Modeling Censored Stability Data Using Tobit Mixed Models

Seth Clark
Research CMC Statistics, Merck & Co., Inc., MAILSTOP WP37C-305
770 Sumneytown Pike, West Point, PA 19486. seth.clark@merck.com

Abstract
It is not uncommon to encounter stability data that are censored due to results being below a quantitation limit of the assay. Often various crude methods are applied to the data, such as imputing values or modeling only the portion of the stability trend that is uncensored. An alternative method is proposed utilizing Tobit regression applied in the mixed model context to account for test-to-test variation. We show the approach allows one to model the distribution mean across the entire stability profile and reduces bias due to censored values.

Key Words: censored data, regression, Tobit, stability, quantitation limit, detection limit

1. Introduction: Stability Designs, Data and Models

Stability studies evaluate many quality features of a drug substance or drug product over time, typically up to 3 or 4 years. The quality features are often related to safety and efficacy and include, for example, % of label claim of the active ingredient or biological potency, impurities, and degradates, or may include characteristics of the drug that are just quality related, such as charge distribution, pH, and moisture, for example. Stability studies often test multiple batches of material stored in different containers or product images, at different conditions (temperature & humidity).

Since each batch of material to be tested in a stability study has been stored at one condition, such as -70ºC or 2-8º/ambient and is then divided into different conditions at the start of the study when put into stability chambers, differences in conditions cannot have an effect on the trend until after time 0. This implies that the model must be constrained to have the same intercept at time 0 for multiple conditions of the same batch. This is also true for other factors where a division has occurred at time 0 and their effect begins after time 0.

1.1 Example of stability data and model

An example of stability data collected over 36 months is shown in the left plot of Figure 1. The time 0 (TZ) data include two replicates that support the trends of both the -70ºC or 2-8ºC/ambient conditions. Subsequent to time 0, the trends of the two conditions diverge with 2-8ºC/ambient showing a statistically significant loss in the quality characteristic.

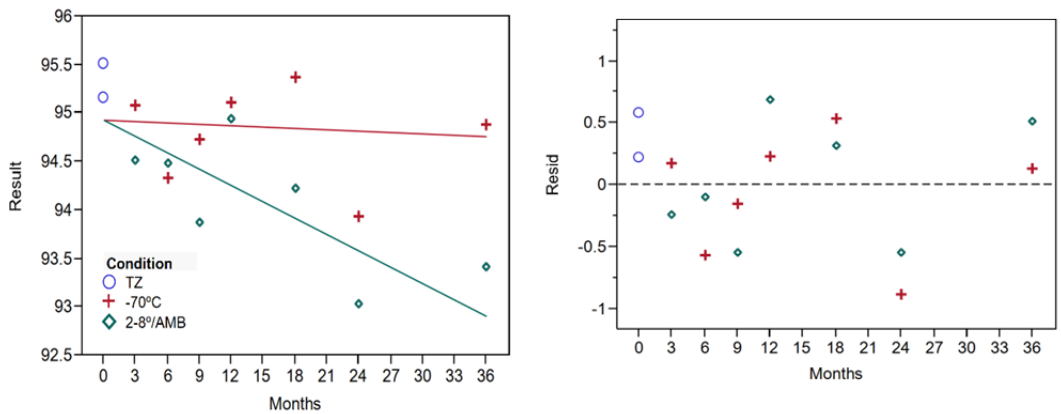

Figure 1: (Left plot) constrained linear model fit to stability data for batch 1, (right plot) corresponding residual plot.

The constraint is invoked in the model by leaving out the condition term from the intercept specification in Model 1. However, when looking at the residual plot (Figure 1, right) we see what appears to be evidence of lack-of-fit. We may consider that the trend is non-linear, but a more plausible explanation comes into focus when we understand that the two conditions for the same batch were tested in the same assay runs at each time point. The apparent lack-of-fit is likely the run-to-run (or test-to-test) variation of the assay. To capture this feature, the stability Model 1 includes an additional error term.

Model 1

$g(\text{Result}_{ij}) = Y_{ij} \; = \mu(\text{Lot,Temp,Months}) + e_{\text{Test}_i} + e_{\text{Residual}_{ij}}$, where

$\mu(\text{Lot,Temp,Months}) = (\mu_0 + \text{Lot}) + (1 + \text{Lot} + \text{Temp} + \text{Lot} * \text{Temp}) f(\text{Months})$,

factors Lot and Temp are appropriately expanded into dummy variables,

$e_{\text{Test}_i} iid \; N(0, \sigma^2_{\text{Test}})$, and independently $e_{\text{Residual}_{ij}} iid \; N(0, \sigma^2_{\text{Residual}})$,

for monotonic transformations g and f. (Model 1)

When there are multiple levels of variation in a regression model, sometimes one can simplify the analysis by averaging over the within variation, and then analyze the averages using only one error term. The example in Figure 1 shows a case where this is not possible since the within run variation is stratified by the condition factor. This latter situation is more likely in a well-designed stability study, because factors that are tested within sources of variation lead to higher precision contrasts than factors whose levels span a source of variation.

1.2 Example of Censoring in Stability

Censoring is another feature of stability data that occasionally occurs in practice. In Figure 2, the data are censored below C= 0.06%, the limit of detection (LOD) for the assay. While the censored values were set to 0, the actual results can occur anywhere between 0 and 0.06%. We do not know where they fall because the assay does not have the resolution to recognize any % peak area below this limit. More commonly, the limit of quantitation (LOQ) acts as a censoring limit. The LOQ is a point beyond which the assay does not yield results of acceptable precision or accuracy.

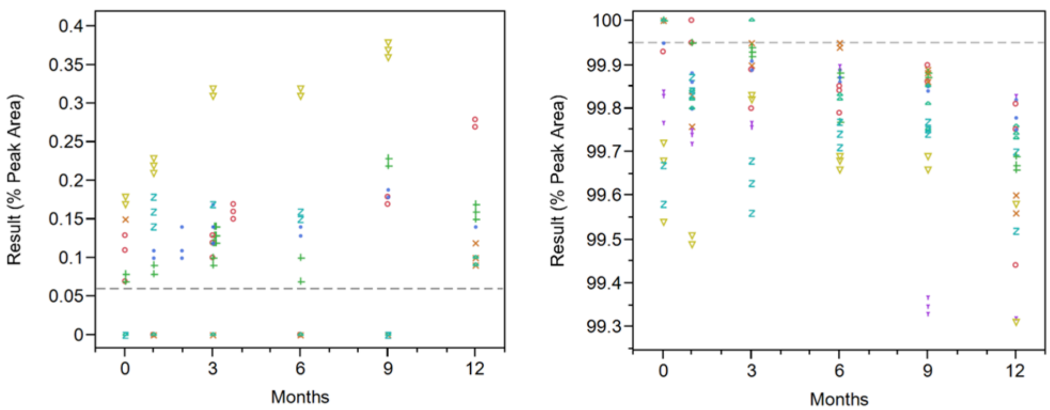

Figure 2: (Left plot) left censored data, censored at C=0.06%. (right plot) Right censored data censored at 99.95%. Data below (above) the censoring limit is set to 0% (100%).

The right plot in Figure 2 shows an example of right censoring that can also occur, though is less common. In this case, the outcome is a % purity measure as measured as 100% − % of peak area of an impurity. Since the impurity cannot be detected below 0.05%, the purity cannot be detected above 99.95%.

Several *ad hoc* methods are used when analyzing censored data that can lead to bias. For example, plugging in a value for the censored data such as C, C/2, $C/\sqrt{2}$, 0, where C = the censoring limit, tends to underestimate the steepness of trend and underestimate the variation. Another approach is to simply model only the time points with completely observed data. If there are enough data available, the uncensored data should give an approximately unbiased estimate of the trend.

To address the bias problem while accounting for test-to-test variation, a mixed model implementation of the Tobit Type I model is proposed in section 2. In section 3 the bias and variance of the method is compared to the other *ad hoc* methods in simulations that demonstrate improved properties of the model parameters when using the Tobit method. A bias problem in the test-to-test component due to ML estimation is noted and addressed with a related method. Section 4 and 5 conclude with further discussion about this bias problem and other methods.

2. Tobit Models

2.1 Tobin's Type I Model

Tobin (1958) introduced a simple concept for modeling censored data. His method, now called the Tobit Type I model, imagines that the regression trend could continue below the censoring limit as if there were no censoring limit. It does away with the censoring problem (temporarily) by supposing a simple latent model (2) for all data, censored and uncensored.

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i^*, \quad e_i^* \sim N(0, \sigma^2) \tag{2}$$

The latent values, of course, are not valid data below the censoring limit, so the Tobit model links the latent data to actual data by censoring it as in (3), where latent values y_i^*

below C are rectified to C and those above C are left as is. That is, the Tobit model assumes that $y_i = C$ in the data is representative of a censored value $y_i^* \leq C$.

$$y_i = CI(y_i^* \leq C) + y_i^* I(y_i^* > C) \quad (3)$$

The censoring in (3) implies that the distribution of y_i is a rectified normal distribution, rectified at the value C. As the latent model approaches C and eventually falls below C, the normal error distribution below C is cumulated into a discrete probability point (Figure 3, right plot). That is, the distributional model for the data is the discrete-continuous distribution (4).

$$F(y_i) = I(y_i > C)\Phi\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\sigma}\right). \quad (4)$$

The Tobit model can similarly be defined for an upper censoring limit or both upper and lower censoring limits. From this point forward we will only work with the lower censoring limit case as this is much more common in stability than other types of censoring.

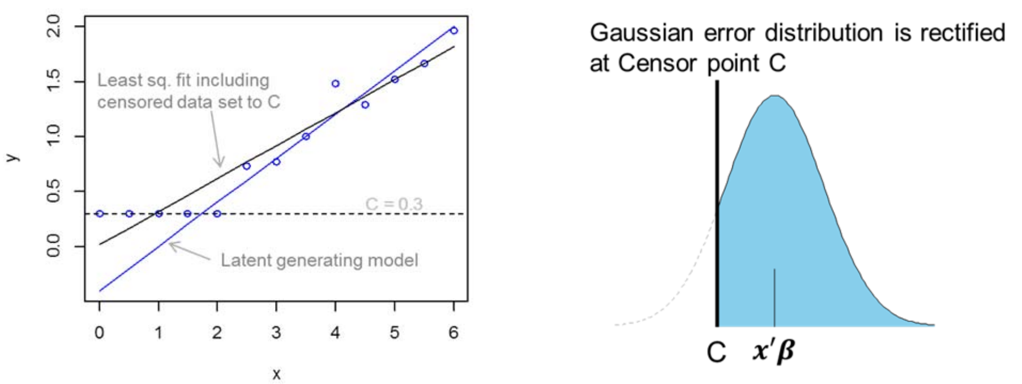

Figure 3: (Left Plot) illustration of data censored at censor limit C=0.3 generated by a latent model (blue) with least squares fit (black). (Right Plot) Latent regression model normal distribution rectified at point C.

2.2 Tobit Mixed Model for Stability Data

The Tobit model is applied to stability data as shown in Model 2. Model 2 takes advantage of the latent model concept of the Tobit model by specifying Model 1 as the latent model, then rectifying the latent result per the censoring limit. An (optional) monotonic transformation g may be applied to the result to meet regression assumptions of normality and constant variance, or to address a 0 boundary in addition to the censoring limit.

Model 2
$Y_{ij}^* = \mu(\text{Lot,Temp,Months}) + e_{\text{Test}_i} + e_{\text{Residual}_{ij}}$, where
$\mu(\text{Lot,Temp,Months}) = (\mu_0 + \text{Lot}) + (1 + \text{Lot} + \text{Temp} + \text{Lot} * \text{Temp})f(\text{Months})$,
factors Lot and Temp are appropriately expanded into dummy variables,
$e_{\text{Test}_i} iid\ N(0, \sigma_{\text{Test}}^2)$, and independently $e_{\text{Residual}_{ij}} iid\ N(0, \sigma_{\text{Residual}}^2)$, and
$g(\text{Result}_{ij}) = Y_{ij} = \begin{cases} Y_{ij}^* & \text{if } Y_{ij}^* > g(C) \\ g(C) & \text{if } Y_{ij}^* \le g(C) \end{cases}$,
for monotonic transformations g and f. (Model 2)

A useful way to express the Model (2) density for Y_{ij} is in a two-stage form (5), conditioning on the test random effect.

$$Y_{ij}|e_{\text{Test}_i} \sim f_1(y_{ij}|e_{\text{Test}_i}) =$$
$$\begin{cases} N\left(y_{ij}|\mu(\text{Lot,Temp,Months}) + e_{\text{Test}_i}, \sigma_{\text{Residual}}^2\right) & \text{if } y_{ij} > g(C) \\ \Phi\left(\frac{g(C) - \mu(\text{Lot,Temp,Months}) - e_{\text{Test}_i}}{\sigma_{\text{Residual}}}\right) & \text{if } y_{ij} = g(C) \end{cases}$$

$$e_{\text{Test}_i} \sim f_2(e_{\text{Test}_i}) \quad = N(0, \sigma_{\text{Test}}^2) \tag{5}$$

The conditional distribution $Y_{ij}|e_{\text{Test}_i}$ is given in two parts, a normal density for a linear model regression if the (transformed) result y_{ij} is above the (transformed) censoring limit, and a discrete part that cumulates the area of the normal density at or below the censoring limit. Given that e_{Test_i} is fixed, it simply becomes part of the mean function along with the stability model mean $\mu(\text{Lot,Temp,Months})$. It is understood by this notation that integration of f_1 involves summing discrete parts and integrating continuous parts.

2.3 Likelihood Estimation

Model specification is comparatively easy compared to model estimation. The likelihood for the Tobit mixed model Mode 2 is given by (6):

$$L = \prod_{i=1}^{\#tests} \int_{-\infty}^{\infty} f_3(y_i|e_{T_i}) f_2(e_{T_i}) de_{T_i}, \text{ where}$$
$$f_3(y_i|e_{T_i}) = \prod_{j=1}^{n_i} f_1(y_{ij}|e_{T_i}),$$
$$y_i = (y_{i1}, y_{i2}, \dots, y_{in_i})' . \tag{6}$$

The density $f_3(y_i|e_{T_i})$ is formed by recognizing that the test vector y_i is conditionally independent when conditioning on the test effect e_{T_i}. f_3 is then multiplied by $f_2(e_{T_i})$ to form a joint density on y_i and e_{T_i} from which e_{T_i} is integrated out to generate the marginal density of the test vector y_i. The tests themselves are assumed independent, so the overall likelihood is formed as the product of the marginal test vector densities.

Estimation with the likelihood in this form requires numerical integration techniques. SAS proc NLMIXED can fit this model when provided with initial estimates. The initial estimates can be taken from the fit to the uncensored data. Example SAS code is given below.

```
proc nlmixed data=stability;
  Parms b0=-0.1 b1=0.033 Cov_test=0.05 Cov_Residual=0.05;
  bounds Cov_Residual>0, Cov_Test>0;
  C     = 0; *censor limit;
  linp  = b0 + b1*Mo + Test; *linear model of the mean;
*normal CDF function for point mass at censor boundary C, conditional on Test
random effect;
  Phi   = probnorm((C-linp)/sqrt(Cov_Residual));
*normal pdf for data > C, conditional on Test random effect;
  f = exp(((Result-linp)**2)/(-2*Cov_Residual))/sqrt(2*3.141592654*Cov_Residual);
*log discrete-continuous mixture likelihood, conditional on Test random effect;
  ll = (censor=0)*log(f) + (censor=1)*log(Phi);
  model Result ~ general(ll);
*random effect for sample/test variance component;
  random Test ~ normal(0,Cov_test) subject=TestID ;
run;
```

2.4 Mean Model

Although the stability model is linear after taking transformations possibly in response or time, the mean model for the Tobit model is non-linear due to the censoring artifact and the convention that the censored data be set to the value C. Recalling that the Tobit model is a discrete-continuous mixture distribution, the approximate marginal expectation of a result is given by (7) (see the Appendix for details). It is clear from expression (7) that as the regression moves away from (above) the censoring limit, z becomes more negative and hence the expectation in (7) on the transformed scale g approaches the usual linear stability model. On the other hand as the regression moves toward and below the censoring limit, z becomes positive and the expectation in (7) on the transformed scale g approaches the censoring limit C.

$$E[\text{Result}] \cong g^{-1}\left(g(C)\Phi(z) + \frac{\sigma}{\sqrt{2\pi}}Exp(-0.5z^2) + \mu \times (1 - \Phi(z))\right), \text{ where}$$
$$z = \frac{g(C)-\mu}{\sigma},$$
$$\mu = \mu(\text{Lot,Temp,Months}) = (\mu_0 + \text{Lot}) + (1 + \text{Lot} + \text{Temp} + \text{Lot} * \text{Temp})f(\text{Months}),$$
$$\sigma = \sqrt{\sigma_{Test}^2 + \sigma_{Residual}^2} \quad (7)$$

It is important to note that setting the censored data to the value C for the Tobit model is not the same as doing so for an ordinary regression. Recall that the Tobit model explicitly accounts for the censoring as cumulative area of the latent model distribution below C. Also, while the approximate expectation in (7) is for the Tobit model data set, it does not necessarily reflect the true mean of the assay results near the censoring limit. Ultimately these are unknown.

The mean function (7) is illustrated on example stability data in Figure 4. The assay for this example had a very low detection limit close to 0. The mean model is approximately the linear latent model when it is away from the censoring limit, but becomes non-linear and limits toward the censoring limit as the latent model (not shown) approaches and drops below the censoring limit.

The Tobit model also provides an estimation of the censoring rate, $\Phi(z)$. The right plot in Figure 4 illustrates the rapid drop in censoring rate for some of the stability trends as they move away from the censoring limit. Interestingly, the censoring rate can increase if the latent model (not shown) has a negative trend. The negative latent trend in this model is a reflection of the partial censoring at early time points, followed by fully censored data at later time points. Likely, the true trend is flat with a constant rate of censoring.

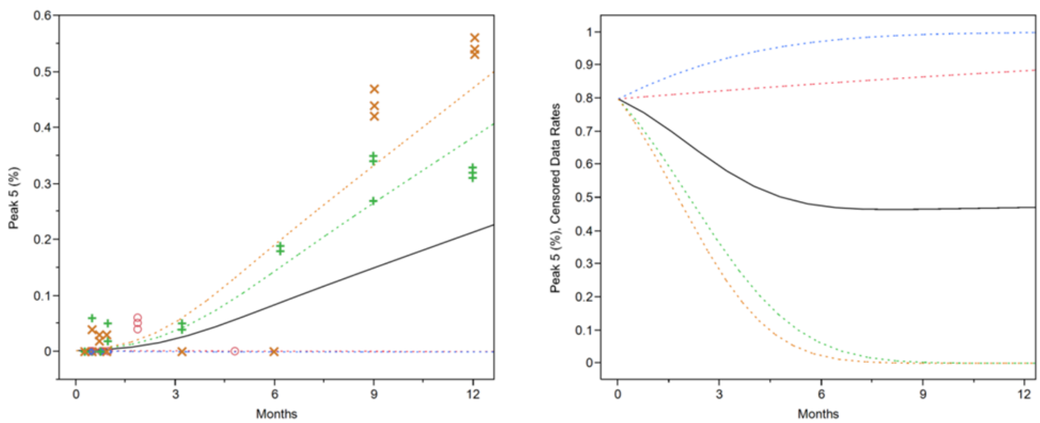

Figure 4: (Left Plot) Approximate expectation based on % peak area stability data fit with a Tobit mixed model. (Right Plot) Plots showing the estimated censoring rates for the fitted models.

3. Simulation Study

A simulation study was conducted to understand the performance of the Tobit mixed model. The Tobit method was compared to other ad-hoc approaches such as plugging in censored values and analyzing with multiple regression, fitting the uncensored data only, and a conditional approach that is explained below.

3.1 Methods compared via simulation

The following methods were compared:

1. **Tobit**: The Tobit mixed model (Model 2)
2. **Plug in**: mixed model fit to data with a single imputation of censored values set to the censoring limit C
3. **Uncensored**: mixed model fit only to the time points that have no censoring
4. **Conditional**: use of the uncensored method to obtain REML estimates of the variance components, followed by Tobit maximum likelihood estimation conditional on the estimated variance components.

Uncensored is only a viable option if the stability slope is sufficiently steep such that a meaningful number of time points remain uncensored. This method cannot be exactly unbiased when taking into account the selection process that identifies uncensored time points, but it would be expected to be approximately unbiased since upon knowing the time points that are unlikely or very unlikely to be censored, the censoring problem is essentially not present. So the issue with this method is not bias, but rather higher variance as the estimates will be less precise being based on less data.

Along these lines, it is of particular interest how much the precision of trend is improved in the Tobit method, having utilized the partial and fully censored data to help inform the model estimates, especially the precision of the slope and predictions at a future time points. The simulation also aims to understand how this precision depends on the degree of censoring and the between and within variance of test results.

3.2 Simulation Set up

The simulation involved two batches tested at time points 0,1,3,6,9,12, and 18 months. One test per time point per batch with two replicates per test were generated using latent model Model 3, with censoring limit C=0.5.

$Y_{ij}^{*} = \beta_0 + 0 * I(Batch\ 2) + 0.3 * Month + 0 * Month * I(Batch\ 2) + e_{\text{Test}_i} + e_{\text{Residual}_{ij}}$, where $e_{\text{Test}_i} \sim N\left(0, \sigma_{\text{Test}}^2\right)$, $e_{\text{Residual}_{ij}} \sim N\left(0, \sigma_{\text{Residual}}^2\right)$, and $I(Batch\ i)$ is an indicator function indicating Batch i. (Model 3)

Two values of $\beta_0 = -0.1, -0.7$, were used to invoke two mean levels of censoring of 28% and 41%, respectively. Two levels of between and within variance ratio, 1 and 4 were set with the total variance, $\sigma_{\text{Test}}^2 + \sigma_{\text{Residual}}^2$ held fixed at 0.1. A variance ratio of 1 corresponds to a cluster correlation of 0.5 and a variance ratio of 4 corresponds to a cluster correlation of 0.8. N = 5,000 simulations were performed.

3.3 Simulation Results and conclusions

The simulation results (Table 1) are divided into a number of sections. The bottom rows of the table focus on prediction at 24 months (6 months past the range of the data) and the top portion of the table considers the parameter estimates themselves. These sections are further divided into 4 sections of variance ratio 1 and 4 and mean censoring rates of 41% and 28%.

Considering the parameter estimates first, it is noted that the plug in method has a high degree of bias from the true parameters for every parameter except those contrasting the batch 1 and 2 slope difference (b2 and b4). The plug in method is clearly not recommended. Next, note that the Monte Carlo expectations of the uncensored method have low bias, with a few exceptions for the intercept. It is not clear why the lower rates of censoring tend to have more intercept bias.

Similarly, the Tobit mixed model has generally low bias. However, the Tobit model also has a disturbing pattern of underestimation of the test-to-test variation that appears to be slightly worse with higher rates of censoring. Much of this bias can be explained by the use of maximum likelihood for estimation instead of REML. Indeed, estimation of the latent model data with ML would also show an underestimation of test-to-test variation. For linear mixed models with balanced data, REML corrects for the degrees-of-freedom (DF) instead of the sample size. If one corrects the test-to-test variance component estimate by the sample size to DF ratio (= 14 tests/10 DF = 1.4x), some of the bias, but not all, is eliminated. Unfortunately we know of no simple way to implement REML in the nonlinear Tobit model context. Current experience suggests that this bias does not have a meaningful impact on the mean and confidence intervals of the mean. For example, prediction at 24 months in the simulation produced low bias for all methods except the plug in approach.

The bias in the test-to-test variance component motivates a hybrid conditional approach which uses REML to estimate the variance components on the uncensored data, but maximizes the likelihood for the Tobit model conditional on these variance component estimates. While there is some information loss in not utilizing all the data for the variance components, it is expected that the censored data provide less information about the variation than the mean. As Table 1 shows, the Monte Carlo expected values of the parameter estimates for the conditional method have overall little bias.

Var. Ratio	Mean Censor Rate	Parameter	True	tobit mean	plug in mean	Uncens. mean	Cond. tobit mean	tobit SE	plug in SE	Uncens. SE	Cond. tobit SE
1	41%	Cov_Residual	0.05	0.05	**0.03**	0.05	0.05	0.02	0.01	0.02	0.02
		Cov_Test	0.05	**0.02**	**0.15**	0.05	0.05	0.03	0.05	0.05	0.05
		b0	-0.7	-0.71	**0.07**	**-0.67**	**-0.74**	0.30	0.07	0.38	0.30
		b1	0	0.01	0.00	0.00	0.01	0.42	0.10	0.54	0.42
		b2	0.3	0.30	**0.24**	0.30	0.30	0.03	0.01	0.03	0.03
		b3	0	0.00	0.00	0.00	0.00	0.04	0.02	0.04	0.04
	28%	Cov_Residual	0.05	0.05	**0.03**	0.05	0.05	0.02	0.02	0.02	0.02
		Cov_Test	0.05	**0.03**	**0.07**	0.05	0.05	0.03	0.03	0.05	0.05
		b0	-0.1	-0.10	**0.22**	**-0.05**	-0.12	0.22	0.09	0.31	0.22
		b1	0	-0.01	0.00	-0.01	-0.01	0.30	0.13	0.43	0.30
		b2	0.3	0.30	**0.27**	0.30	0.30	0.02	0.01	0.03	0.02
		b3	0	0.00	0.00	0.00	0.00	0.03	0.02	0.04	0.03
4	41%	Cov_Residual	0.02	0.02	**0.01**	0.02	0.02	0.01	0.01	0.01	0.01
		Cov_Test	0.08	**0.04**	**0.17**	0.08	0.08	0.03	0.06	0.07	0.07
		b0	-0.7	**-0.73**	**0.07**	-0.69	**-0.76**	0.34	0.08	0.42	0.33
		b1	0	0.00	0.00	0.01	0.01	0.48	0.11	0.59	0.47
		b2	0.3	0.30	**0.24**	0.30	0.30	0.03	0.01	0.03	0.03
		b3	0	0.00	0.00	0.00	0.00	0.04	0.02	0.05	0.04
	28%	Cov_Residual	0.02	0.02	**0.01**	0.02	0.02	0.01	0.01	0.01	0.01
		Cov_Test	0.08	**0.05**	**0.09**	0.08	0.08	0.03	0.04	0.06	0.06
		b0	-0.1	-0.10	**0.21**	**-0.05**	**-0.13**	0.24	0.10	0.32	0.24
		b1	0	-0.01	0.00	0.00	-0.01	0.34	0.15	0.46	0.34
		b2	0.3	0.30	**0.27**	0.30	0.30	0.02	0.02	0.03	0.02
		b3	0	0.00	0.00	0.00	0.00	0.03	0.02	0.04	0.03
1	41%	Pred. 24 mo.	6.50	6.51	5.83	6.48	6.54	0.38	0.29	0.42	0.38
	28%	Pred. 24 mo.	7.10	7.10	6.81	7.06	7.12	0.34	0.29	0.39	0.35
4	41%	Pred. 24 mo.	6.50	6.52	5.82	6.49	6.54	0.42	0.32	0.46	0.42
	28%	Pred. 24 mo.	7.10	7.10	6.81	7.06	7.12	0.37	0.31	0.41	0.37

Table 1: Simulation results comparing for methods: **Tobit**: Tobit mixed model, **Plug in**: mixed model with censored data set to C, **Uncens.**: mixed model fit to only the timepoints that are uncensored, **Cond. Tobit**: REML on uncensored, Tobit conditional on REML var comp. The Monte Carlo expectations and standard errors are reported at the worst observed precisions of the simulations, meaning that any changes in the values at the reported precision reflect more than 2x standard error shifts with respect to Monte Carlo error. The expectations with more noteworthy bias are in bold. About 99.9% of the Tobit fits converged and >97% of the conditional Tobit fits converged among 5,000 simulation runs.

As hoped, utilization of the censored data improves the precision of the parameter estimates beyond the precision that can be achieved modeling only the uncensored data. In all cases the parameter standard errors for the Tobit model were as small or smaller than the uncensored method standard errors. The conditional method also had this property. For the prediction of the mean at 24 months, we see an approximate 10% reduction in standard error for every case for both the Tobit and Conditional methods. While the lowest standard errors are achieved by the plug in method, the bias of this method dominates in comparisons of mean squared error, with both the Tobit and Conditional Tobit methods performing the best.

4. Other Methods and Future Research

For a model with one level of nesting, as with the Tobit model here, numerical integration methods can be used to maximize the likelihood. However, for more complex models, for example a model accounting for random batch effects and test-to-test variation, the integrations become too difficult to do numerically. Hughes (1999) derived the EM algorithm for linear mixed models with censored data and proposed computing expectations for the E step using Gibbs sampling, i.e., using simulation instead of direct numerical integration. Hughes derived both ML and REML methods of estimation, but in his simulations he evaluated only the ML method and noted as I did that the variation of the between variance component is underestimated. He attributed all of the bias to the use of maximum likelihood instead of REML, though he did not use the REML version of his method to improve the estimates.

Another common method addressing missing data is multiple imputation. Multiple imputation, unlike single imputation, accounts for the uncertainty in the model due to censoring. The difficulty is in specifying the model from which imputations will be generated. If there are enough data, modeling the uncensored data may be a good choice, though with a smaller sample size, the imputed values may have excessive variation. The Tobit model should not be used as this already underestimates the test-to-test variation. Wu (2010) proposed using a Bayesian posterior predictive distribution for imputation, although it seems one should move entirely to the Bayesian context in this case.

Since most methods either rely on simulation, or need to if more than two error terms are included in the model, it seems obvious that Bayesian methods should be considered. Gelman et. al. (2004) showed how to implement censoring into a Bayesian model using the same fundamental idea of the Tobit model; cumulating area of the latent density beyond the censoring point. More recently, Dagne and Huang (2013) have proposed a Bayesian method using the Tobit idea, though in a much more complex setting involving a mixture model for censored observations and covariates with measurement error.

In all of these methods, a key question that needs to be answered with respect to the stability model is whether the test-test variation is underestimated as was shown with the Tobit mixed model, and possibly, whether higher level variance components such as batch-to-batch variation are underestimated.

5. Conclusions

Test to test variation, usually as the result of assay run-to-run variation is common in stability studies. Sometimes censoring plays a role in stability too, usually left censoring due to an assay limit of quantitation or a limit of detection. A Tobit mixed model was proposed as a suitable model to account for both test-to-test variation and censoring.

Through simulations we observed that the Tobit model nearly eliminated all bias observed in the plug in, single imputation method that plugs in the censoring limit for the censored values. However, the Tobit model was also observed to underestimate the test-to-test variation. This bias appeared to have little impact on predictions at 24 months, nevertheless, an alternative conditional method was proposed that used REML on the time points with only fully observed data then maximized the likelihood of the Tobit model conditional on these variance component estimates. The conditional method was observed to eliminate the test-test variance bias of the Tobit method, but retained Tobit 's higher precision of model parameters and predictions compared to the uncensored method.

Future research is needed to compare the Tobit mixed model to other methods including EM estimation for mixed models, multiple imputation methods, and Bayesian methods. Of keen interest is whether the test-to-test variance component is underestimated by any of these methods. It is recognized that extensions to additional variance components, such as batch-to-batch variation significantly complicates all of the methods, except perhaps multiple imputation. Additional variance components effectively require most methods to use simulation in one form or another.

References

Dagne and Huang (2013) "Bayesian semiparametric mixture Tobit models with left censoring, skewness, and covariate measurement errors," Statistics In Medicine, 32, 3881–3898.

Gelman A, Carlin J, Stern H, Rubin D (2004), *Bayesian Data Analysis*, Second Edition. Chapman & Hall/CRC.

Hughes (1999), "Mixed Effects Models with Censored Data with Application to HIV RNA Levels," Biometrics, 55, 625-629

Tobin J (1958). "Estimation of relationship for limited dependent variables." Econometrica, 26, pp. 24-36.

Wu, L (2010). *Mixed Effects Models for Complex Data*. CRC Press.

Software

Henningsen (2010) "Estimating Censored Regression Models in R using the censReg Package," R package vignettes.

SAS proc LIFEREG, SAS proc QLIM, SAS proc NLMIXED, SAS Insititute, Inc.

Kiernan, Tao, and Gibbs (2012) "Tips and Strategies for Mixed Modeling with SAS/STAT® Procedures," SAS Global Forum, SAS Institute Inc., Cary, NC, USA.

Appendix

Details of the expectation approximation (7) are shown here. First approximate the Tobit mixed model, Model 2, using the marginal distribution of the latent model and set this marginal distribution up as a Tobit model. This yields the univariate discrete-continuous mixture model

$f(y_{ij}) = \begin{cases} N(y_{ij}|\mu, \sigma^2) & \text{if } y_{ij} > g(C) \\ \Phi\left(\frac{g(C)-\mu}{\sigma}\right) & \text{if } y_{ij} = g(C) \end{cases}$, where

μ is the latent mean model and $\sigma = \sqrt{\sigma_{Test}^2 + \sigma_{Residual}^2}$. Then calculate

$$E[Y_{ij}] = \int_{-\infty}^{\infty} y_{ij} f(y_{ij}) dy_{ij}$$

$$= g(C)\Phi\left(\frac{g(C)-\mu}{\sigma}\right) + \int_{g(C)}^{\infty} y_{ij} N(y_{ij}|\mu, \sigma^2) dy_{ij}$$

The latter quantity is the upper partial expectation of the normal, and via change of variables can be shown to be

$\mu\left(1 - \Phi\left(\frac{g(C)-\mu}{\sigma}\right)\right) + \frac{\sigma}{\sqrt{2\pi}} \int_{z(g(C))}^{\infty} z(y_{ij}) e^{-\frac{1}{2}z(y_{ij})^2} dz(y_{ij})$, where $z(y) = \frac{y-\mu}{\sigma}$.

For the second part, another change of variables yields $\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{g(C)-\mu}{\sigma}\right)^2}$, thus,

$E[Y_{ij}] = g(C)\Phi(z(g(C))) + \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2}(z(g(C)))^2} + \mu\left(1 - \Phi(z(g(C)))\right)$.