

Title: Non-Linear Mixed Modeling of Serial Sacrifice Data: Potential Improvement in Estimation Using Information from Replicated Analytical Determinations

Authors: Kyle Baron* (1) and Richard Brundage (1)

Institutions: (1) Department of Experimental and Clinical Pharmacology, University of Minnesota, Minneapolis, MN, USA

Objectives: The application of non-linear mixed effects modeling (NLMEM) techniques for the analysis of serial sacrifice (one point per animal) data has been attempted in the past with limited success. The difficulty lies in an inability of the method to distinguish between inter-individual variability in pharmacokinetic (PK) parameters and residual unexplained variability (RUV), resulting in unstable estimation and biased parameter estimates. We hypothesized that a duplicate assay of each sample might provide sufficient information to adequately identify the analytical component of the RUV. If analytical variability is a major component of total RUV, the remaining between-subject variability would be reasonably estimated. We used simulation to test this hypothesis. The goals for this study were to: (1) determine if analytical variability information would stabilize the estimation of population parameters from serial sacrifice data using NLMEM; (2) examine bias and precision of those parameter estimates; and (3) evaluate the type 1 and type 2 error rates when attempting to identify differences in exposure (clearance) between two populations.

Methods: NONMEM VI was used for all simulation and estimation runs. *Model:* Datasets were simulated under a one-compartment PK model with IV bolus input and first-order elimination, parameterized in clearance (CL) and volume (V). Variability in PK parameters was modeled as: $P_i = TVP \cdot \exp(\eta_p)$, where TVP is the typical value of the parameter P (CL or V) and $\eta_p \sim N(0, \omega_p^2)$. No covariate effects were simulated. RUV (ϵ) was assumed to be proportional to the individuals' simulated concentrations, with $\epsilon \sim N(0, \sigma^2)$. *Dosing and sampling scheme:* All *in silico* subjects received a dose of 1000 mg at time=0. Under the serial sacrifice design, six subjects were selected for simulated observation(s) at one of seven sampling times (1, 2, 4, 8, 12, 16, 24 hours). Information about RUV was obtained by simulating an observation once (1x) or repeatedly (2x or 3x) for each subject at that given time point. This represented 42 subjects being sampled once, twice or thrice, respectively; in no case was a subject sampled at more than one time point. For comparison, data were also simulated under an intensive sampling design where all 42 subjects were sampled at all 7 time points, with the observation at each time point simulated only once. *Simulation:* In all simulations, TVV and ω_v^2 were set to 10 L and 20%, respectively. To assess bias and precision of model parameters, TVCL was set to 1 L/hr and ω_{CL}^2 was varied from 20 to 50%. To assess the type 1 and type 2 error rates for identifying differences in exposure, two populations were simulated in the same dataset: a reference population with TVCL set to 1 L/hr and a test population with TVCL set to either: 1, 0.9, 0.8, 0.7, or 0.5 L/hr (0 to 50% lower TVCL than reference). The two populations were assumed to have the same ω_{CL}^2 , which varied from 20 to 50%. For all scenarios, 1000 datasets were simulated. *Estimation:* All model parameters were estimated in each simulated dataset using FOCE-I in NONMEM. The structural and error models for estimation were as simulated, with no model misspecification. When two populations were simulated with a given difference in CL, the data were analyzed under two models: a reduced model that includes a common fixed effect for CL across the two populations and a full model that allows for different TVCL for each population. The statistical significance of the extra parameter in the full model was assessed using a likelihood ratio test (LRT; df=1, size=0.05). *Analysis:* The stability of estimation under different simulation scenarios was assessed by recording the number of runs that successfully minimized and/or executed a successful covariance step. Parameter estimates were judged to be unbiased if the 95% confidence interval around the mean of the sampling distribution for that parameter included the true simulated value. The precision of parameter estimates under the serial sacrifice design was evaluated by visual comparison with the sampling distributions arising from the intensive sampling design. Precision was also assessed by calculating the fraction of estimates that were within 20 or 30% of the true simulated value. The Type I error rate of the method for determining the difference in CL across two populations was the fraction of runs where LRT chose the full model when $TVCL_{test}$ was

simulated to be equal to $TVCL_{reference}$. The power to detect a difference in CL across two populations was the fraction of runs where LRT chose the full model when $TVCL_{test}$ was simulated to be different than $TVCL_{reference}$.

Results: With ω^2_{CL} set to 30%, runs with 1x RUV minimized only 40% of the time and the rate of successfully obtaining a variance-covariance matrix was less than 20%. Both measures degraded as ω^2_{CL} increased to 50%. However, when the dataset included 2x or 3x RUV information, the minimization rate was greater than 90% and a variance-covariance matrix was obtained 69% to 84% of the time. Providing 3x residual error information appeared to offer little improvement over 2x residual error information. Minimization and the covariance step were successful 99% of the time under the intensive sampling design. As expected, σ^2 estimates became more precise as more RUV information was provided. When 3x RUV information was provided, the precision of σ^2 estimation approached that of the intensive sampling design. All σ^2 estimates from 2x and 3x serial sacrifice designs as well as the intensive sampling design were unbiased ($p < 0.05$). Under all simulation scenarios, parameter estimates of fixed effects and random effect variances were unbiased ($p < 0.05$). By visual comparison, the serial sacrifice design was somewhat less precise than under the intensive sampling design for both fixed effects and random effects variances. However, $>90\%$ of fixed effect estimates and $>75\%$ of random effect variance estimates were within $\pm 30\%$ of their true simulated values. When the dataset contained two populations with the same simulated CL, the observed Type I error rate was 5-7%. When a difference in CL was simulated between the two populations, the power of the test increased as the simulated difference increased and as the simulated inter-individual variability in CL decreased. For any simulated difference in CL $\geq 30\%$ and $\omega^2_{CL} \leq 30\%$, the power to detect the difference was $>90\%$. Under the intensive sampling design, the NLMEM method had the expected Type I error rate (5%) and power close to unity for all simulated differences in TVCL and levels of ω^2_{CL} .

Conclusions: Under this simple PK model, NLMEM appeared to provide unbiased and acceptably precise estimates of both fixed effects and random effect variances for CL and V when RUV information was provided. Estimation was reasonably stable and the LRT was a powerful test for detecting differences in CL across two populations with the nominal type 1 error rate. These results are encouraging and indicate that samples assayed in duplicate may be useful in the NLMEM analysis of serial sacrifice data. A major limitation that remains to be explored is the decrease in performance that is expected to occur as the fraction of RUV due to analytical error decreases from 100% (the current study) to 75 or 50%. Nonetheless, given the highly controlled experimental conditions associated with serial sacrifice studies, the non-analytical error may be small compared to analytical error.

References:

- Bailer, A. J. (1988). "Testing for the equality of area under the curves when using destructive measurement techniques." *J Pharmacokinet Biopharm* **16**(3): 303-9.
- Ette, E. I., A. W. Kelman, et al. (1995). "Analysis of animal pharmacokinetic data: performance of the one point per animal design." *J Pharmacokinet Biopharm* **23**(6): 551-66.
- Hing, J. P., S. G. Woolfrey, et al. (2001). "Is mixed effects modeling or naive pooled data analysis preferred for the interpretation of single sample per subject toxicokinetic data?" *J Pharmacokinet Pharmacodyn* **28**(2): 193-210.
- Mager, H. and G. Goller (1998). "Resampling methods in sparse sampling situations in preclinical pharmacokinetic studies." *J Pharm Sci* **87**(3): 372-8.
- Nedelman, J. R., E. Gibiansky, et al. (1995). "Applying Bailer's method for AUC confidence intervals to sparse sampling." *Pharm Res* **12**(1): 124-8.
- Pai, S. M., J. R. Nedelman, et al. (1996). "Performance of Bailer's method for AUC confidence intervals from sparse non-normally distributed drug concentrations in toxicokinetic studies." *Pharm Res* **13**(9): 1280-2.
- Wolfsegger, M. J. (2007). "Establishing bioequivalence in serial sacrifice designs." *J Pharmacokinet Pharmacodyn* **34**(1): 103-13.
- Yuan, J. (1993). "Estimation of variance for AUC in animal studies." *J Pharm Sci* **82**(7): 761-3.