J Pharmacokinet Pharmacodyn (2008) 35:101–116
DOI 10.1007/s10928-007-9078-9

Impact of censoring data below an arbitrary quantification limit on structural model misspecification

Wonkyung Byon · Courtney V. Fletcher · Richard C. Brundage

Received: 24 May 2007 / Accepted: 2 October 2007 / Published online: 26 October 2007

Abstract It is not uncommon in pharmacokinetic (PK) studies that some concentrations are *censored* by the bioanalytical laboratory and reported qualitatively as below the lower limit of quantification (LLOQ). Censoring concentrations below the quantification limit (BQL) has been shown to adversely affect bias and precision of parameter estimates; however, its impact on structural model decision has not been studied. The current simulation study investigated the impact of the percentage of data censored as BQL on the PK structural model decision; evaluated the effect of different coefficient of variation (CV) values to define the LLOQ; and tested the maximum conditional likelihood estimation method in NONMEM VI (YLO). Using a one-compartment intravenous model, data were simulated with 10–50% BQL censoring, while maintaining a 20% CV at LLOQ. In another set of experiments, the LLOQ was chosen to attain CVs of 10, 20, 50 and 100%. Parameters were estimated with both one- and two-compartment models using NONMEM. A type I error was defined as a significantly lower objective function value for the two-compartment model compared to the one-compartment model using the standard likelihood ratio test at $\alpha = 0.05$ and $\alpha = 0.01$. The type I error rate substantially increased to as high as 96% as the median of percent censored data increased at both the 5% and 1% alpha levels. Restricting the CV to 10% caused a higher type I error rate compared to the 20% CV, while the

W. Byon · R. C. Brundage (✉)
Department of Experimental and Clinical Pharmacology, University of Minnesota,
308 Harvard St. SE, Minneapolis, MN 55455, USA
e-mail: brund001@umn.edu

Present address:
W. Byon
Pfizer Inc., 558 Eastern Point Rd., Groton, CT 06340, USA

C. V. Fletcher
University of Nebraska Medical Center, 986000 Nebraska Medical Center,
Omaha, NE 68098-6000, USA

error rate was reduced to the nominal value as the CV increased to 100%. The YLO option prevented the type I error rate from being elevated. This simulation study has shown that the practice of assigning a LLOQ during analytical methods development, although well intentioned, can lead to incorrect decisions regarding the structure of the pharmacokinetic model. The standard operating procedures in analytical laboratories should be adjusted to provide a quantitative value for all samples assayed in the drug development setting where sophisticated modeling may occur. However, the current level of precision may need to be maintained when laboratory results are to be used for direct patient care in a clinical setting. Finally, the YLO option should be considered when more than 10% of data are censored as BQL.

Keywords LLOQ · BQL · Censoring · Structural model · Population pharmacokinetics

Introduction

Scientists performing pharmacokinetic (PK) and pharmacodynamic (PD) analyses often rely upon the results from a bioanalytical laboratory as a source of data. It is generally assumed that the analytical laboratory follows Good Laboratory Practices and produces concentration data that are accurate and precise.

It is not uncommon that some concentrations are *censored* by the bioanalytical laboratory since those concentrations are below the lower limit of quantification (LLOQ). The acceptance of a LLOQ in analytical methods development is nearly universal. A common feature of most analytical methods is that as the analyte signal increases, so does the system noise. When the noise is proportional to the analyte signal, a constant coefficient of variation (CV) results. However, as concentrations become very low, this proportionality is lost and noise begins to dominate the signal. This is seen as a CV that increases rapidly as concentrations fall below some level that is unique to the compound being analyzed and the instrumentation. In an effort to report only those concentrations that are considered to have acceptable precision, the laboratory determines a LLOQ and truncates a standard curve so that no concentrations are reported below that limit. The LLOQ is often defined in practice as the lowest concentration on the standard curve that is associated with a CV of no more than 20%. The 20% CV is suggested by the FDA Guidance for Industry Bioanalytical Method Validation and other reports [1,2]. Typically, any sample associated with a signal less than LLOQ is not reported quantitatively, but as below the quantification limit (BQL). When the laboratory has evidence that analytical variability is greater than 20%, it is decided that the value is too imprecise to report.

Although these standard operating procedures are well intentioned, the policy of censoring observations below the LLOQ violates one of the assumptions PK/PD modelers often make. When using the maximum likelihood estimation method in fitting models to data, it is assumed that residual errors are independent and normally distributed with zero mean and a variance. However, censoring data below the LLOQ truncates the tail of this normal distribution and violates the assumption of residual errors.

When designing a PK study, it is recognized that it is waste of resources to sample after you expect the drug to be undetectable. Nonetheless, if the pharmacokinetic characteristics of a drug are not well characterized in advance, the percentage of concentrations reported as BQL may become substantial at later time points if an initial estimate of half-life is upwardly biased. In addition, if the design is intending to capture multiple analytes, e.g., parent compound and metabolite or two different drugs, and one of those compounds has a short half-life while the other is longer, the percentage of concentrations reported as BQL is expected to be substantial at later time points for the drug with the shorter half-life. This becomes more evident if a single analytical method is used to quantify both analytes. Other reasons exist for BQL censoring at early time points. Lower than anticipated concentrations may be observed in multiple-dosing studies with nonadherent subjects. Slowly absorbed drugs and drugs with controlled-release characteristics may exhibit a lag time in which drug cannot be quantified for some period of time after drug administration.

The impact of censoring has been examined in population PK settings and procedures for handling BQL information have been suggested [3–6]. However, these references have focused on bias and precision of parameter estimates when some data were censored as BQL. Since the BQL censoring occurs more frequently at later time points, a visual examination of the cloud of concentration–time data can appear to be associated with a multiple-compartment drug. To our knowledge, structural model misspecification related to BQL censoring has not been examined.

Assuming a one-compartment IV-bolus PK model, the objectives of this paper are to (1) assess the impact of the percentage of data censored as BQL on the PK structural model decision; (2) evaluate the effect of different CV values to define the LLOQ on the structural model decision; and (3) evaluate the use of a maximum conditional likelihood estimation method available in NONMEM VI (YLO/LAPLACIAN). We begin with a description of the motivating example that led us to conduct this simulation study.

Motivating example

The results from a prospective, randomized, open-label trial of concentration-controlled versus standard dose combination therapy of zidovudine (ZDV), lamivudine, and indinavir in antiretroviral-naive HIV-infected subjects have been previously published [7,8]. To summarize the study design, all the participants underwent 8-h intensive PK evaluations on three occasions at weeks 2, 28, and 56. The sampling times were pre-dose, 0.5, 1, 2, 3, 4, 5, 6, 7, and 8 h post-dose. Participants were fasted before and for 2 h after dose for the PK evaluation. The current motivating example centers on the ZDV population PK analysis.

From 40 subjects, 576 ZDV concentrations were available. The range of ZDV doses was 600–900 mg/day. Approximately 33% of plasma concentrations were censored as BQL (<0.02 mg/l). NONMEM VI using the first-order conditional estimation with interaction (FOCEI) was used for the analysis. The BQL censored data were treated as missing, rather than set to $1/2\times$ LLOQ. Based upon ZDV population

pharmacokinetic references [9–11], either one or two compartment structural models were possible. Structural model building was guided by the NONMEM objective function value (OFV) and diagnostic plots to visually inspect goodness/lack of fit. For ZDV, the OFV decreased by 230 units with the addition of the apparent peripheral volume (Vp/F) and inter-compartmental clearance (Q/F) for the second compartment, indicating that two-compartmental model described the dataset significantly better than one-compartment model (χ^2, $\alpha < 0.05$, df = 2). For the one-compartment model, the values of clearance (CL/F) and volume of distribution (V/F) were 143 l/h and 185 l, respectively. For the two-compartment model, the parameters were as follows: CL/F = 132 l/h, V/F = 131 l, Q/F = 18.9 l/h, and Vp/F = 135 l. The absorption rate constant (Ka) was fixed to 5.0 h^{-1} for both structural models. The diagnostic plots were qualitatively similar, but slightly favored the two-compartment model. A plot of the observed data, along with the medians of the individual predicted (IPRED) and the observed concentrations was constructed. Since this diagnostic plot can be biased when only non-censored data are presented, observed concentrations reported as BQL and simulated IPRED less than LLOQ were assigned a value of one-half LLOQ. These one-half LLOQ values were used only when calculating the median values for plotting. Figure 1 displays this plot for both one- and two-compartment fits to the data. The plot visually demonstrates the two-compartment model was missing the central tendency after 6 h of drug administration, while the one-compartment model appears to more adequately predict the central tendency. Therefore, the one-compartment structural model was chosen for ZDV, which was subsequently evaluated to be acceptable and compatible with the current data using a posterior predictive check (result not shown).

Interestingly, one of the references reported two-compartment model parameters for ZDV [9]. The final parameter estimates from that study were similar to those from our ZDV two-compartment model. Both studies suggest ZDV has a beta-phase half life of 6–9 h. However, short elimination half lives of 0.5–3 h in fasting adult patients are typically reported for ZDV [12]. In both studies, a large percentage of observations were censored as BQL at later time points.

When comparing the results from the two different structural models in our analyses, we found the clearance values were comparable (143 l/h vs. 132 l/h). This minimizes the clinical dosing concern related to total daily dose recommendations. However, the choice of structural model would produce different shapes in the concentration–time profiles and could lead to different expectations in terms of time to steady state and drug accumulation. In addition, there may be clinical reasons (e.g., peak or trough considerations) for choosing one particular concentration–time profile shape over another.

Given our PK analyses of ZDV with one- and two-compartment models, along with the previously published analysis, the hypothesis of possible structural model misspecification caused by censoring and ignoring BQL data seemed relevant. Therefore, a simulation study was initiated to evaluate the impact of BQL censored data on structural model misspecification in a population pharmacokinetic study.

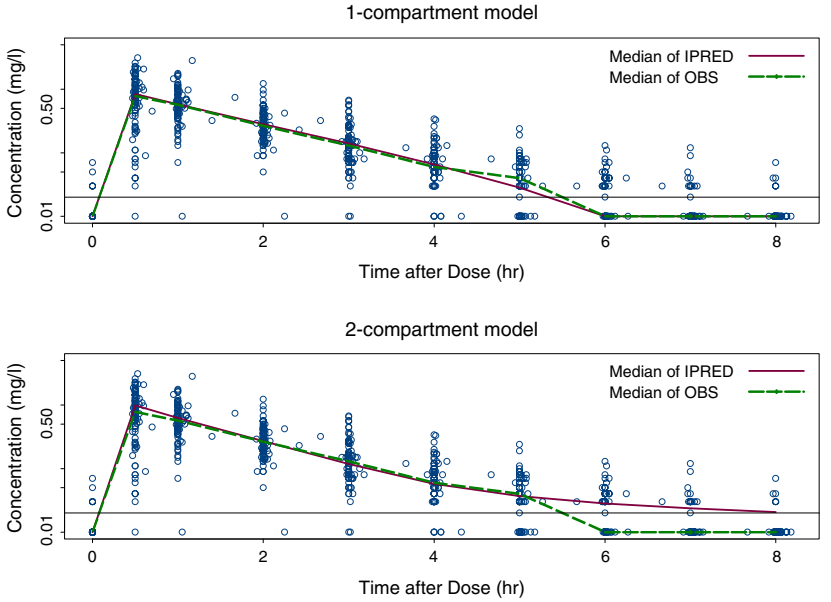

Fig. 1 Plot of observed concentrations (*open circles*) versus time after dose: median of IPRED (*solid line*) and median of OBS (*dashed line*) were plotted at each nominal time point for the one-comparment model (upper panel) and two-compartment model (lower panel). The horizontal line represents the LLOQ of ZDV (0.02 mg/l). All concentrations below the LLOQ (both censored and simulated IPRED) were assigned a value of one-half LLOQ (0.01 mg/l) for the graphical presentation

Methodology

Simulation plans

An intravenous one-compartment pharmacokinetic model was chosen for the simulation. The clearance (CL) and volume of distribution (V) were 0.693 and 1, respectively. A single unit-valued dose was administered at time zero. The PK model becomes

$$C(t) = \frac{\text{Dose}}{V}\exp^{-\frac{CL}{V}t} = \exp^{-0.693\times t}$$

and the units of time can be regarded as half-lives [4]. The between-subject variability on CL and V were assumed to follow a log-normal distribution with an exponential error model, and both were set to a 20% CV. The residual unexplained variability was chosen as a combined proportional/additive error model to represent an analytical proportional component (constant CV), and an absolute additive component (constant standard deviation) of measurement noise. The proportional error component was set to a 5% CV. A different additive error was chosen for each scenario to control the CV at the LLOQ according to the following plans and scenarios.

Eight scenarios were developed to examine two simulation plans. Table 1 summarizes the distinguishing features of each scenario. For each scenario, 500 simulations

Table 1 Summary of simulation plans for eight scenarios

Scenario no.	Proportional error (%)	Additive error	LLOQ	CV at LLOQ (%)	Median of percent data set censored as BQL and negative concentrations	Median of percent data set censored as negative concentrations
Simulation plan 1						
1	5	0.0093	0.0625	<20	10.2	0.0
2	5	0.0132	0.0884	<20	17.3	0.2
3	5	0.0187	0.1250	<20	26.7	0.4
4	5	0.0265	0.1768	<20	37.6	0.9
5	5	0.0374	0.2500	<20	49.1	1.8
Simulation plan 2						
6	5	0.0132	0.2640	<10	51.3	0.2
7	5	0.0132	0.0294	<50	3.1	0.2
8	5	0.0132	0.0139	<100	1.1	0.2

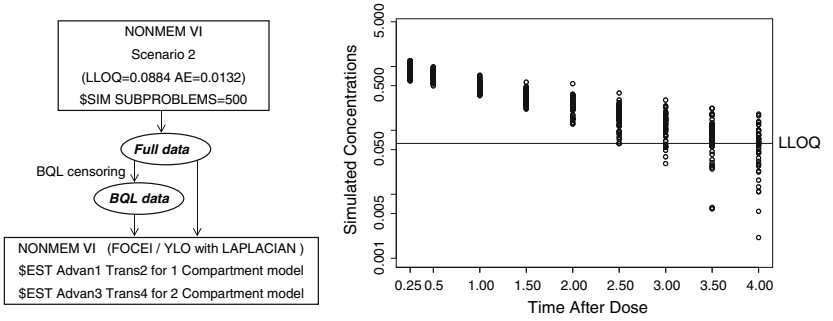

Fig. 2 Simulation flow chart using a typical simulated concentration–time profile from scenario 2 in a logarithm scale. *Open circles* represent concentrations and the *horizontal line* shows the LLOQ value for the scenario 2. While all concentrations are included in *Full data*, concentrations below LLOQ are censored to generate *BQL data*

were conducted. Each simulation consisted of 50 subjects with nine PK observations at 0.25, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, and 4 units of time. In simulation plan 1, scenarios 1–5 examined the influence of the percentage of data censored on the structural model decision when the LLOQ had no greater than a 20% CV. Five different LLOQ values were defined as the concentration at 2, 2.5, 3, 3.5, and 4 half-lives using typical parameter values. Once the LLOQ was decided for each scenario, an additive error was chosen so that the CV at the LLOQ was no more than 20%. Figure 2 displays a schematic of the simulation and illustrates a typical simulation from scenario 2. It shows simulated data along with a horizontal line that represents the LLOQ for the scenario. Scenario 2 tested the case where LLOQ was 0.0884. An additive error (AE) was set to 0.0132, so that the CV at the LLOQ was no more than 20%, as follows.

$$\begin{aligned} \mathrm{CV} = \mathrm{SD/LLOQ} &= (0.05 \times 0.0884 + \mathrm{AE})/0.0884 < 20\% \\ \mathrm{AE} &< 0.15 \times 0.0884 \end{aligned}$$

In simulation plan 2, scenarios 6–8 evaluated the impact of allowing more and less precise CVs at the LLOQ than the current practice of 20%. This was conducted as variations of scenario 2. Three different CV values were chosen as 10, 50, and 100%, and these were analyzed in addition to the 20% CV which was tested as scenario 2. For example, scenario 7 tested the impact of choosing the LLOQ with no greater than a 50% CV on the structural model decision. The additive error was kept at the original value of scenario 2 for the entire simulation plan 2. This caused the percentage of censored data to change with each scenario by allowing different levels of CVs to define the LLOQ.

Software/hardware

The simulations and population analyses were performed using a nonlinear mixed-effects model implemented in NONMEM VI [13] using Compaq Visual Fortran version 6.5. The preparation of BQL censored datasets was performed using SPLUS 7.0 (Insightful Corporation). A single Intel Xeon, Dual Processor CPU, 3 GHz, 2 GB RAM desktop computer was used for these analyses.

Simulation/BQL censoring

Each simulated data set was designated as *Full data* (no BQL censoring). This data set was then used to generate a second data set that excluded data below the relevant LLOQ to the scenario as described above and shown in Table 1. This data set was designated *BQL data*. Each scenario was associated with different levels of additive error, and 500 sets of *Full data* and their corresponding *BQL data* were generated for each scenario. BQL concentrations were censored by coding them as MDV = 1 in the data set. With the combined proportional/additive error model we used, a fraction of data was simulated to be non-positive. Since the percentages of negative concentrations were relatively low (shown in Table 1), these concentrations were deleted from the simulated datasets without replacement under the assumption of minimal impact on simulation results.

Estimation

BQL data and *Full data* were analyzed with a one-compartment model using ADVAN1 and TRANS2 and a two-compartment model using ADVAN3 and TRANS4. When the two-compartment model was tested, the peripheral volume of distribution and the inter-compartmental clearance were added into the model without between subject variability on them. The FOCEI was used for this estimation.

Additionally, a new conditional likelihood estimation feature in NONMEM VI (YLO/LAPLACIAN) was evaluated. In the usual maximum likelihood estimation (MLE) procedure, it is assumed that the probability of an observed concentration comes from a normal distribution with limits of negative infinity and positive infinity. In contrast, the YLO function in NONMEM VI applies the MLE conditioned on accepting the probability of observed concentrations are above the LLOQ [4]. The

LAPLACIAN estimation method is required with the YLO method. The YLO method was only tested on *BQL data* for scenarios 1 through 5, with both a one-compartment model and a two-compartment model.

Type I error rates

For each of the simulated data sets (both *BQL data* and *Full data*), the type I error was defined using the standard likelihood ratio test. A type I error was declared when the OFV was significantly lower for the two-compartment model compared to the one-compartment model. The error rate was determined at a level of significance of 5% and 1%; with two degrees of freedom, the associated drops in OFV from a χ^2 table were 5.99 and 9.21, respectively. It is understood that these significant values at each alpha level are not exact since the entire simulation falls into a constrained one-sided test or boundary condition [14]. However, the likelihood ratio test was chosen since decisions regarding the structural PK model are often naively guided by this test in practice. Furthermore, no clear recommendation for boundary condition testing has been set forth in the pharmacokinetic literature.

The type I error rate was determined from 500 simulations per each scenario with the following rules. First, all runs were considered when determining the type I error rate. Second, only those runs with a successful minimization were counted. Third, only those runs with a successful minimization and a successful covariance step were considered. Last, only runs with reasonable results for the two-compartment model in addition to a successful minimization and covariance step were counted. To be defined as a "reasonable" result, the alpha-phase half-life ($\alpha_{t1/2}$) had to be greater than 0.25 (the first sampling time), and the beta-phase half-life ($\beta_{t1/2}$) had to be less than 10 units (considering concentrations were sampled over 4 units of time). These stipulations were arbitrary, but were intended to remove results that were not likely to be plausible given the study design. Bias and precision of parameter estimates from each scenario were calculated for *BQL data* using the mean error (ME) and the root mean square error (RMSE), respectively [15].

Results

The simulation plan 1 and simulation plan 2 results are shown in Figs. 3 and 4, respectively. Figure 3 shows the type I error rates were elevated when datasets included BQL censoring (*BQL data*, solid line) compared to when all the data were available (*Full data*, dashed line) across all the scenarios. The increasing trend in type I error rate was observed as the median of percent censored data increased when *BQL data* were estimated at both the 5% (upper panel) and 1% (lower panel) alpha levels. For example, in the case of scenario 5 with approximately half of the data censored as BQL, 95.6% and 88.2% of 500 simulations showed a significantly lower OFV for the two-compartment model compared to the one-compartment model at alpha levels of 0.05 and 0.01, respectively.

When the rules of successful minimization, successful covariance step, and reasonable results were applied, the type I error rates were nearly identical to the results

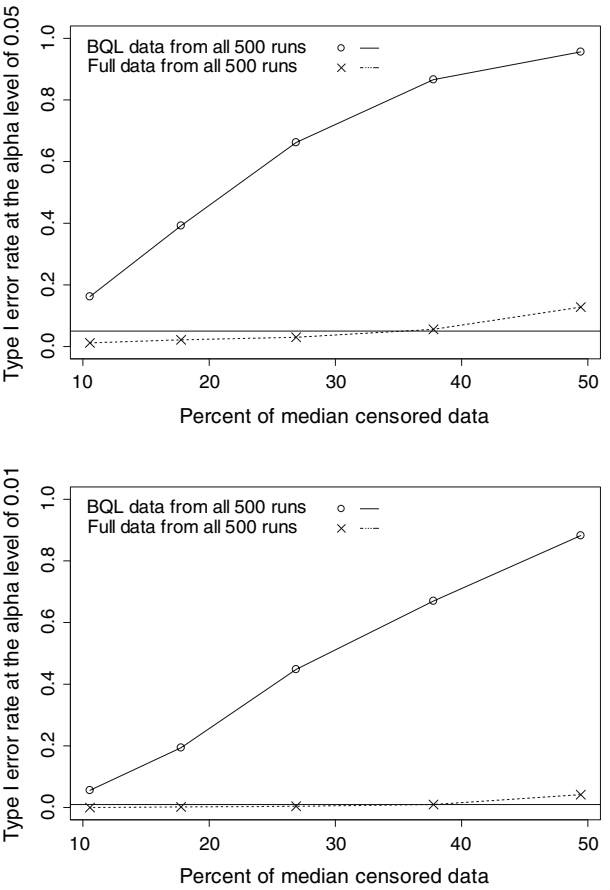

Fig. 3 Type I error rates at the 5% (upper panel) and 1% (lower panel) alpha levels in simulation plan 1 for *BQL data* (*solid lines*) and *Full data* (*dashed lines*). The *horizontal line* shows the nominal type I error rate for each case

from all 500 runs. For example, when only runs with both successful minimization and covariance steps were considered for scenario 5, the type I error rate in these 479 runs remained inflated at 96.0% (compared to 95.6% from above). Further limiting runs considered to only those with reasonable parameter estimates the type I error rate was 96.4% in the 384 runs from this scenario. Therefore, only the type I error results from the 500 runs with no limiting rules are plotted as the curves were nearly superimposable.

Type I error rates in *Full data* without BQL censoring generally stayed close or lower than the expected 5 or 1%. However, the trend was observed that the error rate slightly increased as the median of percent censored data increased, which is believed to be associated with the censoring of the originally simulated non-positive concentrations.

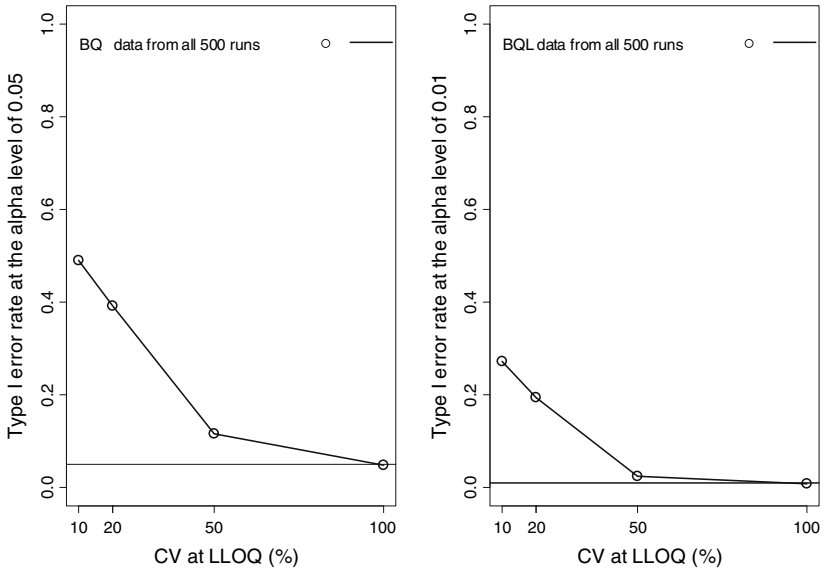

Fig. 4 Type I error rates at the 5% (left panel) and 1% (right panel) alpha levels in simulation plan 2 when different CVs were chosen for the LLOQ. The *horizontal line* shows the nominal type I error rate for each case

Table 2 summarizes the PK parameter estimates from the one- and two-compartment models of *BQL data* for each scenario in simulation plan 1, counting all 500 runs regardless of OFV changes. In the two-compartment model, several trends in PK parameters were noted as the percent censored data increased across the scenarios. There was a trend toward decreasing CL, Vp, and beta-phase half-life, while Q tended to increase.

Figure 4 shows the impact of choosing a more or less precise CV at the LLOQ than the current practice of 20%. While the type I error rate increased to 49% with a 10% CV at the LLOQ (scenario 6), it decreased when the LLOQ was chosen as a less precise value than the current practice of 20%. When the CV was allowed to be 50% at the LLOQ, the type I error rate was 11.6; a CV of 100% at the LLOQ was associated with an error rate of 4.8% at the 5% alpha level.

The result of implementing the conditional likelihood estimation available in NONMEM VI is summarized in Table 3. Only the first 100 simulations of scenarios 1–5 were tested for this new feature because the consistency of results became immediately obvious. When the YLO option was implemented with both one-compartment and two-compartment models for *BQL data*, the type I error rate for structural model misspecification was close to nominal values. Table 4 summarizes the ME and RMSE for the one-compartment fit of *BQL data* to illustrate the bias and precision of parameter estimates for each scenario, respectively. Though there was a trend that biases and precisions were slightly worse with increasing percentage of BQL censoring, they were generally low in all scenarios.

Table 2 Summary of the PK parameter estimates for the one- and two-compartment models from all 500 runs of *BQL data* for each scenario in simulation plan 1 (inf refers to a value greater than 1,000)

Scenario no.	One-compartment model			Two-compartment model		
		Median	Range		Median	Range
1	CL	0.691	(0.639, 0.747)	CL	0.630	(0.235, 0.732)
	V	1.00	(0.913, 1.10)	V	0.997	(0.902. 1.10)
	$T_{1/2}$	1.00	(0.912, 1.16)	Vp	3.79	(0.0419, 79.2)
				Q	0.0638	(0.0011, 0.431)
				$\alpha_{t1/2}$	0.987	(0.856, 1.13)
				$\beta_{t1/2}$	30.1	(1.35, inf)
2	CL	0.690	(0.638, 0.746)	CL	0.596	(0.216, 0.725)
	V	1.01	(0.912, 1.10)	V	0.996	(0.900, 1.10)
	$T_{1/2}$	1.01	(0.914, 1.16)	Vp	1.98	(0.0419, 297)
				Q	0.105	(0.001, 0.468)
				$\alpha_{t1/2}$	0.972	(0.159, 1.11)
				$\beta_{t1/2}$	15.2	(0.986, inf)
3	CL	0.688	(0.637, 0.742)	CL	0.542	(0.203, 0.713)
	V	1.01	(0.911, 1.11)	V	0.994	(0.896, 1.09)
	$T_{1/2}$	1.01	(0.918, 1.16)	Vp	1.95	(0.0419, 44.0)
				Q	0.170	(0.018, 0.495)
				$\alpha_{t1/2}$	0.945	(0.179, 1.11)
				$\beta_{t1/2}$	11	(0.968, 174)
4	CL	0.683	(0.631, 0.738)	CL	0.480	(0.192, 0.688)
	V	1.01	(0.910, 1.11)	V	0.990	(0.744, 1.09)
	$T_{1/2}$	1.03	(0.916, 1.19)	Vp	1.70	(0.0664, 12.2)
				Q	0.240	(0.0335, 1.48)
				$\alpha_{t1/2}$	0.882	(0.0780, 1.027)
				$\beta_{t1/2}$	7.97	(1.02, 53.3)
5	CL	0.669	(0.608, 0.726)	CL	0.399	(0.0525, 0.663)
	V	1.02	(0.922, 1.13)	V	0.979	(0.739, 1.09)
	$T_{1/2}$	1.06	(0.940, 1.26)	Vp	1.26	(0.103, 5.01)
				Q	0.363	(0.0494, 1.95)
				$\alpha_{t1/2}$	0.739	(0.073, 0.943)
				$\beta_{t1/2}$	5.634	(1.15, 69.3)

Table 3 Type I error rates when testing YLO at the 5 % alpha level for the first 100 simulations of scenarios 1–5 in *BQL data*

Scenario no.	Median of percent data set censored as BQL and negative concentrations	Type I error rate
1	10.2	0.00
2	17.3	0.02
3	26.7	0.02
4	37.6	0.05
5	49.1	0.06

Discussion

The 2001 FDA guidance for Bioanalytical Method Validation makes a suggestion for the LLOQ: the "lowest standard on the calibration curve should be accepted as the limit of quantification" where the conditions of "the analyte response at the LLOQ

Table 4 ME (bias) & RMSE (precision) for the one-compartment fit of *BQL data*

Scenario no.	CL		V		ω_{CL}		ω_V		$\sigma_{proportional}$		$\sigma_{additive}$	
True values	0.693		1		0.2		0.2		0.05		See Table 1	
	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE
1	−0.00084	0.01931	0.00429	0.02976	−0.00450	0.02060	−0.00428	0.02010	−0.00005	0.00349	−0.00014	0.00104
2	−0.00203	0.01937	0.00630	0.03033	−0.00493	0.02073	−0.00399	0.02017	−0.00016	0.00440	−0.00019	0.00147
3	−0.00472	0.01977	0.00943	0.03195	−0.00515	0.02088	−0.00328	0.02036	−0.00055	0.00576	−0.00004	0.00206
4	−0.01013	0.02187	0.01368	0.03404	−0.00565	0.02176	−0.00228	0.02079	−0.00130	0.00909	−0.00009	0.00364
5	−0.02404	0.03125	0.02320	0.04057	−0.01050	0.02639	−0.00179	0.02224	−0.00214	0.01661	−0.00055	0.00646
6	−0.00825	0.02095	0.01106	0.03299	−0.00597	0.02172	−0.00266	0.02070	−0.00222	0.00770	−0.00020	0.00690
7	−0.00049	0.01934	0.00369	0.02965	−0.00456	0.02062	−0.00460	0.02015	0.00036	0.00390	−0.00033	0.00103
8	−0.00022	0.01937	0.00289	0.02967	−0.00423	0.02055	−0.00419	0.02009	0.00036	0.00382	−0.00036	0.00102

should be at least 5 times the response compared to blank response" and "analyte peak (response) should be identifiable, discrete, and reproducible with a precision of 20% and accuracy of 80–120%" are met [1]. This 20% definition of precision is frequently considered to be the gold standard to decide the LLOQ in analytical methods development. This, however, is not the only approach; the International Conference on Harmonisation states that "the quantitation limit may be expressed as $10\sigma/S$" where "σ is the standard deviation of the response and S is the slope of the calibration curve" [16]. Nonetheless, the practice of choosing the LLOQ to be associated with no more than a 20% CV seems to have grown into a decree of law, rather than the connotation of a recommendation. Regardless, the current simulation study demonstrates that the censoring of concentrations as BQL can lead to structural model misspecification in population PK analyses. Interestingly, regardless of OFV changes, some PK parameter estimates from the two-compartment models of *BQL data* such as beta-phase half-lives became more plausible as the median of percent censored data increased. Furthermore, relaxing the current practice of censoring data with less than 20% precision can help prevent this misspecification.

With the naïve cut-off values in the χ^2-distribution at two degrees of freedom, the type I error rates from *Full data* (without any BQL censoring) in simulation plan 1 were lower than the nominal value at both the 5% and 1% alpha levels in scenario 1–3. This is a known result under the constrained one-sided test using log likelihood ratio test. It has been observed that the actual distribution of changes in the OFV to test a variance and a covariance of being zero in a linear mixed-effect modeling under a boundary condition did not strictly follow the χ^2-distribution at two degrees of freedom; the distribution was approximately midway between one and two degrees of freedom [14]. Therefore, when some parameters are set to a boundary of zero in linear mixed-effects modeling, it is recommended to adjust the degrees of freedom by putting half the mass on 1 degree of freedom and half the mass on two degrees of freedom ($0.5\chi_1^2 + 0.5\chi_2^2$) [14]. It has also been suggested that a 0.1 significance level be used instead of 0.05 [17]. In our simulation, two additional parameters for the two-compartment model (the peripheral volume of distribution and inter-compartmental clearance) were tested as being zero in the null hypothesis against the alternative hypothesis of testing them being positive. Since these PK parameters cannot be negative, this falls into a boundary condition with a constrained one-sided test. To our knowledge, the actual distribution of OFV changes with parameters under the presence of boundary conditions in nonlinear mixed-effect modeling has not been thoroughly studied. These boundary conditions are encountered with the addition of a second structural compartment, a lag time in absorption, or a variance-covariance term. Our simulation results showed a similar pattern of lower type I error rates with the naïve use of likelihood ratio testing applied to fixed effects as variance–covariance terms did in linear mixed-effect modeling. The naïve application of likelihood ratio testing is a common practice when establishing additional PK structural compartments in NONMEM analyses, and the issue of honest P-values for hypothesis testing needs further investigation.

Another trend in type I error rate in *Full data* was that it increased across scenarios 1 through 5. This is suspected to result from the simulated non-positive data which were removed from the parent datasets. As the additive error increased to allow larger percentages of data to be censored, this also caused higher percentages of concentrations

to be simulated as negative values. From the 500 simulations in scenario 5, the percent of negative concentrations ranged from 0% to 4.4%. This could have been avoided by simulating the logarithm of concentrations using a *transform-both-sides* approach. However, the simulation results of testing the interference of BQL censoring with a structural model decision are still valid by comparing the results from *BQL data* and *Full data*.

The simulation results from scenario 6 to 8 showed that accepting a CV at the LLOQ of 100% would effectively prevent the inflation of type I error in structural model decision with little concern of bias and precision when data are sufficiently informative to estimate the parameters of interest. By increasing the CV at the LLOQ from 20% to 50%, the median of percent data censored as BQL dropped from 17.3% to 3.1% increasing the numbers of data available for analysis. However, the reduced type I error rate is not likely due to simply the addition of more data. Importantly, the data that are included by allowing a higher CV are also allowing a more normal distribution of residual errors at lower predicted concentrations. MLE regression makes the assumption of normally distributed residual errors, and this assumption is violated when concentration data are truncated with BQL censoring.

The maximum conditional likelihood estimation, which was previously discussed as method 2 in a recent paper [4] exploring the effect of BQL censoring on bias and precision, became available with the issue of NONMEM VI. The simulation results show that this new feature minimized the elevation of type I error across all scenarios. Therefore, in a PK analysis that includes a substantial fraction of data being censored, the use of YLO options should be strongly considered to avoid any model misspecification. Bringing this recommendation back to the ZDV data in the motivating example, we did analyze these data with the YLO option. Contrary to the likelihood ratio test result described in the motivating example section that highly favored the two-compartment model, a significant OFV drop was not observed when testing a two-compartment model against a one-compartment model with YLO set at the LLOQ.

As the current simulation study focused on the structural model misspecification rather than exploring bias and precision of parameter estimates, the results in Table 4 show the chosen study design was informative for the one-compartment parameters of interest, regardless of the fraction of data censored as BQL.

This simulation study is limited to a specific case of BQL censoring and structural model misspecification where the simulated model is a one-compartment IV bolus input model. More complicated structural models with other elimination pathways and routes of administration, and the impact on covariate selection still need to be investigated. However, issues related to the number and spacing of the samples are likely to strongly interact with the BQL issue in more complicated models, and that will need to be addressed. Also not addressed in this simulation study is an examination of the effect of BQL censoring on the type II error rate in structural model decisions.

Conclusions

This simulation study has shown that the nonrandom censoring of BQL data can lead to incorrect decisions regarding the pharmacokinetic compartmental model structure.

The practice of assigning a LLOQ during analytical methods development, although well intentioned, in the applications described in this work can cause incorrect pharmacokinetic conclusions and should be revisited. We acknowledge that pharmacokinetic scientists working in different environments may require different assumptions regarding the data provided to them. The clinician working in a patient care setting may require more precise concentration data to confidently use that data in the dosing of a patient. In the clinical setting, the current limit of a 20% CV for the LLOQ may need to be maintained, or BQL concentrations could be reported quantitatively with a suitable cautionary flag indicating less precision. However, for the pharmacokineticist working in a basic/research setting, who is capable of understanding error models and analyzing the data appropriately, then the results from our work would indicate that concentrations associated with CVs of 100% or perhaps even higher should be reported quantitatively rather than be censored. Finally, in a PK analysis that includes more than 10% of the data being censored, the YLO option available in NONMEM VI should be strongly considered to avoid model misspecification.

Acknowledgements Support: RO1 AI33835 from the National Institute of Allergy and Infectious Diseases (to CVF).

References

1. FDA guidance for industry bioanalytical method validation. Available from http://www.fda.gov/cder/guidance/4252fnl.pdf Accessed on may 2001
2. Shah VP, Midha KK, Findlay JW, Hill HM, Hulse JD, McGilveray IJ, McKay G, Miller KJ, Patnaik RN, Powell ML, Tonelli A, Viswanathan CT, Yacobi A (2000) Bioanalytical method validation—a revisit with a decade of progress. Pharm Res 17(12):1551–1557
3. Hing JP, Woolfrey SG, Greenslade D, Wright PM (2001) Analysis of toxicokinetic data using NONMEM: impact of quantification limit and replacement strategies for censored data. J Pharmacokinet Pharmacodyn 28(5):465–479
4. Beal SL (2001) Ways to fit a PK model with some data below the quantification limit. J Pharmacokinet Pharmacodyn 28(5):481–504
5. Duval V, Karlsson MO (2002) Impact of omission or replacement of data below the limit of quantification on parameter estimates in a two-compartment model. Pharm Res 19(12):1835–1840
6. Beal SL (2005) Conditioning on certain random events associated with statistical variability in PK/PD. J Pharmacokinet Pharmacodyn 32(2):213–243
7. Fletcher CV, Anderson PL, Kakuda TN, Schacker TW, Henry K, Gross CR, Brundage RC (2002) Concentration-controlled compared with conventional antiretroviral therapy for HIV infection. AIDS 16(4):551–560
8. Kakuda TN, Page LM, Anderson PL, Henry K, Schacker TW, Rhame FS, Acosta EP, Brundage RC, Fletcher CV (2001) Pharmacological basis for concentration-controlled therapy with zidovudine, lamivudine, and indinavir. Antimicrob Agents Chemother 45(1):236–242
9. Zhou XJ, Sheiner LB, D'Aquila RT, Hughes MD, Hirsch MS, Fischl MA, Johnson VA, Myers M, Sommadossi JP (1999) Population pharmacokinetics of nevirapine, zidovudine, and didanosine in human immunodeficiency virus-infected patients. The National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group Protocol 241 Investigators. Antimicrob Agents Chemother 43(1):121–128
10. Mirochnick M, Capparelli E, Connor J (1999) Pharmacokinetics of zidovudine in infants: a population analysis across studies. Clin Pharmacol Ther 66(1):16–24
11. Capparelli EV, Englund JA, Connor JD, Spector SA, McKinney RE, Palumbo P, Baker CJ (2003) Population pharmacokinetics and pharmacodynamics of zidovudine in HIV-infected infants and children. J Clin Pharmacol 43(2):133–140

12. RETROVIR® prescribing information. Available from http://us.gsk.com/products/assets/us_retrovir.pdf Accessed on Nov 2006
13. Beal SL, Sheiner LB, Boeckmann AJ (eds) (1989–2006) NONMEM users guides. Icon development solutions. Ellicott City
14. Stram DO, Lee JW (1994) Variance components testing in the longitudinal mixed effects model. Biometrics 50(4):1171–1177
15. Sheiner LB, Beal SL (1981) Some suggestions for measuring predictive performance. J Pharmacokinet Biopharm 9(4):503–512
16. ICH harmonized tripartitie guideline: validation of analytical procedures Q2(R1). Available from http://www.ich.org/LOB/media/MEDIA417.pdf
17. Fitzmaurice GM, Laird NM, Ware JH (2004) Applied longitudinal analysis. Wiley