AAPS PharmSciTech, Vol. 12, No. 1, March 2011 (© 2010)
DOI: 10.1208/s12249-010-9558-x

Research Article

Stability Models for Sequential Storage

Emil M. Friedman[1,2] and Sam C. Shum[1]

Received 22 June 2010; accepted 30 November 2010; published online 23 December 2010

Abstract. Some drugs are intended for sequential storage under two different storage conditions. If the data for each condition are analyzed separately, predicting assay and other responses after T1 months at one condition followed by T2 months at the other condition is non-trivial for several reasons. First, the two analyses will give different intercept terms. What should one do about that? Second, how would one calculate the confidence limits for combined storage? Third, what if prior storage at one condition affects the slope at the other condition? This paper proposes a simple ANCOVA model containing two slope terms, one for each storage condition. When multiple batches and/or packages are involved, it is easily generalized to two sets of slope terms. Confidence limits are straightforward and can be calculated using existing commercial software. With properly designed data, one can test whether prior storage at one condition affects the slope at the other condition. If no such effect is significant, very useful extrapolations can be made. Temperature excursions, model reduction and curvilinear dependencies are discussed.

KEY WORDS: dual storage conditions; sequential storage conditions; shelf life; stability; statistical extrapolation.

INTRODUCTION

To market a pharmaceutical, one must determine its shelf life by performing stability studies. The drug is stored under controlled conditions and various characteristics (concentration of assay, specified degradants, etc.) are measured at a series of pre-determined time points. In the simplest possible scenario, regression analysis (1,2) is used to plot a regression line *vs* time on stability and 95% confidence limits for that line. The earliest time at which any of the confidence limits intersects the acceptance criterion for that characteristic is considered to be the shelf life (2). When multiple batches, strengths, and/or package types are involved, analysis of covariance (ANCOVA (2,3)) is used.

Some drugs, particularly biologics, are intended for sequential storage under two different storage conditions. Practical examples of sequential storage might be refrigeration during long-term storage and room temperature in the patient's home or refrigeration during "normal storage" followed by room temperature while traveling.

If the data for each condition are analyzed separately, predicting assay after T1 months at one condition followed by T2 months at the other condition is non-trivial for several reasons. First, the two analyses will give different intercept terms. What should one do about that? Second, how would one calculate the confidence limits for combined storage? Third, what if prior storage at one condition affects the slope at the other condition? Similar questions can be asked when modeling the effect of temperature and/or humidity excursions.

This paper proposes a straightforward solution to the problem. It can be implemented in a straightforward manner using any commercial software package that handles multiple regression and ANCOVA.

In the simple scenario of a single batch and two storage conditions, the simple regression model of an intercept term and a slope term is replaced by an intercept term and two slope terms, one for each storage condition. Such a procedure ensures that there is a single unambiguous intercept instead of one intercept for each set of conditions.

As will be shown, when multiple batches and/or packages are involved, the model is easily generalized to two *sets* of slope terms, one set for each storage condition. Tsong, Chen and Chen's (4) suggested improvements of the FDA/ICH model reduction rules (2) can be generalized to handle this approach. Terms describing non-linear dependence upon time are easily added when needed.

The model is especially useful when one wants to make extrapolations. Suppose one is hoping for a shelf life of 15 or 18 or 24 months at "condition L" followed by some number of months at "condition H." If one is willing to provisionally assume that the changes with time at the two conditions are additive, one does not have to wait until the full time at condition L has elapsed before starting the period at condition H. As used herein, "additivity" means that the slope at the second storage condition does not depend on how long the drug was stored at the first storage condition.

Electronic supplementary material The online version of this article (doi:10.1208/s12249-010-9558-x) contains supplementary material, which is available to authorized users.

[1] Pharmaceutical Analytics, MannKind Corporation, One Casper Street, Danbury, Connecticut 06810, USA.
[2] To whom correspondence should be addressed. (e-mail: efriedman@mannkindcorp.com; emilfrie@alumni.princeton.edu)

Extrapolation, of course, should always be considered "provisional" and subsequent real-time data should be obtained to confirm it. ICH Guidance Q1E "Evaluation of Stability Data" (2) discusses extrapolation and makes detailed recommendations concerning appropriate statistical methods.

To extrapolate reliably, it is important to design the experiment so that the additivity assumption can be tested early in the study. For example, one can use an experimental design in the two storage conditions and start gathering data at condition H after 0, 3, 6 and/or 12 months at condition L. Testing for interactions between the two conditions and/or examination of residuals can then be employed to check the assumption of additivity. If additivity is violated, interaction terms can be included in the model, but one's ability to extrapolate will be limited.

Temperature/humidity excursions can be modeled in the same way. The excursion simply becomes "condition H."

Example 1 below assumes additivity. It considers a single batch, a single package type and uses simulated data. Example 2 checks the data in example 1 to verify that the changes with time at the two conditions are additive. Example 3 involves multiple batches, multiple package types, and non-linear degradant growth at the more highly stressful of the two storage conditions. It uses real data that was rescaled to protect proprietary information. Example 4 reduces the model used in example 3 by pooling non-significant terms and discusses a possible violation of additivity.

EXAMPLE 1. SIMPLE SCENARIO

Consider the simulated data in the first four columns of Table I. A hypothetical drug was stored under two different conditions. "Condition H" was more stressful (*e.g.*, higher temperature and/or humidity) than "Condition L." The total calendar time on stability is the sum of the time spent at condition L and the time spent at condition H. Ten calendar months of data are available. The rows with missing simulated responses are the time points to which one wants to extrapolate. The simulated data were created using the equation

$$Y = 100 - 0.25t_L - 2t_H + N(0,1) \quad (1)$$

where Y is the response after time t_L at storage condition L and time t_H at storage condition H, and $N(0,1)$ is a random variable from a population having a mean of zero and a standard deviation of 1.

An analysis of covariance (ANCOVA) of the data was performed using the statistical model

$$Y = \alpha + \beta_L t_L + \beta_H t_H + \varepsilon \quad (2)$$

α, β_L, and β_H are the initially unknown intercepts and slopes at each of the storage conditions, and ε is an $N(0,\sigma)$ error term. The "usual assumptions" of independently distributed, homogeneous and normally distributed error in the response are made for all examples in this work.

Table I. Simulated Data for Example 1

Time spent at condition L	Time spent at condition H	Total time on stability	Simulated response	Predicted response	Lower 95% confidence limit	Upper 95% confidence limit
0	0	0	99.46	99.5033651	98.3226919	100.684038
0	1	1	95.97	97.5649457	96.5803852	98.5495062
0	2	2	93.80	95.6265263	94.7920463	96.4610062
0	3	3	94.51	93.6881068	92.9298312	94.4463824
0	4	4	91.11	91.7496874	90.9717192	92.5276555
0	5	5	89.60	89.8112679	88.9240731	90.6984627
0	6	6	89.68	87.8728485	86.8142539	88.931443
3	0	3	101.00	98.8014331	97.9287359	99.6741303
3	1	4	97.22	96.8630136	96.1963093	97.529718
3	2	5	95.90	94.9245942	94.396374	95.4528144
3	3	6	93.45	92.9861747	92.4712321	93.5011174
3	4	7	91.45	91.0477553	90.4129774	91.6825332
3	5	8	88.78	89.1093359	88.2772617	89.94141
3	6	9	86.85	87.1709164	86.1063134	88.2355194
6	0	6	97.06	98.099501	97.2417588	98.9572432
6	1	7	96.23	96.1610816	95.4357263	96.8864369
6	2	8	94.78	94.2226621	93.5384206	94.9069037
6	3	9	91.99	92.2842427	91.5346693	93.0338161
6	4	10	88.62	90.3458232	89.4474015	91.244245
6	5	11	–	88.4074038	87.3100932	89.5047143
6	6	12	–	86.4689844	85.1451087	87.79286
9	0	9	97.35	97.397569	96.2502446	98.5448933
9	1	10	95.88	95.4591495	94.3571985	96.5611006
9	2	11	–	93.5207301	92.3965118	94.6449483
9	3	12	–	91.5823106	90.371912	92.7927092
9	4	13	–	89.6438912	88.2955995	90.9921829
9	5	14	–	87.7054717	86.181548	89.2293955
9	6	15	–	85.7670523	84.0412411	87.4928635

Table II. ANCOVA Results for Example 1

Analysis of variance				
Source	Degrees of Freedom	Sum of squares	Mean square	*F* ratio
Model	2	269.26761	134.634	113.1463
Error	18	21.41837	1.190	*p*-value
C. total	20	290.68598		<0.0001
Parameter estimates				
Term	Estimate	Standard error	*t* ratio	*p*-value
Intercept	99.503365	0.561979	177.06	<0.0001
Time at condition L	−0.233977	0.087372	−2.68	0.0154
Time at condition H	−1.938419	0.131058	−14.79	<0.0001

For this simple example, we have assumed additivity between the slopes at the two conditions. In other words, the amount of time spent at condition L does not affect the later slope at condition H. Example 2 tests this data for additivity.

JMP® version 6.0.3 was used to calculate the intercepts, slopes, and other statistics in Table II, along with the predicted values and confidence limits shown in Table I. However, any software that handles regression and ANCOVA should suffice. The data, regression lines and confidence limits are shown in Fig. 1. Note how the model constrains the condition L and condition H regression lines to intersect appropriately. The same is true for the confidence limits.

EXAMPLE 2. CHECKING FOR NON-ADDITIVITY—INTERACTION TERM

The data from example 1 were checked for additivity between the slopes at the two conditions by adding a 3-degree-of-freedom interaction term between *time at L treated as nominal* and *time at condition H* to the model used in example 1. The main effect for time at L treated as nominal was *not* included because that would over specify the model and make it impossible to estimate a slope for time at condition L.

Conceptually, the equation to fit became

$$Y = \alpha + \beta_L t_L + \beta_H t_H + (\beta_{H0} t_H \text{ or } \beta_{H3} t_H \text{ or } \beta_{H6} t_H \text{ or } \beta_{H9} t_H) + \varepsilon \quad (3)$$

where the β_{H0} term was used when there was no prior low severity storage, the β_{H3} term was used when the drug was stored for 3 months at low severity before being placed under high severity conditions, etc., and $(\beta_{H0}+\beta_{H3}+\beta_{H6}+\beta_{H9})$ was constrained to be zero. $(\beta_H+\beta_{H0})$ estimates the high severity slope after no storage at low severity, $(\beta_H+\beta_{H3})$ estimates the high severity slope after 3 months at low severity, etc., and β_H estimates the average of the four high severity slopes.

Numerically, JMP version 6.0.3 accomplished the fit using "indicator" or "dummy" variables (5) and tested whether the extra three degrees of freedom improved the fit enough to say that there was a statistically significant difference among the four high severity slopes. The results (Table III) show that the interaction term was not significant ($p=0.57$) so the additivity assumption in example 1 was not rejected.

For the purposes of this paper, it is not necessary to understand the "sum to zero" indicator variables used by JMP, but they are shown in Eq. 4 for the interested reader:

$$Y = \alpha + \beta_L t_L + \beta_H t_H + D_0\beta_{H0} t_H + D_3\beta_{H3} t_H + D_6\beta_{H6} t_H + \varepsilon \quad (4)$$

Samples with no prior low severity storages are assigned ($D_0=1$, $D_3=0$, $D_6=0$), samples with 3 months prior low severity storage are assigned ($D_0=0$, $D_3=1$, $D_6=0$), samples with 6 months prior low severity storage are assigned ($D_0=0$, $D_3=0$, $D_6=1$), samples with 9 months prior low severity storage are assigned ($D_0=-1$, $D_3=-1$, $D_6=-1$), and β_{H9} is set equal to $(1-\beta_{H0}-\beta_{H3}-\beta_{H6})$.

Other parameterizations can be used to accomplish the same results (3). For example, the PROC GLM model in SAS, "Response=time at L (continuous), time at H (continuous), time at L (nominal)*time at H (continuous)" fits equation (5):

$$Y = \alpha + \beta_L t_L + \beta_H t_H + D_0\beta_{H0} t_H + D_3\beta_{H3} t_H + D_6\beta_{H6} t_H + D_9\beta_{H9} t_H + \varepsilon \quad (5)$$

It assigns ($D_0=D_3=D_6=0$, $D_9=1$) to samples with 9 months prior low severity storage, adds the assignment

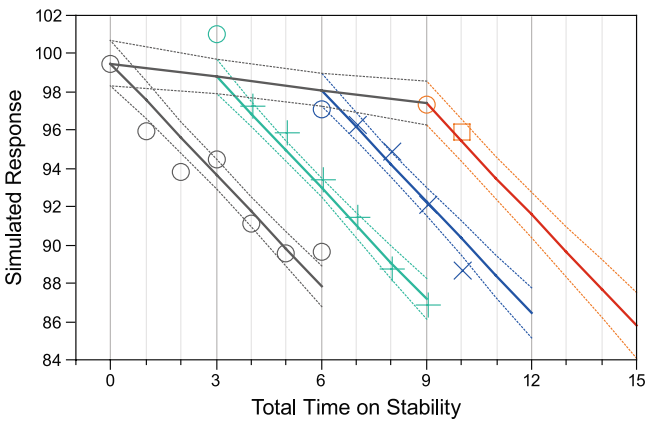

Fig. 1. Results of example 1. *Solid lines* are predicted response. *Dashed lines* are 95% confidence limits. *Black circles* are at storage condition H with no prior storage at condition L. *Green plus signs* are after 3 months at storage condition L. *Blue X*s are after 6 months at L. *Red square* is after 9 months at L. *Green*, *blue*, and *red circles* experienced storage at condition L but were not stored at condition H. *Black lines* are at pure condition L or pure condition H. *Green*, *blue*, and *red lines* are after 3, 6, or 9 months at condition L

Table III. ANCOVA Results for Example 2

Analysis of variance				
Source	Degrees of Freedom	Sum of squares	Mean square	*F* ratio
Model	5	271.87717	54.3754	43.3643
Error	15	18.80881	1.2539	*p*-value
C. total	20	290.68598		<0.0001
Effect tests				
Source	Degrees of Freedom	Sum of squares	*F* ratio	*p*-value
Time at condition L (treated as continuous)	1	1.162148	0.9268	0.3510
Time at condition H	1	34.401093	27.4348	0.0001
Interaction between time at L treated as nominal and time at condition H	3	2.609559	0.6937	0.5701

$D_9=0$ to the JMP assignments for the other samples, and its generalized matrix inversion sets $\beta_{H9}=0$ (5). In a SAS output, β_H estimates the high severity slope after 9 months of low severity storage, $(\beta_H+\beta_{H0})$ estimates the high severity slope after no storage at low severity, $(\beta_H+\beta_{H3})$ estimates the high severity slope after 3 months at low severity, etc. The JMP and SAS parameterizations give different parameters but calculate the same predictions and confidence limits, and the same *p*-value for the significance of the interaction term. Draper and Smith (6) discuss the non-uniqueness of dummy variable parameterizations.

Not including the main effect for *time at L treated as nominal* gives one a non-hierarchical model (7–9). More specifically, the model contains an interaction effect without its corresponding main effect. If one's software defaults to polynomial centering and sum-to-zero parameterization (*e.g.*, JMP), it is important to turn polynomial centering off so that the regression lines will be constrained to cross at the appropriate time points.

METHODOLOGY FOR MULTIPLE BATCHES, PACKAGE TYPES, ETC. (EXAMPLE 3)

Consider a situation where there are all 12 possible combinations of three lots of material, two packages types, and two dosages. Figure 2 illustrates such a situation using real data [(supplementary file 1) in Excel format, (supplementary file 2) in JMP format] which have been rescaled to protect proprietary information. The curves in the figure will be discussed shortly. The statistical model (Table IV) includes "combination" (the 12 combinations of different lots, package types, and dosages), time at the high severity condition, its interaction with combination, time at the low severity condition, and its interaction with combination (The "usual" model for a single storage condition would have one set of time terms instead of two sets). A quadratic term in time was included because plots of the residuals for many of the responses measured without the quadratic term suggested curvature. The high severity quadratic term's *p*-value confirms its statistical significance. Some pooling may be permissible under the ICH guidance (2), but pooling will be discussed in Example 4. Testing the additivity assumption is also postponed to example 4. The solid curves represent the predicted response averaged over the 12 combinations. In accord with the guidance (2), the short dashed curves represent one-sided 95% upper confidence limits of the mean for the "worst combination" at each time point.[1] The "short–long dashed" curves are discussed in the next section.

METHODOLOGY FOR POOLING NON-SIGNIFICANT TERMS (EXAMPLE 4)

It is sometimes desirable to reduce the full model described above by pooling non-significant terms. Section B.3.2 of the ICH guidance Q1E "Evaluation of Stability Data" (2) recommends a methodology for doing so. Tsong, Chen, and Chen (4) proposed a more rigid procedure and pointed out that such rigidity is desirable in regulated environments. Neither their procedure nor the guidance address the dual storage condition situation described herein, but both can easily be generalized to cover dual storage conditions. For example, the guidance recommends testing the equality of a set of slope terms before testing the equality of the corresponding intercepts. A physically reasonable generalization of that restriction would be to test the equality of slope terms under less stressful conditions before testing corresponding slopes under more stressful conditions.

The example 3 data already shown in Fig. 2 were used for this analysis. The degrees of freedom associated with combination were broken into the main effects and interactions of lot, package type, and dosage. Table V lists the model reduction steps. The above-described generalization of Tsong, Chen, and Chen's (4) restriction on the order removing terms via pooling tests and the cutoffs recommended in the guidance (2) ($p=0.25$ for terms involving lot, $p=0.05$ otherwise) were used. The final model is shown in Table VI. The short–long dashed curves in Fig. 2 are one-sided 95% confidence limits of the mean for the "worst combination"[1] at each time point using the reduced model.

When the additivity assumption was checked, the interaction between time at condition L (treated as nominal) and time at condition H was statistically significant (Table VII, $p=0.02$). However, as can be seen from Fig. 2, the data for samples that spent 12 months at condition L were sparse at the time of the analysis, so the black plus sign for 12 months at

[1] "Worst" is defined as the combination with the highest upper confidence limit at that particular time point. This is slightly more conservative than choosing the upper confidence limit for the worst combination at the target shelf life, but the worst combination can vary with time point, and target shelf lives can change. When two-sided limits are involved, there are two worst combinations, one with the highest upper and one with the lowest lower confidence limit.

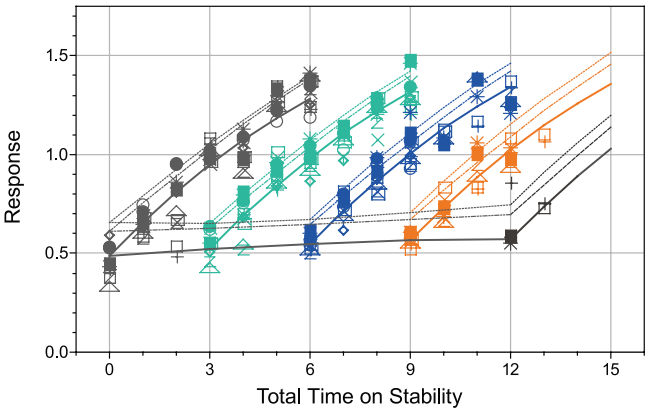

Fig. 2. Results of examples 3 and 4. *Solid lines* are averages of the predicted responses for each combination using the full model of example 3. *Short dashed lines* are one-sided upper 95% confidence limits ("UCL's") for the worst combination at each time point using the full model (example 3). *Short–long dashed lines* are analogous UCL's using the reduced model (example 4). *Black* indicates pure storage condition L or pure condition H. *Black* is also used for 12 months at condition L followed by condition H. *Green*, *blue*, and *red* are after 3, 6, or 9 months at condition L. The 12 different symbols are the data for the 12 different combinations of lots, package types, and doses. No legend is provided for the symbols because the identities of the combinations are not relevant to the material covered in this paper

condition L without any storage at condition H appears potentially influential. When that single point was excluded, the test became non-significant ($p=0.089$). Taken together, one might be suspicious of the additivity assumption but not ready to reject it out of hand.

Because there was also a quadratic term in time at condition H, the dependence of the *initial* slopes at condition H as a function of time at condition L (treated as nominal) were calculated (Table VII). They showed monotonic behavior, so a prudent investigator would be well advised to be highly suspicious of the additivity assumption. Some might also include an interaction term between time at L treated as nominal and the quadratic term in time at H, but others might not want to introduce a third order term into a second order model.

Fortunately, the initial slopes *decreased* with increasing time at condition L. Since the response in this example is clearly one where growth is undesirable, one could simply say that incorrectly assuming additivity for this response is a safe and conservative approximation.

DISCUSSION

Extrapolation is provisional whether one is dealing with a single storage condition or dual storage conditions, but there is a difference in the nature of the extrapolations. Imagine that one is part way through a single storage condition study. One needs to assume that the linear behavior seen thus far is representative of a truly linear mechanism. If the extrapolated confidence limit is near the acceptance limit that assumption may be risky. However, in the dual storage situation, the stability limiting factor is apt to be condition H, not condition L and the experimental designs in Figs. 1 and 2 give one *real-time* data for condition H quite early in the study. Therefore, extrapolation may actually be less risky in the dual storage condition scenario than in the single storage condition scenario. In either case, it is assumed that one will continue the study and obtain real-time data.

The model can be used for fitting real-time data or for making extrapolations like those shown in Figs. 1 and 2 or for modeling the effect of temperature excursions. If one is willing to assume that the changes with time at the two conditions are additive, extrapolation and temperature excursions are straightforward. Examination of residuals and/or testing for interaction terms between condition L and condition H can be employed to check the assumption of additivity.

If significant non-additivity is found, one's ability to extrapolate or model excursions may be seriously limited. When trying to extrapolate in the presence of non-additivity, the interaction term can take several forms:

1. "Time at condition L" multiplied by "time at condition H."
2. "Time at L treated as nominal" by time at H (as in example 2).
3. "Was it at condition L, yes/no" by time at condition H?

Interaction model 1 makes an assumption that may be too strong. If model 2 is used for extrapolation, the slope estimates at high severity following different time periods at

Table IV. ANCOVA Results for Example 3

Analysis of variance				
Source	Degrees of Freedom	Sum of squares	Mean square	*F* ratio
Model	37	18.290755	0.494345	89.6006
Error	232	1.279991	0.005517	*p*-value
C. total	269	19.570745		<0.0001

Effect tests				
Source	Degrees of Freedom	Sum of squares	*F* ratio	*p*-value
Combination	11	0.252928	4.1676	<0.0001
Time at high severity	1	14.950818	2,709.856	<0.0001
(Time at high severity)^2	1	0.145419	26.3573	<0.0001
Combination* time at high severity	11	0.096441	1.5891	0.1028
Time at low severity	1	0.151567	27.4717	<0.0001
(Time at low severity)^2	1	0.006758	1.2250	0.2695
Combination* time at low severity	11	0.045599	0.7514	0.6883

Asterisk's denote interaction

Table V. Model Reduction Steps for Example 4

Step	Term removed from model	*p* value for removal	RSquare after removal	Number of parameters (including intercept)
Full model			0.9346	38
1	Lot*dosage*package*time at low severity	0.7161	0.9344	36
2	Lot*dosage*package*time at high severity	0.4101	0.9339	34
3	Lot*dosage*package	0.7376	0.9337	32
4	Lot*dosage*time at low severity	0.7194	0.9336	30
5	Lot*dosage*time at high severity	0.8013	0.9334	28
6	Lot*dosage	0.3141	0.9328	26
7	Dosage*package*time at low severity	0.6990	0.9327	25
8	Dosage*package*Time at high severity	0.8645	0.9327	24
9	Dosage*package	0.5137	0.9326	23
10	Lot*package*time at low severity	0.6959	0.9324	21
11	Dosage*time at low severity	0.7039	0.9324	20
12	Dosage*time at high severity	0.6292	0.9323	19
13	Dosage	0.1756	0.9318	18
14	Package*time at low severity	0.1020	0.9311	17
15	Time at low severity*time at low severity	0.5472	0.9310	16

Asterisk's denote interaction

low severity are not mutually supportive. Model 3 might make sense in a situation involving freeze–thaw effects. It could take the form

$$Y = \alpha + \beta_L t_L + \beta_H t_H + D_{FT}\beta_{HFT} t_H + \varepsilon \quad (6)$$

where D_{FT} is 1 when the sample experienced prior freezing and 0 when it did not. In that case β_H would represent the high severity slope for samples that were not pre-frozen and (β_H+β_{HFT}) would represent the slope of samples that were pre-frozen. As with example 2, different dummy variable parameterizations could be used. In some cases (*e.g.*, example 4) one may be able to argue that incorrectly assuming additivity is conservative.

One can make cogent arguments for and against combining the analyses for two different conditions when sequential storage is not envisioned. Proponents of combining analyses can argue that combining forces them to have a common intercept or set of intercepts. This makes physical sense because the drug does not "know" what conditions it will be stored at before storage begins. On the other hand, opponents can argue that, if the analyses are separate, an inappropriate model for one storage condition (*e.g.*, failing to detect that a curvature term is needed) will not distort the results for the other storage conditions. The opponents can also argue that an "astute statistician" may notice and wonder why the separate analyses approach yields contradictory intercepts. On the other hand, proponents can reply that the condition with the inappropriate model will distort the residuals associated with the other condition and lead an "astute statistician" to notice the problem. Combining the analyses increases statistical power because it pools their error terms. On the other hand, it assumes that the variance is the same for both storage conditions.

The guidance (2) discusses experimental situations where it may be appropriate to omit the intercept term for package type. As with conventional models, if this is done using software that automatically centers polynomials (*e.g.*,

Table VI. Reduced (Pooled) Model for Example 4

Analysis of variance				
Source	Degrees of Freedom	Sum of squares	Mean square	*F* ratio
Model	15	18.220313	1.21469	228.4681
Error	254	1.350432	0.00532	*p*-value
C. total	269	19.570745		<0.0001

Effect tests				
Source	Degrees of Freedom	Sum of squares	*F* ratio	*p*-value
Lot	2	0.1030296	9.6893	<0.0001
Package	1	0.0565908	10.6440	0.0013
Lot*package	2	0.0132123	1.2425	0.2904
Time at high severity	1	2.6185428	492.5163	<0.0001
Time at high severity*time at high severity	1	0.1532119	28.8173	<0.0001
Lot*time at high severity	2	0.0171894	1.6166	0.2006
Package*time at high severity	1	0.0328766	6.1837	0.0135
Lot*package*time at high severity	2	0.0227098	2.1357	0.1203
Time at low severity	1	0.2295664	43.1787	<0.0001
Lot*time at low severity	2	0.0173683	1.6334	0.1973

Asterisk's denote interaction

Table VII. Model of Table VI with Interaction Term Added

Analysis of variance (all observations included)

Source	Degrees of Freedom	Sum of squares	Mean square	*F* ratio
Model	19	18.281650	0.962192	186.6021
Error	250	1.289096	0.005156	*p*-value
C. total	269	19.570745		<0.0001

Effect tests

Source	Degrees of Freedom	Sum of squares	*F* ratio	*p*-value
Lot	2	0.11245696	10.9046	<0.0001
Package	1	0.06520488	12.6455	0.0005
Lot*package	2	0.01565987	1.5185	0.2211
Time at high severity	1	0.74902459	145.2616	<0.0001
Time at high severity*time at high severity	1	0.20034756	38.8543	<0.0001
Lot*time at high severity	2	0.01847395	1.7914	0.1689
Package*time at high severity	1	0.05098128	9.8870	0.0019
Lot*package*time at high severity	2	0.02509476	2.4334	0.0898
Time at low severity	1	0.24541873	47.5951	<0.0001
Lot*time at low severity	2	0.02628713	2.5490	0.0802
Time at low severity (categorical)*time at high severity	4	0.06133636	2.9738	0.0200

Parameter estimates related to slope at condition H

Term	Estimate	Standard error	t ratio	*p*-value
Intercept	0.4576857	0.014882	30.76	<0.0001
Parameters not related to high severity slope omitted from table				
Time at high severity	0.1622858	0.013465	12.05	<0.0001
Time at high severity*time at high severity	−0.008516	0.001366	−6.23	<0.0001
Parameters not related to high severity slope omitted from table				
Time at low severity (categorical)[0]*time at high severity	0.0297259	0.012853	2.31	0.0216
Time at low severity (categorical)[3]*time at high severity	0.0228799	0.012163	1.88	0.0611
Time at low severity (categorical)[6]*time at high severity	0.0142547	0.011604	1.23	0.2205
Time at low severity (categorical)[9]*time at high severity	−0.001228	0.011957	−0.10	0.9183
Time at low severity (categorical)[12]*time at high severity	−0.065632	0.045244	−1.45	0.1481

Condition H initial slopes for example 4 (averaged over all combinations)

Time spent at condition L	Parameters using JMP parameterization	Initial slope at condition H	
		Including all data	Excluding one influential observation
0	$\beta_H+\beta_{H0}$	0.192012	0.191341
3	$\beta_H+\beta_{H3}$	0.185166	0.185543
6	$\beta_H+\beta_{H6}$	0.176540	0.178121
9	$\beta_H+\beta_{H9}$	0.161058	0.165800
12	$\beta_H+\beta_{H12}$	0.096653	0.127160

Asterisk's denote interaction

JMP) it is necessary to turn off polynomial centering when doing so.

It would be straightforward to model more than two storage conditions by adding (an) additional set(s) of slope terms. On the other hand, an Arrhenius approach (10,11) might be more useful when three or more storage conditions are involved because it would add fewer degrees of freedom to the model. The Arrhenius approach does not convey that advantage with only two storage conditions. In particular, the models herein are not meant for the accelerated stability assessments studied by Porter (10) and by Carella (11) wherein high severity experiments were used in screening experiments to predict low severity shelf life.

If mixed (random and fixed) effects models (Murphy and Hofer, (12); Shao and Chen, (13)) are desired, the slope terms described herein might be treated as random effects. Incorporating a second set of storage time terms may be compatible with the quantile regression models proposed by Quinlan, Schwenke, and Stroup (14,15). However, we have not attempted those generalizations.

Shao and Chow (16,17) proposed an interesting two-step method for dual storage conditions. It is unclear whether their method could be easily implemented using commercial software.

ACKNOWLEDGEMENTS

James S. Bergum is thanked for the excellent mentorship during the first author's tenure at Bristol-Myers Squibb. MannKind Corporation is thanked for permission to publish. The reviewers are thanked for the many valuable suggestions. One reviewer is particularly thanked for his "eagle eye."

REFERENCES

1. Box GEP, Hunter WG, Hunter JS. Statistics for experimenters. New York: Wiley; 1978. Chapter 14.
2. ICH Guidance for Industry. Q1E evaluation of stability data. Food and Drug Administration, 2004. http://www.fda.gov/RegulatoryInformation/Guidances/ucm128092.htm. Accessed 10 June 2010
3. Neter J, Wasserman W, Kutner MH. Applied linear statistical models, 2nd edn. In: Irwin RD, editor. Chapter 10. Chicago: Homewood; 1985.
4. Tsong Y, Chen WJ, Chen CW. ANCOVA approach for shelf life analysis of stability study of multiple factor designs. J Biopharm Stat. 2003;13:375–93. doi:10.1081/BIP-120022761 Accessed 11 Dec 2010.

5. SAS Institute. JMP statistics and graphics guide release 6. Cary: SAS; 2005. p. 862–3.
6. Draper NR, Smith H. Applied regression analysis. 2nd ed. New York: Wiley; 1981. p. 246.
7. Nelder JA. The selection of terms in response-surface models—how strong is the weak-heredity principle? Am Stat. 1998;52:315–8. http://www.jstor.org/pss/2685433. Accessed 10 June 2010.
8. Peixoto JL. Hierarchical variable selection in polynomial regression models. Am Stat. 1987;41:311–3. http://www.jstor.org/pss/2684752. Accessed 10 June 2010.
9. Peixoto JL. A property of well-formulated polynomial regression models. Am Stat. 1990;44:26–30. http://www.jstor.org/pss/2684952. Accessed 10 June 2010.
10. Porter WR. Solid state degradation mechanisms: impact on drug product design and stability test design. Presented at the 33rd Annual Midwest Biopharmaceutical Statistics Workshop, 25 May 2010, Muncie, IN. http://www.mbswonline.com/presentationyear.php?year=2010. Accessed 22 June 2010.
11. Carella, AJ. Accelerated stability assessment program for packaged solid dosage forms. Presented at the 33rd Annual Midwest Biopharmaceutical Statistics Workshop, 25 May 2010, Muncie, IN. http://www.mbswonline.com/presentationyear.php?year=2010. Accessed 22 June 2010.
12. Murphy JR, Hofer J. Establishing shelf life, expiry limits, and release limits. Drug Inf J. 2002;36:769–81.
13. Shao J, Chen L. Prediction bounds for random shelf-lives. Stat Med. 1997;16:1167–73. doi:10.1002/(SICI)1097-0258(19970530)16:10<1167::AID-SIM524>3.0.CO;2-5 Accessed 15 December 2010.
14. Quinlan W, Schwenke J, Stroup W. Direct approach to shelf life estimation. Poster presentation at the 31st Annual Midwest Biopharmaceutical Statistics Workshop 5/19-21/08, Muncie IN. http://www.pqri.org/commworking/minutes/pdfs/dptc/sslwg/Addl/Quinlan_-_Poster_-_Shelf_Life_Estimation.pdf Accessed 10 June 2010.
15. Schwenke J. Reconsidering shelf life: an update from the PQRI Stability Shelf Life Working Group, Presentation at 31st Annual Midwest Biopharmaceutical Statistics Workshop. 19–21 May 2008, Muncie, IN.
16. Shao J, Chow SC. Two-phase shelf-life estimation. Stat Med. 2001;20:1239–48. doi:10.1002/sim.783 Accessed 11 Dec 2010.
17. Chow SC, Shao J. Statistics in drug research: methodologies and recent developments. In: Chapter 4.4. London: CRC Press; 2002.