University of California, Berkeley

U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2003 *Paper* 127

Rank Regression in Stability Analysis

Ying Qing Chen* Annpey Pong†

Biao Xing‡

*Division of Biostatistics, School of Public Health, University of California, Berkeley, yqchen@stat.berkeley.edu

†Department of Biostatistics, Forest Laboratories, Inc., Jersey City, NJ

‡Division of Biostatistics, School of Public Health, University of California, Berkeley

Rank Regression in Stability Analysis

Ying Qing Chen, Annpey Pong, and Biao Xing

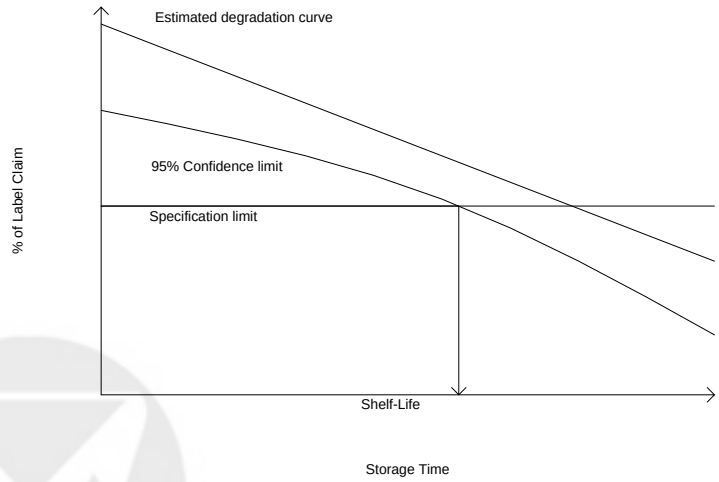
Abstract

Stability data are often collected to determine the shelf-life of certain characteristics of a pharmaceutical product, for example, a drug's potency over time. Statistical approaches such as the linear regression models are considered as appropriate to analyze the stability data. However, most of these regression models in both theory and practice rely heavily on their underlying parametric assumptions, such as normality of the continuous characteristics or their transformations. In this article, we propose and study some rank-based regression procedures for the stability data when the linear regression models are semiparametric with unspecified error structure. Numerical studies including Monte Carlo simulations and practical examples are demonstrated with the proposed procedures as well.

1 Introduction

According to the *Guidance for Industry: Guideline for Submitting Documentation for the Stability of Human Drugs and Biologics* by the Food and Drug Administration (FDA, 1987), a stability study is to determine the shelf-life within which the capacity of a drug product remains its identity, strength, quality, and purity. In the United States, the shelf-life is required by the FDA for every proposed market formulation or post approval market product. The stability data are hence often collected on the product samples stored in the controlled conditions to measure these properties changing over time. To be specific, if the drug capacity is hypothesized by a degradation curve in time as illustrated in Figure 1, the shelf-life is usually determined as the time point at which one of the 95% confidence bounds of the estimated degradation curve intersects the specification limit.

Figure 1: Illustration of shelf-life with decreasing mean degradation curve



Apparently there are two logical stages in the process of determining the reasonable shelf-lives:

(1) develop statistical models to characterize and estimate the product's outcomes of interest changing over time;

(2) determine the shelf-life based on the allowable specification limit of drug products.

Although stage (2) is the ultimate goal of the stability analysis, a sensible model in stage (1) is critical to the entire analysis. For the models in stage (1), as indicated in a draft consensus guideline of *Evaluation of Stability Data* recently issued by the International Conference on Harmonization (ICH) Steering Committee (2002), "regression analysis is considered an appropriate approach to evaluating the stability data for a quantitative attribute and establishing a retest period or shelf life" (p. 6). In fact, the linear regression models have been used extensively in both research and industrial practice. For example, in an article by Shao and Chow (2001), when the batch-to-batch variation of the tested product is not concerned, the following simple linear regression model can be used to describe the continuous characteristics over time:

$$Y = \alpha + \beta t + e, \tag{1}$$

where α and β are unknown parameters, t is the fixed measurement time determined by the study design and e is zero-mean normal deviate. When the batch-to-batch variation is concerned, the random effects models, which often assume that (α, β) in (1) varies among batches according to some parametric bivariate normal distribution, are suggested to use.

There are variations of the linear regression models similar to (1), e.g., see Chen, Hwang and Tsong (1995) and Chen, Ahn and Tsong (1997). Nevertheless, most of the linear regression models in stability analysis have been used with resort to certain parametric assumptions. In reality, however, it may not be always true that these parametric assumptions are appropriate. For example, when the number of manufacturing batches is actually limited for the drug product, then the continuous bivariate normal assumption on (α, β) is questionable in dealing with the batch-to-batch variations.

The parametric models and their associated maximum likelihood methods do have much advantage in gaining efficiency and are generally easy to compute with today's computer resources. The trade-off is that their subsequent inferences rely heavily upon the parametric assumptions. Without strong belief in these assumptions, however, there is imminent need to develop alternative flexible models with less restrictive assumptions.

In this article, the semiparametric linear regression models are proposed and studied as alternatives to the parametric models such as (1), without restrictive normality assumptions on the errors while the parametric linear trend over time is preserved. Accordingly, the rest of this article is organized as follows. In §2.1, the semiparametric regression models are presented and discussed. The rank-based inference procedures to determine shelf-life

are studied and investigated in §2.2. Numerical studies including Monte-Carlo simulations and applications of the new methods in practice are in §3. Some remarks and discussion on potential methodology extensions are in §4.

2 Stability Analysis Based on Rank Regression

2.1 Semiparametric regression models

The linear regression models are frequently used in the stability analysis to describe the product's characteristics of interest. Because of the products manufactured with variablity by batches, it is suggested that the samples selected in stability studies should "constitute a random sample from the population of production batches" (FDA, 1987). Suppose that there are n batches in the stability analysis. Denote Y_{ij} the continuous outcome of the characteristic for the jth measurement of the ith batch, $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, m_i$. Then in general the linear regression model assumes that

$$Y_{ij} = \alpha_i + \beta_i X_{ij} + e_{ij}, \tag{2}$$

where X_{ij} are the associated covariates, $(\alpha_i, \beta_i)^{\mathrm{T}}$ are the parameters and e_{ij} are the independent random errors. Specifically in stability analysis, since the shelf-life need to be finally determined, X often includes a variable of time in certain scale. In addition, the random errors are usually assumed to be identically zero-mean Normal deviates with constant standard variance σ_e^2. The superscript T denotes the transpose of vector or matrix.

In practice, when the batch-to-batch variation is of little concern, it is often assumed that $\alpha_i = \alpha_0$ and $\beta_i = \beta_0$ for any i. One special situation of such, for example, is that there is only one single batch, i.e., $n \equiv 1$. The usual multiple linear regression models with the method of the ordinary least squares are suggested by Chow and Shao (2002, p. 83) to obtain the estimates of $(\alpha_0, \beta_0)^{\mathrm{T}}$.

When the batch-to-batch variation is suspected, there are indeed three occasions of model (2) that may allow certain degree of batch-to-batch variation:

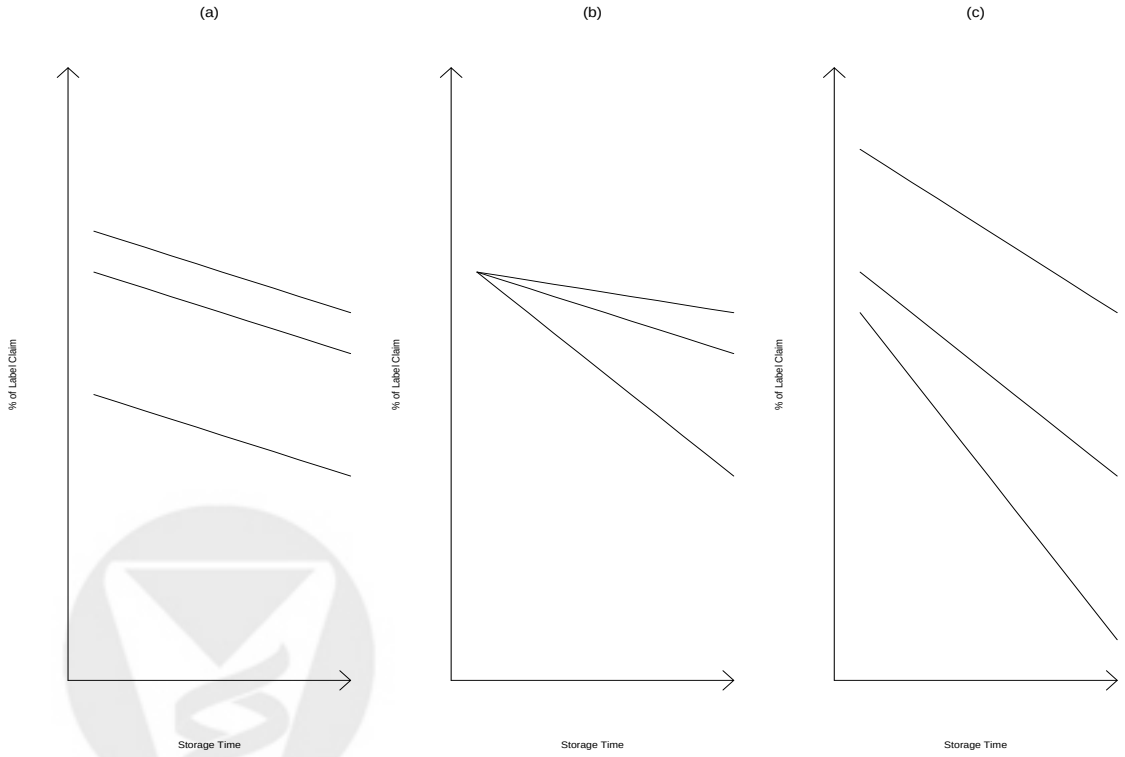
(1) intercept terms but not slopes, or

(2) slopes but not intercept terms, or

(3) both,

are different from batch to batch, as demonstrated in Figure 2(a), 2(b) and 2(c), respectively. The random effects models have been advocated to deal with the batch-to-batch variation. For example, in the afore-mentioned occasion (3) of model (2), (α_i, β_i) can be further assumed to follow a Normal distribution, $N((\alpha_0, \beta_0)^{\mathrm{T}}, \Sigma)$, say (Chow and Shao, 1991). In addition to the straightforward estimation procedure using the maximum likelihood of the final mixture distributions, the random effects models also bear the implication of describing the "future" batches.

Figure 2: Illustration of three potential batch-to-batch variations



It is well known that the fully parametric approaches carry much benefit in model interpretation, estimation and readily available softwares. However, they also have to heavily depend on the validity of their parametric assumptions. For example, a normal approximation of the batch-to-batch varying $(\alpha_i, \beta_i)^{\mathrm{T}}$s has to be degenerated in either α or β to

accommodate the first two afore-mentioned occasions of model (2), respectively. Therefore, there is need in developing more flexible models with less assumptions. One of such effort can be made by semiparametric modeling. Specifically, the following semiparametric model is assumed:

$$h(Y_{ij}) = \alpha_i + \beta_i X_{ij} + e_{ij}, \tag{3}$$

where e_{ij} are independent zero-mean deviates with unknown density function of $f(\cdot)$ and $h(\cdot)$ is some known monotonic transformation function.

The model in (3) is quite flexible without assuming the underlying distribution of normality or constant variance. Moreover, as shown in the following section it does not necessarily assume the parametric distributions on $(\alpha_i, \beta_i)^{\mathrm{T}}$ as random effects but leaving them as fixed effects, when appropriate rank-based inference procedures are utilized for the corresponding models. The parametric component in model (3) is still the linear combination of parameter β_i and covariates X_{ij}. So the sign and quantity of β_i characterizes the direction and magnitude of linear trend of Y_{ij} with respect to X_{ij}, respectively. For example, when X_{ij} are the times at which Y_{ij} are measured, β_i may stand for the "degradation" of measurements over time for its negative sign.

2.2 Rank-based shelf-life determination

We first consider model (3) for the stability analysis when one batch of stability data is to be analyzed. That is, $n \equiv 1$. So the index i is ignored to simply the notations. Furthermore, without loss of generality, we assume $h(\cdot)$ is an identity link. Then the model becomes

$$Y_j = \alpha + \beta X_j + e_j.$$

Under the Null Hypothesis of $\beta = 0$, a standard rank test statistic is then

$$U = \sum_{j=1}^{m} (X_j - \bar{X}) R(Y_j),$$

where $\bar{X}$ is mean of X's and $R(Y_1), R(Y_2), \ldots, R(Y_m)$ are the ranks of $Y_1, Y_2, \ldots, Y_m$, respectively. When β is not necessarily zero, a straightforward extension is to consider the residuals of $e_j = Y_j - (\alpha + \beta X_j)$:

$$U(\beta) = \sum_{j=1}^{m} (X_j - \bar{X}) R(e_j).$$

Apparently $E[U(\beta);\beta] = 0$ for any $\beta \in \mathcal{B} \subset \mathbf{R}$, where $\mathcal{B}$ is the parameter space and $\mathbf{R}$ is the real line. Therefore it is reasonable to solve the unbiased estimating equation of $U(\beta)$ to obtain appropriate estimates of β:

$$U(\hat{\beta}) = 0.$$

It is noticeable that the parameter of α is not estimable from $U(\beta)$, because it serves as baseline effect for every observation and therefore the ranks are invariant to the constant intercept. However, it is critical in stability analysis, because it has to be determined for further estimation of the shelf-life. One remedy to determine α is in use the median of $(Y_j - \hat{\beta}X_j)$, $j = 1, 2, \ldots, m$. When the distribution of error terms are further symmetric, α can be estimated by the median of the Walsh average of residuals of $(e_i + e_j)/2, 1 \leq i \leq j \leq m$ (Hettmansperger, 1984). Denote $\hat{\alpha}$ the estimate of α.

Since $U(\beta)$ is monotonically decreasing step function in β, there may not exist values of β to allow $U(\beta) = 0$ exactly. One alternative is to define $\hat{\beta}$ that satisfies $U(\beta+)U(\beta-) \leq 0$. The other alternative is to obtain the rank estimate $\hat{\beta}$ by minimizing the

$$D(\beta) = \sum_{j=1}^{m} w_j R(e_j) e_j,$$

where w_j are nonconstant sequence of scores such that $w_j + w_{m-j+1} = 0$. Then $D(\beta)$ is a nonnegative, continuous and convex function of β. Suppose that β_0 is the true value of β, according to Theorem 5.2.3 in Hettmansperger (1984),

$$m^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{D} N(0, \tau^2 \Sigma^{-1}),$$

where $\tau = (\sqrt{12} \int_0^\infty f^2(u)du)^{-1}$ and Σ is the variance-covariance matrix of $e = (e_1, e_2, \ldots, e_m)^{\mathrm{T}}$. Furthermore, when $\hat{\alpha}$ is the median of $\hat{e}_j = Y_j - \hat{\beta}X_j$, it can be shown that $m^{1/2}[(\hat{\alpha}, \hat{\beta})^{\mathrm{T}} - (\alpha_0, \beta_0)^{\mathrm{T}}]$ would have an asymptotic distribution of multivariate normal with mean zero and variance-covariance

$$V = \tau^2 \begin{bmatrix} [2f(0)\tau]^{-2} + \mu^{\mathrm{T}}\Sigma^{-1}\mu & -\mu^{\mathrm{T}}\Sigma^{-1} \\ -\mu^{\mathrm{T}}\Sigma^{-1} & \Sigma^{-1} \end{bmatrix},$$

where μ is the mean vector of Y's and $f(\cdot)$ is the density function of e_i's. When $\hat{\alpha}$ is the median of the Walsh averages, $m^{1/2}[(\hat{\alpha}, \hat{\beta})^{\mathrm{T}} - (\alpha_0, \beta_0)^{\mathrm{T}}]$ would have the asymptotic distribution of multivariate normal with mean zero and variance-covariance of

$$V = \tau^2 \Lambda^{-1},$$

where $\Lambda^{-1} = \lim_{m\to\infty}[\mathbf{1}, \mathbf{X}]^{\mathrm{T}}[\mathbf{1}, \mathbf{X}]$. Here $\mathbf{1}$ is unit vector and $\mathbf{X}$ the design matrix.

As a result, an approximately 95% pointwise confidence bound for EY_i is then $(\mathrm{LB}_i, \mathrm{UB}_i)$ where

$$\mathrm{LB}_i = \hat{\alpha} + \hat{\beta}X_i + Z_{0.025}\sqrt{(1, X_i)\hat{V}(1, X_i)^{\mathrm{T}}}$$

and

$$\mathrm{UB}_i = \hat{\alpha} + \hat{\beta}X_i + Z_{0.975}\sqrt{(1, X_i)\hat{V}(1, X_i)^{\mathrm{T}}}.$$

Therefore, an estimate of shelf-life is the point of time, T_{shelf}, at which L_i or U_i intersects the pre-defined most allowable specification, η, say, depending on whether β is negative or positive, respectively.

When the batch-to-batch variation is present as in occasion (1):

$$Y_{ij} = \alpha_i + \beta X_{ij} + e_{ij}$$

where β is the common slope for every batch. This model is most plausible when the degradation of product substances is uniform and the batch-to-batch variation only comes from the difference in initial value after manufacturing of each batch. If the aforementioned rank-based estimating equation is used for this model, then it is not necessarily to assume that α_i have to follow some parametric random effects, but rather be treated as unknown fixed quantities:

$$U(\beta) = \sum_{i=1}^{n} U_i,$$

where $U_i = \sum_{j=1}^{m_i}(X_{ij} - \bar{X}_i)R(e_{ij})$. Here $\bar{X}_i = m_i^{-1}\sum_j X_{ij}$ and $e_{ij} = Y_{ij} - (\alpha_i + \beta X_{ij})$. Similarly, $U(\beta)$ is not able to estimate α_i's. However, the medians of $Y_{ij} - \beta X_{ij}$, $j = 1, 2, \ldots, m_i$, can be used to estimate α_i, $i = 1, \ldots, n$, individually.

When the batch-to-batch variation is present as in occasion (2):

$$Y_{ij} = \alpha + \beta_i X_{ij} + e_{ij}$$

where α is the common slope for every batch. This model is most plausible when the degradation of product substances does not share identical rate of change although its initial value of characteristics are controlled to be the same. To estimate individual β_i, consider:

$$U_i(\hat{\beta}_i) = 0$$

for every individual i to solve for $\hat{\beta}_i$. Then the medians of $Y_{ij} - \hat{\beta}_i X_{ij}$, $j = 1, 2, \ldots, m_i$, $i = 1, \ldots, n$, can be used to estimate α.

When the batch-to-batch variation is present as in occasion (3):

$$Y_{ij} = \alpha_i + \beta_i X_{ij} + e_{ij}$$

where the intercepts and slopes are both different from batch-to-batch. This model is most plausible when the batches of products are quite heterogeneous in either the initial value of characteristics or the degradation rates. To estimate individual β_i, consider:

$$U_i(\hat{\beta}_i) = 0.$$

And the medians of $Y_{ij} - \hat{\beta}_i X_{ij}$, $j = 1, 2, \ldots, m_i$ can be used to estimate α_i, $i = 1, \ldots, n$, respectively. An interesting observation is that the $\hat{\beta}_i$ solved from $U_i(\hat{\beta}_i)$ should be identical to the estimates in occasion (2), demonstrating some robustness of rank-based estimating equation in terms of estimating the slope of degradation rate.

If the batch-to-batch variation exists, computing the appropriate shelf-life becomes challenging and controversial, because there is not one set of $(\hat{\alpha}, \hat{\beta})$ but several $(\hat{\alpha}_i, \hat{\beta}_i)$ to determine the confidence bands. One straightforward approach is to use the method suggested by the FDA guidelines. That is, first compute T_{shelf} for each individual batch, and the final shelf-life is then $\min(T_{\text{shelf},i}, i = 1, 2 \ldots, n)$. Or, when the number of batches are enough large and potentially the so-called "future" batches may also show variability, then the shelf-life can be computed based on the prediction bounds of the average values of α_i and β_i, following the approach by Shao & Chen (1997).

To test the batch-to-batch variation, the method in Chow & Shao (2002, p. 98), which is similar to the Woolfe's test for stratified samples in strata homogeneity, can be extended to the obtained rank-based estimates. Specifically, suppose that $(\hat{\alpha}_i, \hat{\beta}_i)$ are obtained in occasion (3) for individual batches and $(\hat{\alpha}, \hat{\beta})$ in the occasion without considering the batch-to-batch variation. Then the statistics

$$\text{TS} = \sum_{i=1}^{n} (\hat{\alpha}_i - \hat{\alpha}, \hat{\beta}_i - \hat{\beta})^{\text{T}} \hat{V}_i^{-1} (\hat{\alpha}_i - \hat{\alpha}, \hat{\beta}_i - \hat{\beta}),$$

where $\hat{V}_i$ are the estimates of the variance of $(\hat{\alpha}_i, \hat{\beta}_i)$, respectively, under the null hypothesis that there is no batch-to-batch variations. It approximately follows χ^2-distribution with degrees of freedom of $n-1$. The null hypothesis is rejected if TS tends to be unusually large, for example, when its associated p-value is not bigger than 0.25.

3 Numerical Studies

Like other rank-based linear regression approaches, the computation of the proposed rank-based approaches in previous sections are generally tedious. In fact, there are some challenging issues in actual implementation. One challenge involved is to solve the estimating function itself. When the covariates are of one dimension or even low dimensions, the usual bisection or grid search should be suffice. But when the covariates are of relatively higher dimensions, then the finding of solution using grid search may be intractable. One way to solve this is using the so-called "recursive bisection" programming (Chen and Jewell, 2001). The idea is, given the solution at the kth step, the $(k+1)$st solution can be obtained by bisectional approach. In practice, this approaches well within reasonable time frame.

Another challenge is to compute the appropriate confidence bands of $(\mathrm{LB}_i, \mathrm{UB}_i)$. Although they are of usual simple form similar to their counterparts in linear regression models with the least-squares, the actual computation is not as straightforward because of the parameter $\tau = (\sqrt{12}\int f^2(u)du)^{-1}$ involved with unknown density function f. Without an appropriate estimate of $\hat{\tau}$, it is foreseeable that $\hat{V}$ would be difficult to estimate in any straightforward sense. One estimate of $\int f^2(u)du$ is by Schuster (1974). That is, first obtain a kernel type of estimate of $f(u)$:

$$\tilde{f}(u) = \frac{1}{mh}\sum_{j=1}^{m} k\left(\frac{u-e_j}{h}\right),$$

where $k(\cdot)$ is a uniform kernel function

$$k(u) = \begin{cases} 1 & u \in [-1/2, 1/2] \\ 0 & \text{otherwise} \end{cases},$$

and h is kernel bandwidth. Then $\delta = \int f^2(u)du = \int f(u)dF(u)$ can be estimated by

$$\hat{\delta} = \frac{1}{m^2 h}\sum_{i=1}^{m}\sum_{j=1}^{m} I\left(|e_i - e_j| < \frac{h}{2}\right).$$

It is shown to be consistent as in Aubuchon (1982). A modified version of the above estimate, $\hat{\delta}_c$, is proposed in Hettmansperger (1984) to ease the computation,

$$\frac{1}{mc} + \frac{1}{m(m-1)h}\sum_{i=1}^{m}\sum_{j\neq i} k\left(\frac{e_i - e_j}{h}\right),$$

where c is any fixed constant. The modified $\hat{\delta}_c$ is also consistent of δ when $h = cm^{-1/2}$ (Theorem 5.2.4, Hettmansperger, 1984). To actually implement $\hat{\delta}_c$, the following algorithm can be applied.

Algorithm[1]:

1. obtain $\hat{\beta}$ and residuals $\hat{e}_i = y_i - \hat{\beta}x_i$, $i = 1, 2, \ldots, m$;
2. compute the inter-quartile range, d, of $\hat{e}_i$, $i = 1, 2, \ldots, m$;
3. compute $\hat{\delta}_c$ as

$$\frac{1}{4.11md} + \frac{2}{m^{1/2}(m-1)4.11d} \sum\sum_{i<j} I\left(\frac{m^{1/2}(\hat{e}_i - \hat{e}_j)}{4.11d}\right).$$

In our simulations, we study the performance of rank-based estimation procedure under different error structures. For simplicity, we only consider the following single covariate data generation model of occasion (1)

$$y_{ij} = \alpha_i + \beta x_{ij} + e_{ij}, \quad \text{for } i = 1, \ldots, n \text{ and } j = 1, \ldots, m,$$

where y_{ij} is the characteristic of interest of the jth measurement for the ith batch, x_{ij} is the time of the jth measurement for the ith batch, e_{ij} is random error of the jth measurement for the ith batch, α_i is batch-specific intercept and β is common slope. Note that this is the common slope model. But it is straightforward to expand all the simulations described below to other occasions with batch-to-batch variation.

The FDA (1987) recommends at least three batches are tested to obtain the shelf-life for an NDA submission. In our simulation study, we considered both $n = 1$ and $n = 3$. Furthermore, we assume that the characteristic of interest decreases over time, as $\beta = -0.5$, i.e., 0.5 unit decreasing change per month, say. In addition, $\alpha = 100$ for $n = 1$ and $\alpha = \{100, 101, 102\}$ for $n = 3$. The sampling time points are of a full FDA plan, i.e., $x = (0, 3, 6, 9, 12, 18)$. Three error distributions are considered: (1) Normal ($e_{ij} \sim N(0, 1)$); (2) Uniform ($e_{ij} \sim U(-2, 2)$); and (3) Log-normal ($\log e_{ij} \sim N(0, 1)$). Note that the error distributions of (2) and (3) are not the commonly used normal distributions.

[1]Here $c = 4.11d$ is calculated as if $f(\cdot)$ were normal density, when the bias in $\hat{\delta}_c$ is set to be zero. The actual formula for c is $c = d \cdot \left\{0.5 \int [f_1'(x)]^2 dx \cdot \int x^2 I(|x| \leq 0.5) dx\right\}^{-1/3}$, where $f_1(x) = d \cdot f(d \cdot x)$. The different choices of density function $f(\cdot)$ should not lead to dramatic change in c.

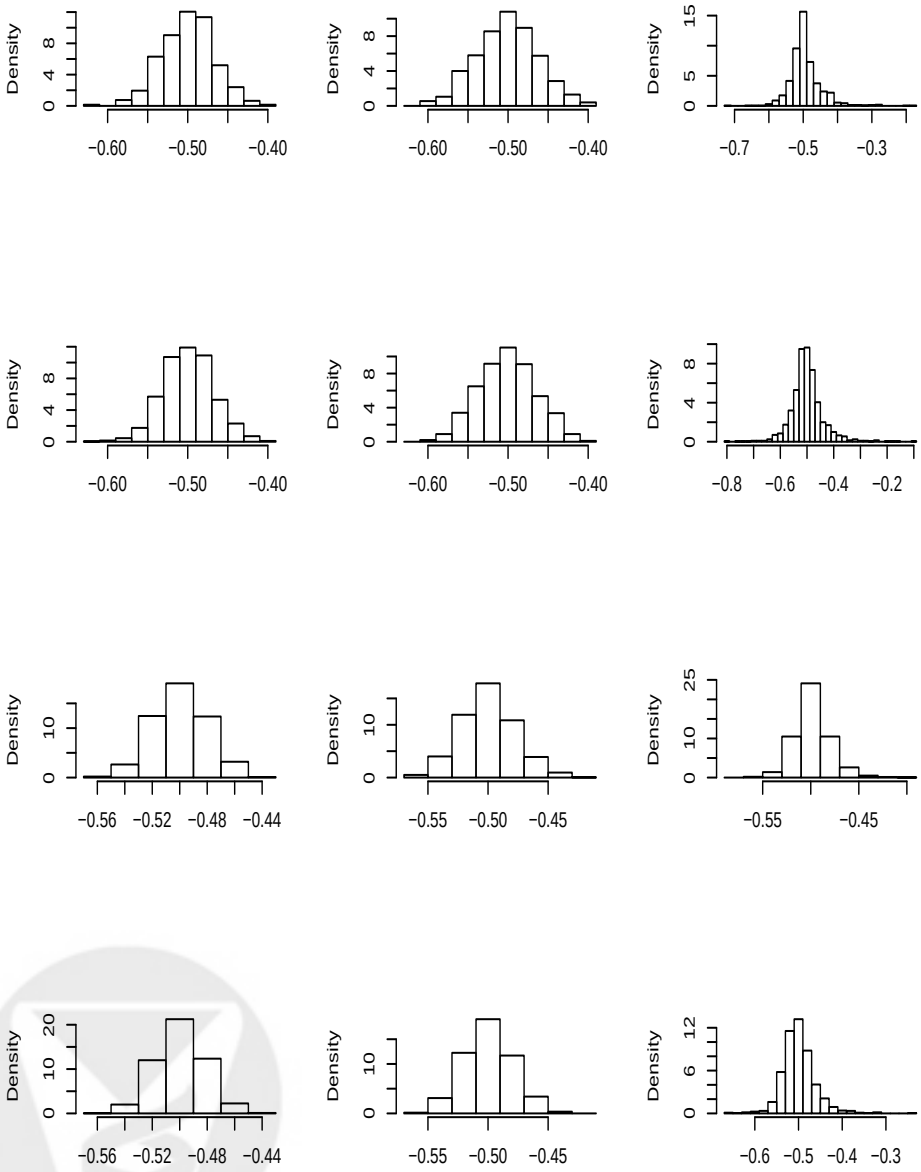
One thousand sets of data for each error distribution and for number of batch $n = 1$ and $n = 3$, respectively. The mean and standard deviation of the estimated intercepts and slopes are reported in Table 1. Histograms of the estimated slopes are shown in Figure 3. It is noticeable that the parameter estimates of rank regression models are comparable in terms of efficiency with those of normal linear regression models with the ordinary least squares (OLS) for data with normal errors, or other errors that deviate not too much from normal errors. In these cases, the OLS estimates for normal linear models tend to be a bit more efficient than the rank regression estimates with the normal assumption, as can be seen from their smaller standard deviations. However, the rank regression is more robust and efficient for data with errors that deviate much from normal errors (e.g., log-normal errors). It can be seen from Table 1 that the standard deviations for the estimated intercepts and slopes for the rank regression models are much smaller than those correspond to the estimates for the normal linear models fitted by OLS. Note that since the expectation of log-normal error is positive, the estimated intercepts for both rank regression models or normal linear models are biased for data with log-normal errors.

Table 1: Mean and Standard Deviation of Estimated Model Parameters

	Parameter	$e_{ij} \sim N(0,1)$	$e_{ij} \sim U(-2,2)$	$\log e_{ij} \sim N(0,1)$
[n=1]				
RRM	α	100.0048 (0.6057)	99.9829 (0.8052)	101.1343 (0.7491)
RRM	β	-0.4995 (0.0328)	-0.5010 (0.0384)	-0.4941 (0.0462)
NLM	α	100.0077 (0.5489)	99.9992 (0.6230)	101.6468 (1.1458)
NLM	β	-0.4994 (0.0316)	-0.5012 (0.0359)	-0.4997 (0.0623)
[n=3]				
RRM	α_1	99.9860 (0.5051)	100.0252 (0.6642)	101.0751 (0.4958)
RRM	α_2	100.9893 (0.5012)	101.0363 (0.6622)	102.1150 (0.5241)
RRM	α_3	101.9936 (0.4877)	102.0038 (0.6615)	103.1259 (0.5476)
RRM	β	-0.4996 (0.0193)	-0.4998 (0.0227)	-0.4981 (0.0182)
NLM	α_1	99.9915 (0.4411)	100.0097 (0.4955)	101.6286 (0.9150)
NLM	α_2	100.9961 (0.4302)	101.0180 (0.4847)	102.6333 (0.9056)
NLM	α_3	101.9827 (0.4275)	102.0043 (0.4960)	103.6305 (0.8960)
NLM	β	-0.4995 (0.0176)	-0.4997 (0.0205)	-0.4993 (0.0382)

Note: RRM: Rank Regression Model. NLM: Normal Linear Model (fitted by OLS). Standard deviations are enclosed in parenthesis. $n = 1$ or $n = 3$ represents number of batches. True parameters are: $\beta = -0.5$, $\alpha = 100$ for $n = 1$, and $\alpha_1 = 100$, $\alpha_2 = 101$, and $\alpha_3 = 102$ for $n = 3$.

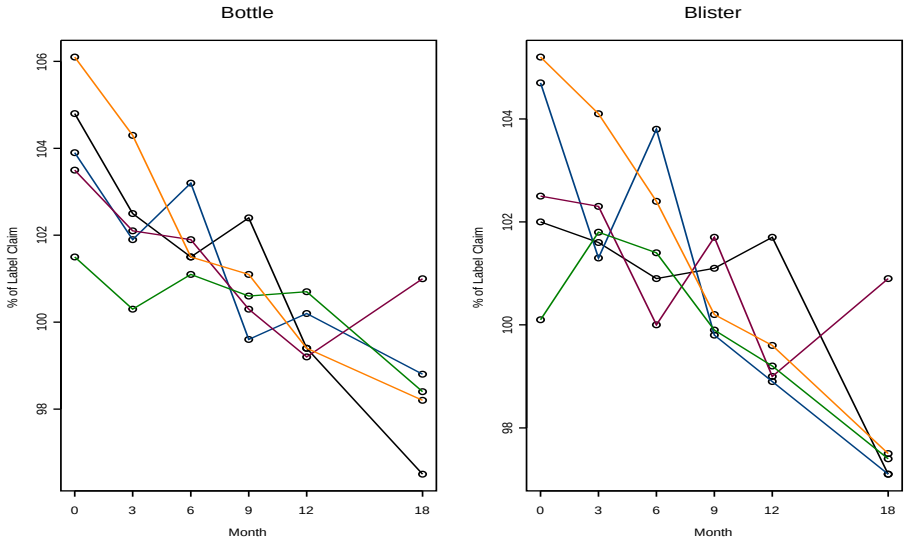
Figure 3: Histograms of Estimated Slopes



Note: Row 1: Rank Regression (n=1); Row 2: Normal Linear Regression (OLS) (n=1); Row 3: Rank Regression (n=3); Row 4: Normal Linear Regression (OLS) (n=3). Column 1: $e_{ij} \sim N(0,1)$; Column 2: $e_{ij} \sim U(-2,2)$; Column 3: $\log e_{ij} \sim N(0,1)$. True slope is $\beta = -0.5$.

In Shao and Chow (1994), a data set of the stability study on a 300-mg tablet of a drug product to determine appropriate labeled shelf-life was given and analyzed. The tablets from five batches were stored at room temperature in two types of containers (high-density polyethylene bottle and blister packages). Tests of potency were conducted at 0, 3, 6, 9, 12 and 18 months. The data are plotted in Figure 4.

Figure 4: Shao and Chow (1994) data



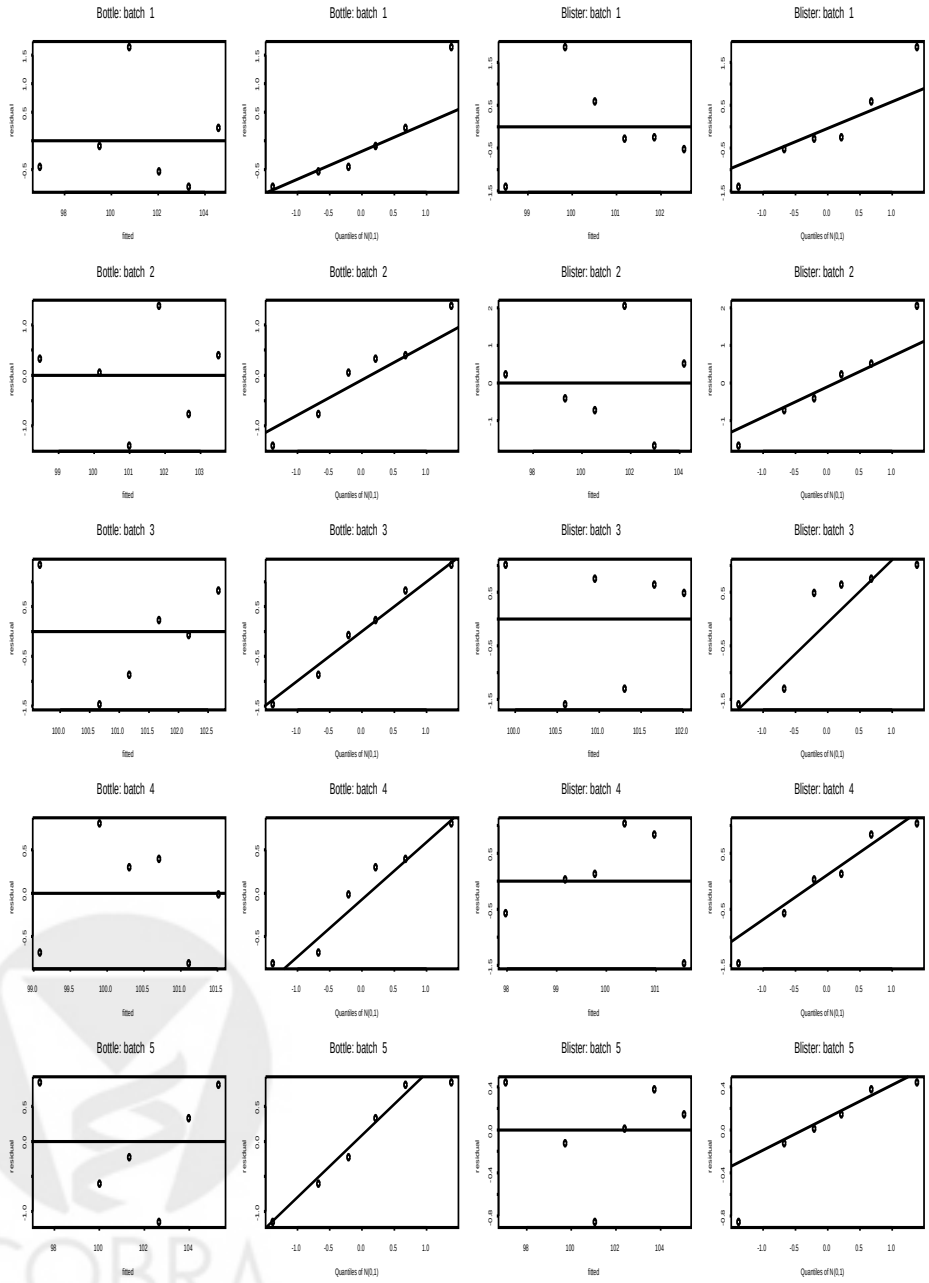
It can be seen in the figure that the potency is generally degrading as time progresses. Furthermore, simple linear regression models are fitted with the OLS to each batch. Their residuals are plotted in Figure 5. The residuals are plotted against the fitted values and their corresponding normal quantiles, respectively, for each batch. As seen in Figure 5, the residual plots does not always show random pattern, but some of them may suggest the violation of normal assumption.

A closer look at Figure 4 would find that the variation of degradation curves tends to be bigger in intercepts while the slopes of these curves tend to be similar. Therefore, the models of occasion (1) of common slope with different intercepts

$$y_{ij} = \alpha_i + \beta X_{ij} + e_{ij}$$

are fitted to the data. Its associated rank-based estimating function of β is plotted in Figure 6. As shown in the figure, the estimating function is non-increasing with β and the estimate

Figure 5: Shao and Chow (1994) data: Residual Plot (OLS Fit)



of β are -0.32 and -0.29 for the bottle container and blister package, respectively. For the bottle container, the estimates of intercept for five batches are 103.45, 103.98, 103.35, 103.27 and 104.00. For the blister package, the estimates of the intercept for five batches are then 102.57, 102.43, 102.85, 102.68 and 103.64.

In addition, the model of varying (α_i, β_i)

$$y_{ij} = \alpha_i + \beta_i X_{ij}$$

are fitted to the data as well. The estimates of (α_i, β_i)'s for bottle container are $(104.20, -0.42)$, $(103.75, -0.28)$, $(103.25, -0.30)$, $(101.50, -0.10)$ and $(105.90, -0.53)$. Similarly, the estimates of (α_i, β_i)'s for blister package are $(102.00, -0.13)$, $(102.50, -0.30)$, $(102.52, -0.09)$, $(102.47, -0.28)$ and $(105.20, -0.47)$. The estimated mean degradation curves from both models for their respective package type are also plotted in Figure 7. Based on these estimates, the estimated shelf-lives for the bottle container and blister package with specification limit of 90% are approximately 22 and 21 months, respectively, which are comparable to the situation of $\epsilon = 0.05$ given in Shao and Chow (1994).

Figure 6: Shao and Chow (1994) data: Rank-based Estimating Function

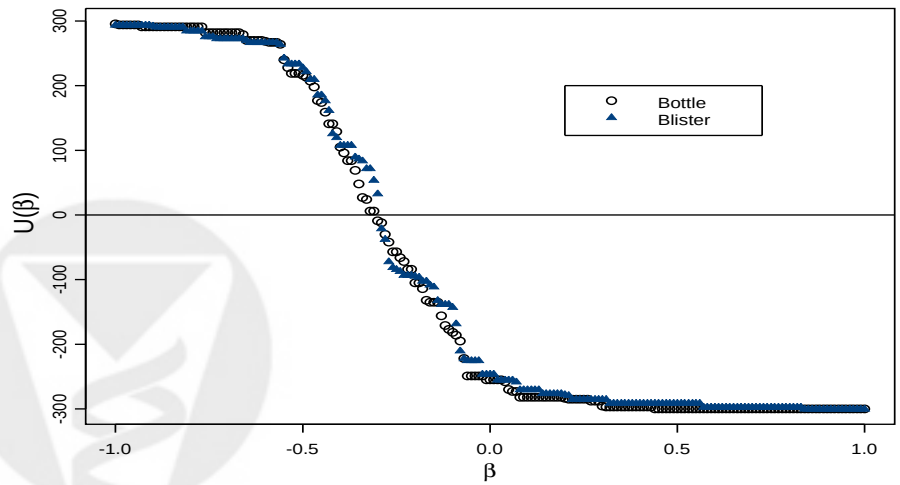
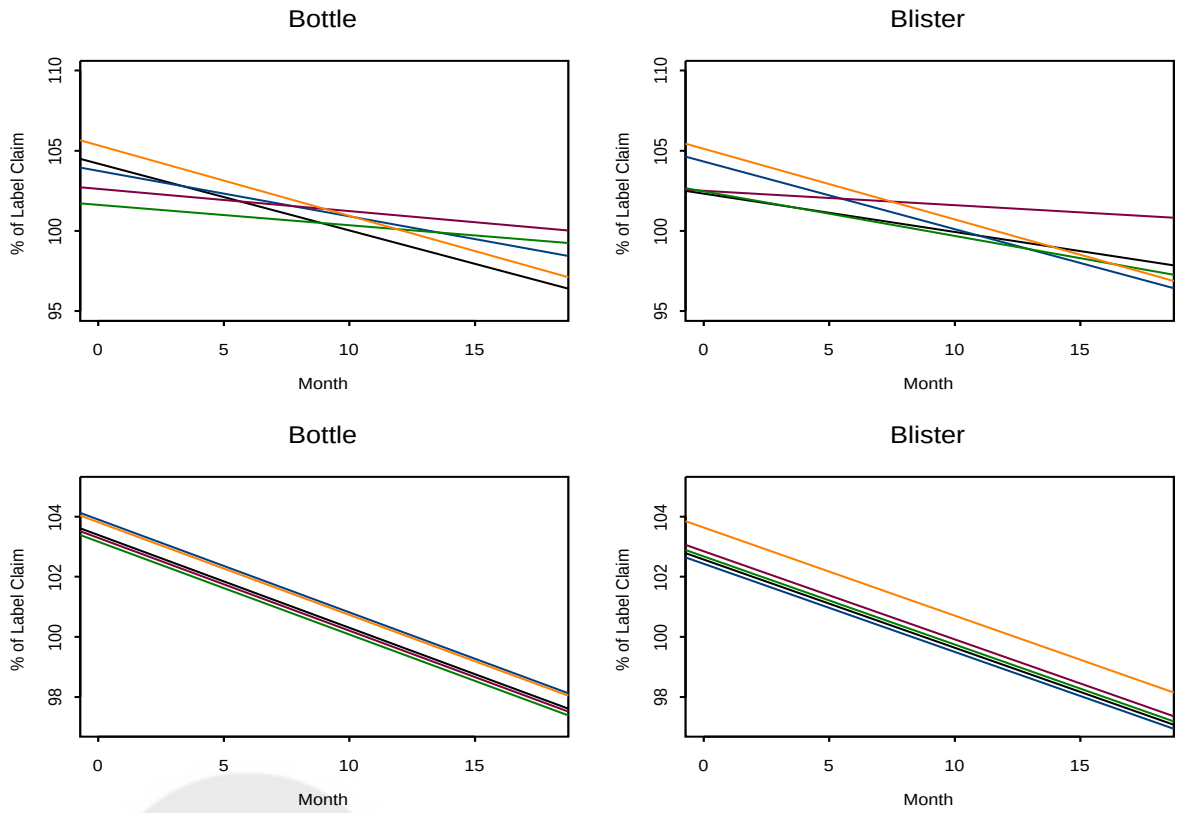


Figure 7: Shao and Chow (1994) data: Estimated Mean Degradation Curves



Note: Row 1: fitted mean degradation curves with batch-to-batch variation in both intercept and slope; Row 2: fitted mean degradation curves with batch-to-batch variation in intercept only.

4 Discussion

In this article, the semiparametric regression models and their associated rank-based inference procedures in obtaining the self-life provide an alternative approach when the parametric assumptions are not necessarily true in the fully parametric models. Although they may carry disadvantages in efficiency or computation, its flexibility to allow non-standard continuous outcomes in stability data is appealing in practice when the regulatory agencies or drug manufactures tend to be conservative in determining shelf life of actual products.

In fact, the feature of the nonparametric components in this model also have profound practical interpretation. They do not need parametric assumptions but rather are treated as fixed. By doing so, however, we do have the risk of falling into the classical Neyman-Scott problem of great number of nuisance parameters. Without the usual way of further assumptions in reducing the dimensionality of parameter space, the estimation and inference procedures may lead to biased results. Therefore, the approaches in this article may be of more advantage when the existence of the general batch-to-batch variation has limited range. But, it may be extended to the situation of extensive batch-to-batch variations when only the intercepts are different between batches, i.e., in occasion (1). In this case, the proposed approaches echoes the semiparametric models for stratifies samples when the baseline parameters are not of concern.

As argued in Shao and Chow (1994), when the batch-to-batch variation presents, the FDA guidelines (1987) on choosing the shelf life "lacks statistical justification." While the parametric approaches were further developed in their article with more reasonable justification, its practical implementation may be hindered by the choice of the ϵ in their method. Therefore, the final determination of shelf-life would be still up to the subjective discretion of regulatory agency itself with several reported choices under different circumstances. In this article, the proposed approaches does not explore the statistical justification whether or not the determined shelf-life estimates its true counterpart, but simply adopts the FDA guidelines because of their simplicity and uniform as an algorithm in choosing an intuitive "shelf-life." Thus, the results are also comparable to the available ones that are conformed with the FDA guidelines.

There are some potential extensions of the proposed methodologies in different scenarios. Some of the scenarios are listed as follows.

1. The approaches discussed in this article are for the so-called "full sampling" plan in obtaining measurements. There are alternative designs such as "bracketing" and "matrix" design as alternative (ICH, 1993). When the sampling points of response are sparse, the nonparametric estimates of variance may not be stable and hence jeopardize the semiparametric gesture of flexibility. So, how will the proposed methodologies indeed perform under different stability design schemes?

2. The approaches discussed in this article are for single major ingredient of a pharmaceutical product. When multiple active ingredients have to be considered to reach a single shelf-life for the product, can the proposed approaches accommodate it?

3. The outcomes of the proposed approaches are limited to continuous outcomes. Can the proposed methods overcome the obstacle of generalizing themselves to other types of outcomes, discrete or mixed outcomes, say?

4. The linear regression models are usually assumed for the degradation curves of outcome measurements, because of the belief in prior knowledge of decreasing benefits or increasing hazards. In fact, with such prior knowledge, it is more general to assume the degradation curves following

$$h(EY(t)) = \beta X + \mu_0(t),$$

where $\mu_0(t)$ is nonparametric monotone baseline function in t and X are other concomitant covariates such as temperature. Although this kind of partial linear models may be of more interest, its difficulty in developing sound estimation and inference procedures may be formidable, which would incur tremendous research effort and needs to be developed in theory and computation.

References

Aubuchon, J.C. (1982) *Rank Tests in the Linear Models: Asymmetry Errors.* Ph.D. Thesis. Pennsylvania State University, Department of Statistics.

Chen, J.J., Ahn, H., and Tsong, Y. (1997). Shelf-life estimation for multifactor stability studies. *Drug Information Journal*, 31:573-587.

Chen, J.J., Hwang, J.-S. and Tsong, Y. (1995) Estimation of the shelf-life of drugs with mixed effects models. *Journal of Biopharmaceutical Statistics*, 5(1):131-140.

Chen, Y.Q. and Jewell, N.P. (2001) On a general class of semiparametric hazards regression models. *Biometrika*, 88:687-702.

Chow, S.-C. and Shao, J. (1991) Estimating drug shelf-life with random batches. *Biometrics*, 47:1071-1079.

Chow, S.-C. and Shao, J. (2002) *Statistics in Drug Research: Methodologies and Recent Developments.* New York: Marcel Dekker.

Food and Drug Administration (1987) *Guideline for Submitting Documents for the Stability of Human Drugs and Biologics.* Rockville: FDA.

Hettmansperger, T.P. (1984) *Statistical Inferences Based on Ranks.* New York: John Wiley.

International Conference on Harmonization (1993). *Stability Testing of New Drug Substances and Products.* Tripartite International Conference on Harmonization Guideline.

International Conference on Harmonization (2002) *Evaluation of Stability Data.* Draft consensus guideline.

Shao, J. and Chen, L. (1997) Prediction bounds for random shelf-lives. *Statistics in Medicine*, 16:1167-1173.

Shao, J. and Chow, S.-C. (1994) Statistical inference in stability analysis. *Biometrics*, 50:753-763.

Shao, J. and Chow, S.-C. (2001) Drug shelf-life estimation. *Statistica Sinica*, 11:737-745.

Schuster, E. (1974) On the rate of convergence of an estimate of a functional of a probability density. *Scandinavian Actuarial Journal*, 1:103-107.