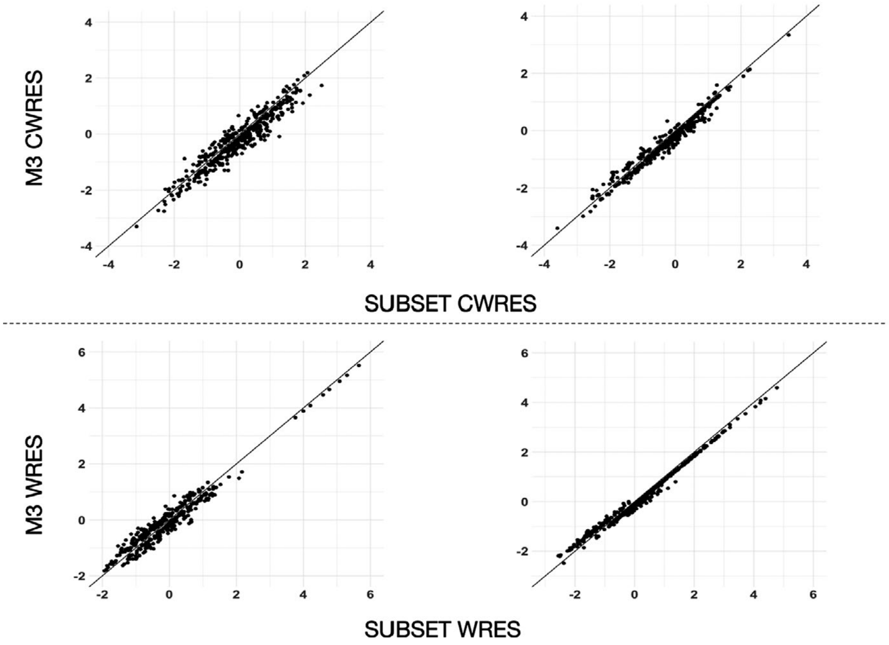

FIGURE 1 Weighted residuals for both the SUBSET and the M3 data in both scenarios. Upper panels present CWRES; lower panels present WRES. Left panels present scenario 1; right panels present scenario 2. The absolute values of the WRES and CWRES less than 0.1 units in all plots were censored to improve clarity. CWRES, conditional weighted residuals; WRES, traditional weighted residuals

TABLE 1 Bias calculations

Bias	SUBSET, mean (95% CI)	M3, mean (95% CI)	Pairwise deviations, mean (95% CI)
Population level			
Scenario 1: intensive sampling			
CWRES	−0.100 (−0.133 to −0.067)	−0.138 (−0.170 to −0.106)	−0.038 (−0.042 to −0.034)
WRES	−0.036 (−0.069 to −0.003)	−0.055 (−0.088 to −0.023)	−0.019 (−0.022 to −0.016)
Scenario 2: sparse sampling			
CWRES	−0.305 (−0.369 to −0.242)	−0.390 (−0.451 to −0.329)	−0.085 (−0.094 to −0.076)
WRES	−0.188 (−0.255 to −0.122)	−0.249 (−0.308 to −0.187)	−0.059 (−0.068 to −0.051)
Individual level			
Scenario 1: intensive sampling			
CWRES	−0.106 (−0.139 to −0.073)	−0.144 (−0.177 to −0.111)	−0.038 (−0.041 to −0.035)
WRES	−0.041 (−0.073 to −0.009)	−0.059 (−0.092 to −0.027)	−0.018 (−0.022 to −0.016)
Scenario 2: sparse sampling			
CWRES	−0.352 (−0.415 to −0.288)	−0.435 (−0.497 to −0.374)	−0.084 (−0.091 to −0.076)
WRES	−0.227 (−0.285 to −0.169)	−0.286 (−0.341 to −0.230)	−0.058 (−0.065 to −0.051)

Abbreviations: CI, confidence interval; CWRES, conditional weighted residuals; WRES, traditional weighted residuals.

sets. Also contributing to these deviations may be the nonnormality of these weighted residual distributions.

We conclude that bias in CWRES and WRES can be detected but is small and unlikely to impact decisions made based on weighted residual–based diagnostic plots when the M3 method with MDVRES is performed to accommodate BLQ observations in the scenarios we studied. However, the scope of the scenarios examined in this Perspective is limited and does not explore situations such as multicompartment models, alternative random effects models, the range of CVs in the RUV model, other sampling designs, or varying amounts of BLQ data. Hence, it is always good practice to evaluate goodness of fit using several approaches when looking for reasons to revise a particular model.

FUNDING INFORMATION

This publication was partially supported by the Office of Orphan Products Development of the US Food and Drug Administration (R01FDR0006100).

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

Mutaz M. Jaber
Shen Cheng
Richard C. Brundage

Department of Experimental and Clinical Pharmacology, University of Minnesota, Minneapolis, Minnesota, USA

Correspondence
Richard C. Brundage, University of Minnesota, 717 Delaware St. SE, Room 464, Minneapolis, MN 55455, USA.
Email: brund001@umn.edu

ORCID

Mutaz M. Jaber https://orcid.org/0000-0002-7536-8753
Shen Cheng https://orcid.org/0000-0002-7493-4784
Richard C. Brundage https://orcid.org/0000-0003-4465-5036

REFERENCES

1. Nguyen THT, Mouksassi M-S, Holford N, et al. Model evaluation of continuous data pharmacometric models: metrics and graphics. *CPT Pharmacometrics Syst Pharmacol*. 2017;6(2):87-109. https://doi.org/10.1002/psp4.12161
2. Beal SL. Ways to fit a PK model with some data below the quantification limit. *J. Pharmacokinet. Pharmacodyn*. 2001;28(5):481-504. https://doi.org/10.1023/A:1012299115260
3. 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. https://doi.org/10.1007/s10928-008-9094-4
4. Bergstrand M, Karlsson MO. Handling data below the limit of quantification in mixed effect models. *AAPS J*. 2009;11(2):371-380. https://doi.org/10.1208/s12248-009-9112-5
5. N. Discussion Group. RES and WRES output with Beal's M3 method. https://www.cognigen.com/nmusers/2010-April/2444.html. Published 2010.
6. Bauer RJ. *NONMEM User Guide. Introduction to NONMEM 7.5.0*. Gaithersburg, MD: ICON plc; 2019.
7. N. Discussion Group. M3 method–WRES, and CWRES. https://www.cognigen.com/nmusers/2020-September/7971.html. Published 2020
8. Elmokadem A, Riggs MM, Baron KT. Quantitative systems pharmacology and physiologically-based pharmacokinetic modeling with mrgsolve: a hands-on tutorial. *CPT Pharmacometrics Syst Pharmacol*. 2019;8(12):883-893. https://doi.org/10.1002/psp4.12467

SUPPORTING INFORMATION

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