

ALTERNATIVE SHELF LIFE ESTIMATION METHODOLOGIES

Walt Stroup and Michelle Quinlan, Department of Statistics, University of Nebraska,
Lincoln, NE 68583-0963, wstroup@unl.edu

Abstract: PQRI Stability Shelf Life Working Group has worked to advance statistical methods to estimate shelf life. We develop a modeling context that allows forming a quality statement. Crucial insights: 1) there is a batch response distribution that projects to shelf life distribution – batch effects are, by definition, random; 2) quality statement suggests targeting a lower shelf life distribution quantile: statistical methods should target this lower limit. Quinlan (2010) developed procedures to translate shelf life distribution quantile to corresponding criteria for mixed model based tolerance interval and quantile regression. Comparing shelf life estimation methods reveals the limitations of current practice, why unadjusted mixed model procedures are unsuitable and demonstrate the promise of mixed model based quantile regression and tolerance interval estimation.

Key Words: ICH Guidance for Shelf Life Estimation; Batch Response Distribution; Batch Shelf Life Distribution; Linear Mixed Model; Tolerance Interval; Quantile Regression

1. Introduction

The PQRI Stability Shelf Life Working Group’s experience underlines the fact that, as an essential pre-condition for selecting statistical methodology to estimate shelf life, the modeling context needs to be carefully articulated. Our discussions have established the modeling context as unusually subtle and subject to misunderstanding. Unless these subtleties are well-understood, the consequences of competing methodologies – what they *actually do* versus what we would *like to think they are doing* – are difficult to characterize.

George Bernard Shaw famously said, “The Americans and English are two peoples separated by a common language.” So it has often seemed in our working group’s internal discussions as well as our interaction outside the group. Sandell (2010) described the need for a common vocabulary regarding shelf life and the current lack of such a vocabulary. Terms associated with shelf life estimation – e.g. “product shelf life,” “distribution of the shelf life mean,” “tolerance interval,” “quantile,” etc. – do not necessarily mean the same thing to all who work in this area. As a result, conversations about shelf life are prone to going astray unless participants take care at the beginning of the discussion to make sure that they agree on a common understanding of all key terminology. We have repeatedly been surprised at how often this common understanding *is not there*, even among people who have been working in the shelf-life area for years. Sandell suggested a standardized terminology intend to alleviate these problems.

Given that we have an unusually complex modeling problem that is unusually subject to differing understandings of key terminology and therefore unusually subject to breakdowns in communication, we begin this presentation with a review of the shelf life modeling context. In this review, we emphasize visual illustration and avoid the use of technical terms – especially those contaminated by a history of multiple definitions – wherever possible.

Littell, et al. (2006), on page 2 of *SAS for Mixed Models*, write, “...Statistical models ... are mathematical descriptions of how the data are produced and how they vary from

observation to observation.” This means clearly identifying the *processes* and their associated *probability distributions* that give rise to the data we observe. In developing the modeling context presented here, the following premises guided the Working Group:

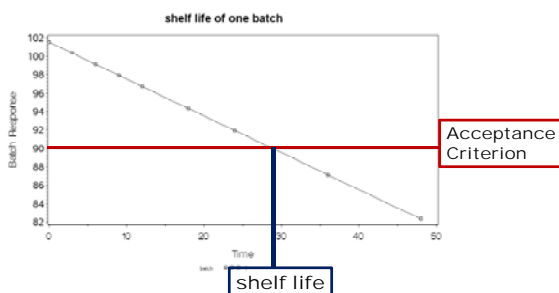
- the two primary processes giving rise to observations are
 - variation among batches
 - changes in a batch over time
- the observed response is a function of these two processes – mathematically, $response = f(batch, time)$
- following ICH guidance, changes over time follow a linear regression

In principle, our modeling context can be extending to nonlinear regression over time, but in this presentation we limit our focus to the linear case.

2. Modeling Shelf Life

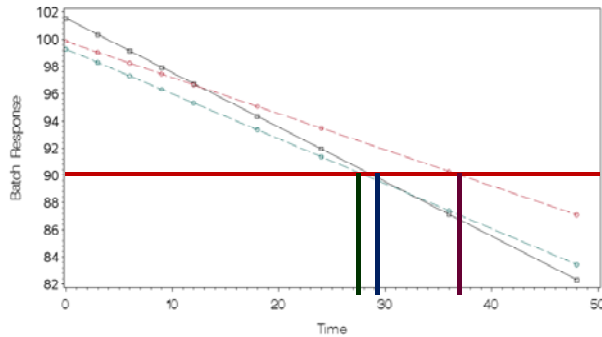
We begin by considering a single batch. Consistent with the premises listed above, the mean response of this batch changes linearly over time – that is, $\mu = \alpha + \beta X$, where μ denotes the expected batch response, α denotes the intercept (the expected batch response at time 0), β denotes the slope and X denotes time. If the stability-limiting response is a characteristic that decreases over time (e.g. efficacy) until it reaches the point at which it is no longer acceptable, then we would expect $\beta < 0$. If the stability-limiting characteristic increases over time until it exceeds an acceptable upper limit, then we would expect $\beta > 0$. Subsequent illustrations use the decreasing-response case, but can easily be visualized for the increasing-response case as well. Figure 1 illustrates the shelf life context for a single batch.

Figure 1. Illustration of the Acceptance Criterion, Response of a Single Batch over Time and Resulting Shelf Life



Letting A denote the acceptable limit, i.e. the lowest acceptable response, then the shelf life of this batch is $\frac{A - \alpha}{\beta}$. Denote this by SL .

Forenzo (2010) described the ICH guidelines for stability testing for shelf life. Current practice in shelf life estimation, following ICH guidance, typically entails observing 3 batches. Figure 2 illustrates the modeling context of such studies.

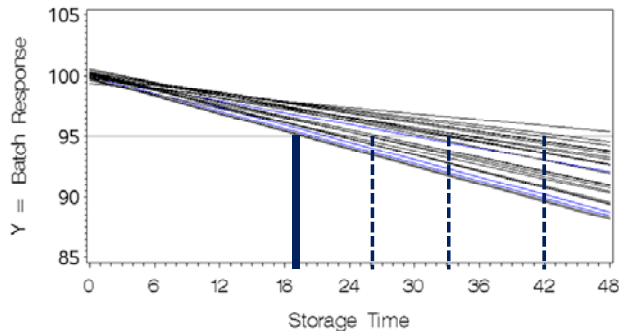
Figure 2. Illustration of Responses of 3 Batches in Stability Study

In figure 2, we denote the expected response over time for the i^{th} batch as $\mu_i = \alpha_i + \beta_i X$.

Thus the shelf life for the i^{th} batch is $SL_i = \frac{A - \alpha_i}{\beta_i}$. The crucial insight from Figure 2 is

that, in general, batches do not have identical intercepts or slopes and hence do not have identical shelf lives. Estimating shelf life requires a decision about how to understand variation among shelf lives of the observed batches.

The 3 batches observed in a given stability study represent a larger population of batches. There is no reason why *these 3* batches were observed; in principle, *any 3* that represent the population will do. More importantly, following ICH guidance, these 3 batches *must* represent *all future* batches stored under similar conditions. In other words, this group of 3 batches is a sample from a population. The population can be visualized as in Figure 3.

Figure 3. Illustration of a Population of Batches.

The key points drawn from Figure 3 are

- the responses over time vary among batches
- as a result, shelf lives vary among batches

We can frame this in probability language as follows. The intercept and slope parameters each have a probability distribution, i.e. for each batch, the random vector $\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}$ has a

distribution. Under linear model theory for Gaussian responses, for example,

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} \sim NI \left(\begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix}, \begin{bmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{bmatrix} \right) \text{ where "NI" signifies a normal distribution with mutually}$$

independent batches, μ_α denotes the population mean of the intercepts, μ_β denotes the population mean of the slopes, σ_A^2 denotes the variance among batch intercepts, σ_B^2 denotes the variance among batch slopes, and σ_{AB} denotes covariance between batch intercepts and slopes. It follows that the batch shelf lives, $SL_i = \frac{A - \alpha_i}{\beta_i}$ follow a

distribution determined by the joint distribution of $\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}$.

The probability model for the expected batch responses can alternatively be characterized as $\mu_i = \mu_\alpha + a_i + (\mu_\beta + b_i)X$, where $\begin{bmatrix} a_i \\ b_i \end{bmatrix} \sim NI\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{bmatrix}\right)$. This is important, as it establishes a rationale for modeling shelf life using random coefficient linear mixed models – at least as a starting point.

Note that the probability model describing the processes affecting stability limiting response specifically does *not* include within-batch variability. This is not an oversight. It is a deliberate omission. Omitting this term has been the subject of much controversy, both within the group and in our discussions external to the group. Our premise is that, for the purpose of shelf life estimation, there is a *single value* that characterizes a given batch at a given time. By the time a product has reached this point in its development, content uniformity will have been established, the manufacturing process should be under statistical control, and within-batch variance should be negligible. In any event, the acceptance limit, A , should have a built-in margin of safety so that any likely within-batch variance should be well within this margin.

3. Making Use of the Shelf Life Probability Model

Given the probability model illustrated in Figure 3, the shelf life question is this: how are we to understand shelf life – how do we define it – in terms of this model? Recall from the last section that batch shelf lives have a distribution that is a consequence of the distribution of batch response over time. Figure 4 illustrates the shelf life distribution as a projection of batch response distribution.

Figure 4. Relationship between Batch Response and Shelf Life Distribution

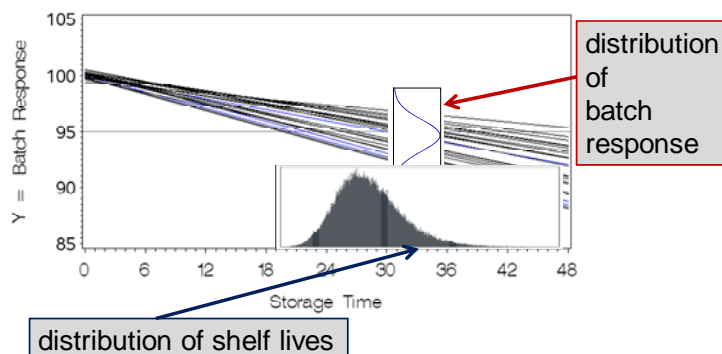


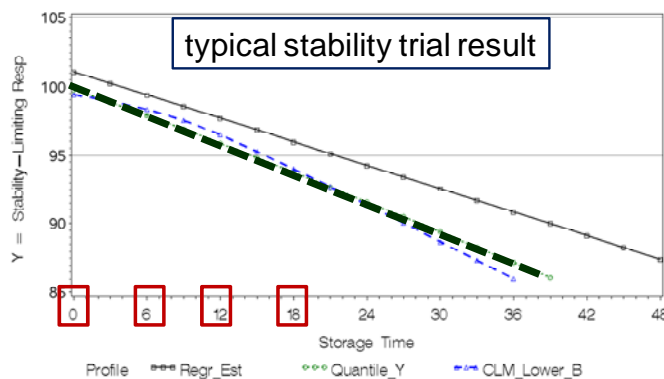
Figure 4 became the subject of lively debate within the working group about what is meant by shelf life. The crux of the debate focused on whether *mean* of the shelf life distribution or an agreed-upon lower *quantile* best captures the goal implicit in ICH guidance for shelf life estimation.

Ultimately, we used a mountain climbing metaphor to clarify the problem. Suppose you are climbing at extreme altitude, e.g. climbing Mt. Everest. Do you want the time your oxygen bottle is guaranteed to last to be characterized in terms of the *mean* of the distribution of oxygen bottle lifetimes or do you want it characterized in terms of the *lower limit* of that distribution? If the mean characterizes oxygen bottle lifetime, you have a 50-50 chance of taking an oxygen bottle that will not last the time it takes to summit and return safely to camp. If you use a lower quantile, there is an assignable probability (hopefully as close to 1 as is practical) that your bottle will last. If you value your life, using the lower quantile as the defining criterion seems obvious.

Because the consequences of an oxygen bottle not lasting are potentially fatal, this metaphor seems applicable to situations for which using a drug after time $SL_i = \frac{A - \alpha_i}{\beta_i}$ has serious health implications.

Visualizing the relationship between the distributions of batch response and batch shelf life provides insight into what information a regression-based confidence interval can and cannot provide for estimating shelf-life. Figure 5 shows the mean of the batch response distribution, the lower 5% quantile of that distribution and the expected result of estimating a 95% confidence interval about the regression line.

Figure 5. Illustration of Difference between lower Quantile of Batch Response Distribution and lower Regression-Based Confidence Limit



The red boxes identify the times (0, 6, 12 and 18 months) when observations have been taken. Solid line starting at $Y=102$ and decreasing to below 90 by month 48 is the population mean of the batch responses ($\mu_{\alpha} + \mu_{\beta}X$ using notation given above). The heavy green dashed line is the lower 5% quantile of the batch response distribution. In theory, from the above discussion, this (or possibly a different quantile, e.g. 1%) is our target. The lighter, blue dashed line with the hour-glass shape below the mean batch response line is the expected 95% regression-based confidence interval. Three features that have consequences for shelf-life estimation should be apparent:

1. The shape of the lower confidence bound over time does not correspond to the target – i.e. the lower quantile of the batch response distribution.
2. The width of the hour-glass associated with the 95% confidence interval is minimized at the mean of the times included in the stability trial (in this example, 0, 6, 12 and 18 months) then flares outward as months increase. The shape of the hour-glass is a function of the stability trial design – add, subtract or change the times of observation and the shape of the confidence bound profile changes.
3. Given (1) and (2) there is no transparent relationship between the target of the shelf life estimation exercise (the lower quantile) and the lower bound of the regression-based confidence interval.

Figure 5 actually depicts an optimistic relationship between the regression-based confidence interval and the lower quantile. In practice, the fixed-batch version of the regression $\mu_i = \alpha_i + \beta_i X$ can only be pooled into a single regression equation $\mu_i = \alpha + \beta X$ if we fail to reject H_0 : all $\alpha_i = \alpha$ and all $\beta_i = \beta$. However, because within-batch variance is typically negligible relative to among batch variance (σ_A^2 and σ_B^2 as defined above) only rarely will pooling be allowed (and when pooling is possible, it would appear to speak poorly of within-batch consistency). When pooling is not possible, we see behavior described in detail by Schwenke (2010): the regression-based lower confidence bound depends on the worst batch only, shifting the lower confidence profile to the left and angling it more sharply toward the X (month) axis, exacerbating the disconnect between the confidence bound and the lower quantile. Visualized this way – i.e. with the modeling context clearly laid out – it is easy to see that using fixed-batch, regression-based confidence bounds to estimate shelf life does *not* result in an estimator with a transparent relationship to the objective of the shelf life estimation exercise. Instead, in many cases, it amounts to extreme overkill.

4. What *Should* We Do? Alternative Shelf-Life Methodologies

Two primary realities emerge from the modeling context described in the previous section:

1. The batches observed in a stability trial represent a population and therefore have an associated probability distribution. In modeling term, batch effects are *random*, not *fixed*.
2. The target of shelf life estimation should be the lower bound of the shelf-life distribution.

While no statistical procedure can provide a 100% guarantee, there should be a transparent relationship between the shelf life estimation procedure and the target probability. In other words, when ICH guidance suggests a 95% confidence interval, this appears to imply that the target is the 5% quantile (as visualized via Figure 4). While we have established that the regression-based confidence interval does not dependably address this target, the target itself is not in question: this *is* a target that a preferable estimation method *should* address.

To address the first of these realities, we might ask what happens if we use a linear mixed model procedure, specifically the random coefficient model, given as follows. Let y_{ijk} denote the k^{th} observation on the i^{th} batch at time j . The random coefficient model is

$y_{ijk} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})X_j + e_{ijk}$ where β_0 and β_1 denote the mean intercept and slope parameter, b_{0i} and b_{1i} are the random batch-specific intercept and slope, assumed independent among batches and bivariate normal within batches, i.e.

$\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}\right)$, X_j denotes the j^{th} storage time and e_{ijk} denotes within batch error, assumed i.i.d. $N(0, \sigma^2)$. Note that this model corresponds to the modeling context depicted in Figure 3. Estimation of this model can easily be implemented by standard mixed model software, e.g. SAS[®] PROC MIXED or PROC GLIMMIX.

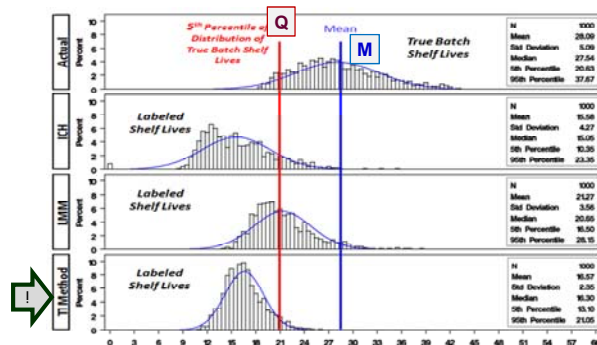
Unfortunately, standard mixed model procedures estimate the expected value of y_{ijk} . As visualized in Figure 4, this is not what we want. However, the parameters estimated by the linear mixed model procedure can be used either to implement mixed-model based tolerance interval procedures or quantile regression procedures using the variance component estimates from the mixed model procedure as weights. To do this accurately, however, we need to refer back to Figure 4. Figure 4 depicts the batch shelf life distribution as a projection of the batch response distribution onto the time axis. More precisely, the batch shelf-life distribution is the distribution of $SL_i = \frac{A - \alpha_i}{\beta_i}$. In other words, it is the distribution of the ratio of two correlated random variables, α_i and β_i .

Assuming normality, recall $\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} \sim NI\left(\begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix}, \begin{bmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{bmatrix}\right)$. Hinkley (1969) derived the

distribution of the ratio of correlated normal random variables. We can use this to determine the relationship between the batch response distribution and the batch shelf-life distribution. Quinlan (2010) showed that this relationship is not in lock-step in the sense that a given quantile of the shelf life distribution does not necessarily correspond to the same quantile of the batch response distribution.

The practical implication of Quinlan's result is that a mixed model tolerance interval or a quantile regression procedure using the lower 5% of the batch response distribution as a criterion does not necessarily target the lower 5% quantile of the batch shelf life distribution. In fact, in most practical situations such a strategy would result in a substantial underestimate of the target – exactly the kind of overkill we seek to avoid. However, if we have a reasonable idea of the parameters of the joint distribution of α_i and β_i we can determine the quantile of the batch response distribution that we should target given the quantile of the batch shelf life distribution that we are trying to estimate. Figure 6 shows simulation results from Quinlan (2010) showing the relative performance of the ICH-prescribed procedure, the unadjusted linear mixed model procedure using the random coefficient model described above, and a mixed-model based tolerance interval procedure using the relationship between batch response quantile and batch shelf-life quantile from Quinlan (2010).

Figure 6. Simulation Results Comparing Three Shelf Life Estimation Procedures



There are four distributions shown in Figure 6.

- The top (labeled “Actual”) is the distribution of batch shelf lives in a simulated population of 1000 batches, generated using parameters suggested by live data from stability trials collected by the working group.
- The second (labeled “ICH”) is the observed sampling distribution of estimated shelf lives applying the ICH guidance to each simulated data set. Note two undesirable features of the ICH sampling distribution: first, there are a number of extreme results – shelf life estimates of 0 or estimates well above the mean of the actual shelf life distribution; second, the proportion of shelf life estimates above the lower 5th quantile of the actual shelf-life distribution is well over 5%. Even though the ICH guidance purports to target the lower bound of the shelf life distribution, it clearly does not do so effectively. It also has a relatively high sampling distribution variance.
- The third (labeled “LMM”) is the sampling distribution of shelf life estimates based on unadjusted linear mixed model estimates. The sampling distribution is better behaved than ICH in the sense that the variance is lower and there are fewer extreme results. However, the estimate focuses on the mean of the shelf-life distribution. This would be desirable if the target for shelf life estimate was the mean of the shelf-life distribution, but it is not. A high proportion of shelf life estimates are above the lower 5th quantile of the shelf-life distribution.
- The fourth (labeled “TI Mixed”) is a mixed-model based tolerance interval procedure using Quinlan’s adaptation of Hinkley’s result. Specifically, the tolerance interval used the 20th quantile of the batch response distribution to target the 5th quantile of the batch shelf life distribution. The result is an estimator whose sampling distribution matches the variance and absence of extreme values of the mixed model procedure but has all but 5% of the shelf life estimates at or just below the target. This estimator is clearly superior to either the ICH or unadjusted mixed model estimates.

Quantile regression weighted by the batch intercept and slope variance components would procedure similar results to the mixed model based tolerance interval shown in Figure 6. Quinlan (2010) gives technical details of how to implement both of these estimation procedures. The main limitation of both the mixed model tolerance interval and the weighted quantile regression procedure is that both require reasonably good approximate knowledge of the parameters of the joint distribution of α_i and β_i in order to

approximate the relationship between the target quantile of the shelf life distribution and the quantile to be used in the estimation procedures. If these are not known and must be estimated based on the three batches prescribed in the ICH guidance, sampling distributions such as the last two shown in Figure 7 are typical.

Figure 7. Simulation Results Including Mixed Model Tolerance Interval and Quantile Regression Procedures that Require Estimates of the Response Quantile-to-Shelf Life Quantile Relationship based on information from 3 batches

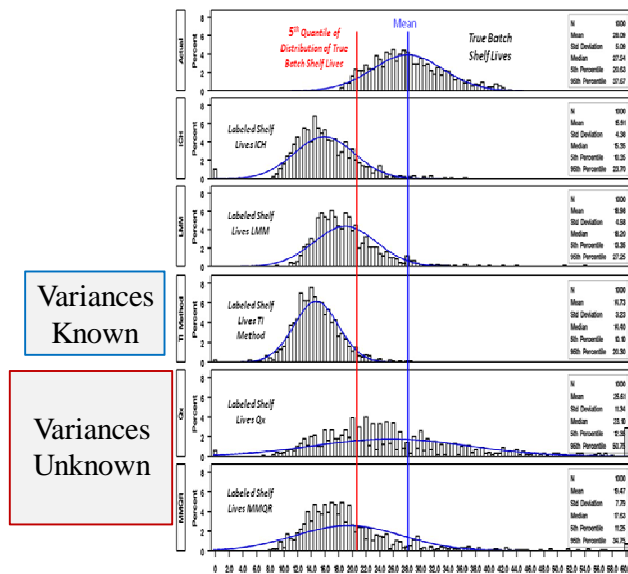


Figure 7 underlines the fact that, absent reasonably accurate information about the batch response distribution – specifically the random intercept and slope variances – accurate estimation of shelf life is problematic.

The good news about the mixed-model based procedures is that they reward obtaining additional data, whereas obtaining additional data under the ICH guidance is counterproductive in the sense that it simply makes a worse worst-case batch inevitable – a strong disincentive to obtaining additional information when otherwise warranted. The question is where would additional data come from? One obvious source is to observe more batches. This may be impractical and, in fact, it may be unnecessary. In drug development, the stability trial occurs late in the sequence of events. By this time, much is known about the drug. Specifically, in establishing content uniformity and getting the manufacturing process under statistical control, much is probably already known about the mean and variance of the joint distribution of the intercept and slope, α_i and β_i , of the batch response distribution. Assuming the availability of this information, it can be used to determine the batch response-to-shelf life quantile relationship needed to accurately estimate shelf life via mixed model based tolerance interval or variance-weighted quantile regression.

5. Summary and Conclusions

The main points established in this presentation are:

- The shelf life modeling context has to be clearly articulated prior to the selection of an estimation procedure suitable for determining labeled shelf life
- All batches are not created equal (at least, not exactly equal) – they have a probability distribution. The batches observed in stability trials represent a sample of a target population. By ICH guidance, that population includes all future batches.
- Therefore, in modeling terms, batch effects are, by definition, *random*, not *fixed*.
- Data are observed from the batch response distribution.
- The batch response projects to a distribution of batch shelf lives. In linear regression, this projection occurs through $SL_i = \frac{A - \alpha_i}{\beta_i}$. Assuming the joint distribution of α_i and β_i is bivariate normal and the parameters are known, the shelf life distribution can be determined from the batch response distribution.
- When the health implications are serious for a product whose storage time exceeds $SL_i = \frac{A - \alpha_i}{\beta_i}$, the appropriate target for defining shelf life is the lower limit of the shelf life distribution. Because this distribution is continuous, this means determination of labeled shelf life requires targeting an agreed-upon lower quantile.
- For the purposes of mixed model based quantile regression or tolerance interval estimation, the target quantile of the shelf life response distribution does not translate to the same quantile for the response distribution, but the relationship can be determined.
- Determining this relationship requires good information about the parameters of the joint distribution of α_i and β_i .
- Assuming this is available, both the mixed model, variance-weighted quantile regression and mixed model based tolerance interval procedures represent a big improvement over existing methods used to estimate shelf life.

The remaining issues appear to be 1) identifying processes to obtain the information needed to determine the shelf life quantile-to-response quantile relationship without adding burdensome extras to the current regulatory process and 2) developing “friendly” software to implement the mixed model tolerance interval and quantile regression procedures. The software would need to include determining the appropriate quantile for the estimation procedures to work with given the agreed-upon target quantile of the shelf life distribution.

6. References

- Christopher, David. 2010. The Philosophy and Intent of Stability Shelf Life. *JSM Topic Contributed Session: Pharmaceutical Stability Shelf Life: Philosophy, Intent and Estimation*.
- Forenzo, Pat. 2010. An Industry Perspective on Shelf Life and the PQRI Initiative. *JSM Topic Contributed Session: Pharmaceutical Stability Shelf Life: Philosophy, Intent and Estimation*.

Hinkley, D. V. 1969. On the Ratio of Two Correlated Normal Random Variables. *Biometrika* **24**: 428-440.

Littell, R. C., G. A. Milliken, O. Schabenberger, W. W. Stroup and R. D. Wolfinger. 2006. *SAS for Mixed Models, 2nd ed.* Cary, NC. SAS Institute, Inc.

Quinlan, Michelle. 2010. On Issues in the Estimation of Quantiles in the Presence of Random Effects with Applications to Shelf Life Estimate. Unpublished PhD Dissertation. University of Nebraska, Lincoln.

Sandell, Dennis. 2010. Some Reflections About Shelf Life. *JSM Topic Contributed Session: Pharmaceutical Stability Shelf Life: Philosophy, Intent and Estimation.*

Schwenke, James. 2010. Current Practices in Shelf Life Estimation. *JSM Topic Contributed Session: Pharmaceutical Stability Shelf Life: Philosophy, Intent and Estimation.*

Slides from Christopher, Forenzo, Sandell and Schwenke Topic Contributed presentations available on request.