Available online at www.sciencedirect.com
ScienceDirect
Journal of Pharmaceutical and Biomedical Analysis 45 (2007) 171–175
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Short communication

Uncertainty of measurement and error in stability studies

Robert T. Magari *

Beckman Coulter Inc., 11800 SW 147 Avenue, Miami, FL 33196-2500, USA

Received 12 January 2007; received in revised form 30 April 2007; accepted 2 May 2007
Available online 10 May 2007

Abstract

It is required that shelf life be determined based on the lower limit of the confidence interval of the estimate from the stability tests. Simulations indicate that a 1-year prediction of shelf life will have approximately 1 month of error. However, this is product specific and is related to the uncertainty of measurement and experimental design. Factors associated with product and experimental design, such as degradation rate, number of time points, implementing a full versus a reduced design, etc., can significantly affect the error of shelf life. Uncertainty in measurement is positively correlated to the amount of error through the manufacturing lot-to-lot variability, precision of the analytical method and calibrator. Experimental design can control random variability and actually can reduce error by increasing number of lots and replicates in stability tests. The decision on the number of lots and replicates will be a balancing act between the uncertainty of the measurement, design and other practical considerations.

Keywords: Uncertainty of measurements; Accelerated stability; Shelf life; Experimental design; Degradation; Confidence intervals

1. Introduction

Shelf life of a product that is estimated from the stability tests has a certain degree of error. To minimize any risk associated with this estimation it is required that shelf life be determined as the earliest time at which the 95% confidence limit for the mean response intersects the proposed criterion [1,2]. For products that degrade with time (i.e. the characteristic of interest decreases as time increases) shelf life should be reported as the lower one-sided 95% confidence limit of the estimate [3]. While calculations of confidence intervals are addressed in several studies [3,4], it is common in biopharmaceutical practice to consider that the measured characteristic is normally distributed and to use the percentiles from normal or student distribution to calculate confidence limits for the estimates. Meeker and Escobar [4] used simulation to obtain parametric bootstrap confidence limits of the time-to-failure distribution.

The difference between the point estimate of shelf life and its lower confidence limits depends on the width of the confidence interval, which is positively related to the amount of error. For relatively wide intervals, the determination of shelf life may be conservative and sometimes practically meaningless. Optimum allocation in an experiment will reduce the variability associated with error. FDA [1,5] has outlined some fundamental design considerations. Designs based on different optimality criteria are given in Lin and Chen [5], Boulanger and Escobar [6], and Hedayat et al. [7] while Zhang and Meeker [8] described a Bayesian approach that uses the estimation precision at specified use conditions to find optimum test plans. Appropriate modeling [9–11] of random effects like batch-to-batch differences, will also improve the accuracy for estimating shelf life. The design approach is relatively efficient for reducing error but there are certain practical limitations on the numbers of test to be conducted, time intervals, etc.

Other factors affecting the amount of error are the precision of measurement, calibration of methods or measuring devices, sample handling, manufacturing variability, etc. The cumulative variability of these factors is recognized as uncertainty of measurement and is defined in ISO documentations [12] as, 'a parameter associated with the results of a measurement that characterizes the dispersion of the value that could reasonably be attributed to the measurand.' Furthermore, uncertainty is interrelated to another meteorological concept, traceability, which relates the measured result to a recognized reference through an unbroken chain of calibrations or comparisons. Each step of this process involves a measurement procedure that contributes

* Tel.: +1 305 380 4239; fax: +1 305 380 3699.
E-mail address: Robert.Magari@Coulter.com.

0731-7085/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2007.05.003

some uncertainty. Cumulative combinations of these individual uncertainties represent the uncertainty of measurement that provides information about the closeness of the result to the true value, as well as a quantitative estimate of the quality of a result generated from an analytical method or measuring device. There are several article published on this subject [13–17], but we are not aware of any publication that addresses error in stability studies from the perspective of the uncertainty of measurement. Thus, the purpose of this paper is to show how the random variability in the measurement is related to the amount of error of estimated shelf life and to provide information that would assist in designing stability tests and predict stability within certain limits of error.

Paper is partitioned into three main sections, uncertainty of measurement, degradation model, and experimental design. Relationships between these sections are established by considering a hypothetical product X, with certain performance characteristics, that is undergoing stability testing. Degradation data for product X are simulated for different experimental design scenarios and statistically analyzed to estimate shelf life.

2. Uncertainty of measurement

2.1. Model

There are three sources that can contribute to total uncertainty. First, the uncertainty of the calibrator is the direct link to the traceability chain. This value is usually provided by the manufacturer of the calibrator. The second contributor of uncertainty is the imprecision of the analytical method. This refers to repeatability as well as reproducibility as described in the ISO literature. For a given calibration cycle, the calibrator determines the average result, while the degree of imprecision determines the width of the scatter around this average. Sample is the third source of uncertainty. The constitution and nature of sample may vary from raw material to biological specimen, and itemizing the sources of this uncertainty can be intricate. Sample uncertainty may includes manufacturing variability of the material, pre-analytical effects, and losses in time caused by metabolism, simple chemical reaction or denaturation, sample non-homogeneity, storage, stability, other specific conditions, etc. Manufacturing variability of the material represented in the sample is a 'true' random component of sample uncertainty. The contribution of the other elements may be excluded from the uncertainty by a careful definition of the measuring system in space and time. For example, if we are measuring a biological sample that drifts in time, we can remove the amount of drift from the sample uncertainty by defining the time of the measurement.

Measurement result from an analytical procedure can be expressed as:

$$M = C + P + S \quad (1)$$

where M is the measurement, C the calibrator, P the analytical procedure, and S is the sample effect. We assume that all effects are independent from each other and normally distributed. When there is no need to calibrate, C has mean zero and standard deviation, σ_c^2 that represents the uncertainty of the calibrator. $c \neq 0$ if calibration is needed. This will only shift the average up/down but will not affect the uncertainty. The result from an analytical procedure, P, is a function of the sample being measured and the precision of the procedure, $P \sim N(\mu_P, \sigma_P)$, where μ_P is the mean result of the sample and σ_P is the imprecision of the analytical procedure. The average of sample effects, S is zero for samples coming from a stable and in control manufacturing process, as well as stored and handled under recommended conditions. Thus, $S \sim N(0, \sigma_S)$, where σ_S is the manufacturing variability, mainly represented as lot-to-lot variation. Since all individual effects are independent and normally distributed, M will be normally distributed as well, with mean, μ_M, and σ_M standard deviation of uncertainty. Furthermore, $\mu_M = \mu_P$ and σ_M is calculated from its components as:

$$\sigma_M = \sqrt{\sigma_c^2 + \sigma_S^2 + \sigma_P^2}$$

Uncertainty can also be expressed as ±range, $3\sigma_M$ or as coefficient of variation:

$$\mathrm{CV}\,(\%) = \frac{\sigma_M}{\bar{X}} \times 100$$

2.2. Performance data

Let us consider a hypothetical product X with an average result of one for the characteristic that we are interested ($\mu_M = 1$). Let also assume that this product is available from five different vendors that have manufacturing (lot-to-lot) variability from 1% to 5%, with 1% being the best process and 5% the worse process. These values are detailed in Table 1. We are going to use a certain analytical procedure to measure the characteristic of interest. This procedure has an imprecision of 2%. We also ran a commercial calibrator that has an uncertainty of 1%. Analytical procedure did not need any recalibration.

Model (1) is used to calculate the uncertainty of measurement presented as CV% in Table 2 or as standard deviation and ranges in Table 3. Lot-to-lot manufacturing variability is the only contributor to sample uncertainty. The other causes of sample uncertainty are considered to be negligible. Manufacturing variability of vendors is important in determining the amount of uncertainty associated with a single measurement when the same analytical procedure and calibrator is used. The choice of vendors cannot be based solely on the amount of

Table 1
Uncertainties of different sources

Source	CV (%)	S.D.	±Range
Lot-to-lot	1.0	0.01	0.03
	2.0	0.02	0.05
	3.0	0.03	0.08
	4.0	0.04	0.10
	5.0	0.05	0.13
Precision	2.0	0.02	0.05
Calibration	1.0	0.01	0.03

Table 2
Coefficient of variations of uncertainties

Lot-to-lot (%)	Precision (%)	Calibration (%)	Measurement (%)
1.0	2.0	1.0	2.4
2.0	2.0	1.0	3.0
3.0	2.0	1.0	3.7
4.0	2.0	1.0	4.6
5.0	2.0	1.0	5.5

uncertainty but a vendor should not be selected if the range of uncertainty of measurement is greater than a tolerable range for that characteristic.

3. Degradation model

We consider product X degrades in time according to a first-order degradation reaction [9]. This pattern is modeled as:

$$M = \alpha \exp(-\delta t) + \varepsilon \quad (2)$$

where M is the measured result, α the result at time zero, δ the degradation rate, t the time ($t > 0$), and ε is the experimental error. Error is a pooled estimate of residuals at each time points considering that variances are homogenous. Using Arrhenius relationship, the degradation of the product at elevated temperatures will be accelerated by the following acceleration factor:

$$\lambda = \exp\left[\frac{E_a}{0.00199}\left(\frac{1}{T_s} - \frac{1}{T_e}\right)\right] \quad (3)$$

where T_s, T_e are storage and elevated temperatures, respectively and E_a is the activation energy (kcal mol^{-1}). Degradation at storage temperature as a function of elevated temperatures can be expressed as:

$$M = \alpha \exp(-\delta_s \lambda t) + \varepsilon \quad (4)$$

δ_s is the degradation rate at storage temperature. At time zero $\mu_M = \alpha$, while after 100 days ($t = 100$), $\mu_M = \alpha \exp(-\delta_s \lambda 100)$. A maximum likelihood method can be used to estimate the parameters of the models. The standard errors of the estimates are computed based on the inverse of the Hessian matrix (the matrix of second derivatives). Shelf life of the product is calculated as:

$$\hat{t}_{Stab} = \frac{\log(Crit) - \log(\hat{\alpha})}{\hat{\delta}_s} \quad (5)$$

where Crit is a critical level where the essential performance characteristics of the product are within the specification. We are using the value of Crit = 0.9 in this paper. Standard error of shelf life is actually the standard deviation of the non-liner function (5), and it is consequently dependent on the errors of the estimates of α and δ. There is no closed form equation to calculate the standard deviation of the above function. Delta method [18] based on the Taylor series of the first derivatives of the function can be used to obtain the approximate error of shelf life.

Table 3
Standard deviation and range of uncertainty of measurement

CV (%)	S.D.	±Range	Lower	Upper
2.4	0.0245	0.06	0.94	1.06
3.0	0.0300	0.08	0.92	1.08
3.7	0.0374	0.10	0.90	1.10
4.6	0.0458	0.12	0.88	1.12
5.5	0.0548	0.14	0.86	1.14

Table 4
Error of shelf life for different uncertainty of measurement and number of lots in real time stability test

Uncertainty of measurement (%)	Number of lots				
	2	3	4	5	6
2.4	3.2%	2.7%	2.4%	2.1%	1.9%
3.0	3.9%	3.4%	3.0%	2.7%	2.5%
3.7	4.7%	4.2%	3.8%	3.5%	3.2%
4.6	5.8%	5.2%	4.8%	4.4%	4.1%
5.5	6.8%	6.3%	5.7%	5.3%	4.9%

4. Experimental design

4.1. Real time stability test

Product X is stored at normal storage temperature (25 °C) and monitored for a period of time until it fails. One to six lots are tested at 22 different time points. At each time point a random sample from each lot is tested in two to five replicates. Time points are: 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 130, 160, 200, 250, 300, 350, 400, 450, 500, 550, and 600 days. As it is mentioned before the average of product X is $\mu_M = 1$. Data for combinations of lots, replicates and time points are simulated using a lot-to-lot variability of 1%, 2%, 3%, 4%, and 5%, imprecision of 2% and calibrator uncertainty of 1%. We used SAS® 9.1.3 [19] to analyze the data based on model (2) and calculating shelf life and its error as described in (5).

Errors of estimated shelf life are shown in Tables 4 and 5 in terms of coefficient of variations. Errors presented in Table 4 are calculated for three replicates while calculations in Table 5 are based on 3.7% uncertainty of measurement. All the correlation coefficients between the uncertainty of measurements and error of shelf life for all number of lots presented in Table 4 are above 0.94. Regression slopes range from 1.16 for two lots to 0.97 for six lots experiments indicating that the increase of error due to uncertainty can be controlled by testing more lots. The relationship between the uncertainty of measurements, number

Table 5
Error of shelf life for combination of different number of replicates and lots in real time stability test

Number of replicates	Number of lots					
	1	2	3	4	5	6
2	5.6%	5.5%	5.0%	4.5%	4.1%	3.8%
3	5.3%	4.7%	4.2%	3.8%	3.5%	3.2%
4	4.9%	4.2%	3.8%	3.4%	3.1%	2.9%
5	4.5%	3.8%	3.4%	3.1%	2.8%	2.6%

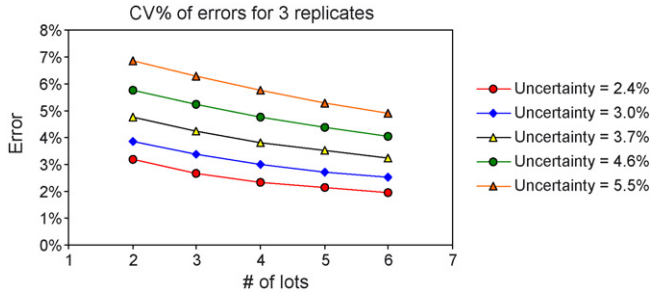

Fig. 1. Error of shelf life in real time stability tests as a function of uncertainty of measurements and number of lots.

of lots, and the number of replicates is also shown graphically in Figs. 1 and 2. There is a curvature associated with the increase on the number of lots and replicates in figures. However, for the levels of uncertainties that we selected, there is no evidence on the formation of a plateau indicating that further increase will not reduce error. In real time stability tests with three replicates and 3.7% uncertainty of measurement, the error of shelf life will be reduced by 0.6% when testing two lots instead of one (Table 5). Error of shelf life will be reduced by 0.5% when the number of lots is three, 0.4% when testing four lots and it is reduced by 0.3% for consecutive increases up to six lots. ICH [2] recommends using at least three primary batches (lots) for stability studies. Results in Tables 4 and 5 support the fact that using three lots would significantly reduce error for estimating shelf life in comparison to one lot. However, there is no evidence to conclude that increasing the number of lots beyond 6 will not reduce error.

Let assume that it is required that product *X* has a shelf life of 365 days. To determine that this requirement is valid, product *X* will be subjected to real time stability by testing three lots and three randomly selected replicates for each lot. Because of a manufacturing variability of 3%, imprecision of the analytical method that measures the characteristic of 2%, and 1% uncertainty of the calibrator, the measured results for product *X* will have an uncertainty of 3.7%. This converts to a range of ±0.1 units. Consequently, it is expected that a measured result to be in the range of 0.9–1.1 units at time zero, but as time progresses this range will shift downwards because of the degradation. Estimating shelf life involves the process of changing from the original units of the characteristic to time units like hours, days, months, etc. This is accomplished by models (2) and (5). Actual simulation for product *X* indicated that the estimated shelf life has an error of 4.2% (Table 5). This converts to a range of ±39.5 days, indicating that product *X* should perform within specification for an estimated 365 + 39.5 = 404.5 days in real time stability test to meet the 1 year requirement.

Table 6
Error of shelf life for combination of different number of replicates and lots in accelerated stability test

Number of replicates	Number of lots		
	1	2	3
2	5.7%	5.5%	4.8%
3	5.0%	4.7%	4.1%
4	4.4%	4.2%	3.6%

4.2. Accelerated stability test

Product *X* is stored at three elevated temperatures (35 °C, 45 °C, and 55 °C) in accelerated stability test and monitored until it fails. One to three lots are tested at each temperature at different time points, where a random sample from each lot is tested in two to four replicates. Data for the combinations of lots, replicates and time points are simulated using a lot-to-lot variability of 3%, imprecision of 2%, and calibrator uncertainty of 1%. As before we used SAS® 9.1.3 [19] to analyze the data based on model (4) and calculating shelf life and its error as described in (5).

Errors of estimated shelf life are shown in Table 6 and graphically in Fig. 3. For three replicates, error of shelf life is reduced by only 0.3% when testing two lots instead of one lot, but this reduction is doubled (0.6%) when three of lots are tested. Increasing the number of replicates is also important for reducing error and monitoring precision of the method during the course of the study [20]. The biggest gain is achieved when this number is increased from two to three replicates. Trends of error change are consistent for different lots and replicates numbers (Fig. 3). Let consider that accelerated stability testing at three elevated temperatures is going to be used to validate the requirement that product *X* has a shelf life of 365 days. Based on the simulated data and statistical calculations the estimated shelf life from this test will have an error of 4.1% when three lots of the product are tested in three randomly selected replicates for time point at each temperature. This converts to ±38.5 days indicating that

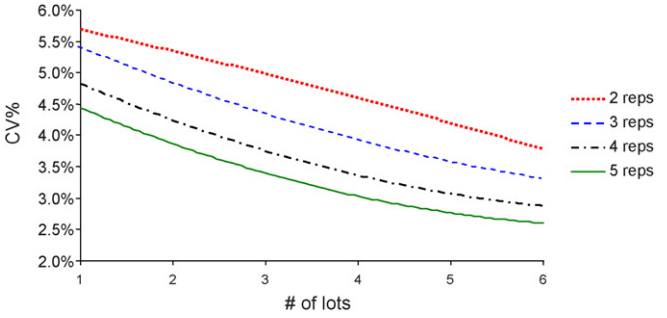

Fig. 2. Error of shelf life for different replicates and lots in real time stability tests.

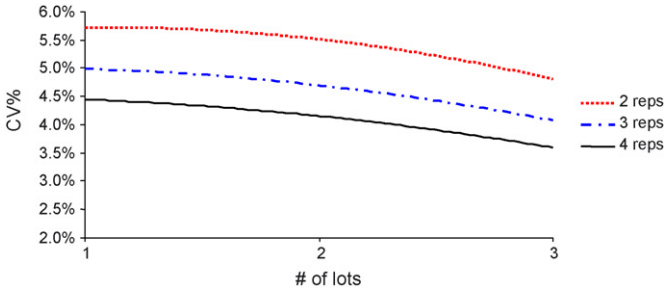

Fig. 3. Error of shelf life for different replicates and lots in accelerated stability tests.

the point estimate of shelf life from the accelerated stability test should be at least 365 + 38.5 = 403.5 days for product *X* to meet the 1 year requirement.

5. Discussion

The requirement that shelf life be reported as the lower confidence limit of the estimate signifies the importance of reducing the amount of error and the width of the confidence interval in stability studies. Our models and simulations indicate that a 1 year prediction of shelf life will have approximately 1 month of error. This certainly depends on the specifics of the product, testing protocol, and other related issues. In this paper we grouped them in terms of uncertainty of measurement and experimental design. Degradation model is a part of this general equation as well. However, we have considered only the simple exponential model since the majority of biopharmaceutical products exhibit some sort of approximation to the first order kinetics.

Uncertainty of measurement is strongly associated with error. Controlling the variability of the testing material and using an analytical procedure that delivers accurate results with high degree of precision would considerably reduce the uncertainty of measurement and consequently, the amount of error and the width of the confidence interval. The amount of error is also associated with experimental design. Experimental design is a planning of stability tests in time and space to make possible the statistical analysis of the data based on the degradation model and to accommodate the uncertainty associated with the measurement. Number of lots and replicates are two important elements of the experimental design in stability tests. Actually, error can be reduced below the level of uncertainty of measurement by increasing the number of replicates and lots. The decision on the number of lots and replicates will be a balancing act between the uncertainty of the measurement, experimental design, required tolerance on the final estimates and practical issues of conducting the tests.

References

[1] Food and Drug Administration, Guidelines for Submitting Documentation for the Stability of Human Drugs and Biologics, Center for Drug and Biologics, Rockville, Maryland, 1987.
[2] International Conference on Harmonization. Q1E. Evaluation for stability data, 2003, Rockville, Maryland.
[3] J. Shao, S. Chow, Biometrics 50 (1994) 753–763.
[4] W.Q. Meeker, L.A. Escobar, C.J. Lu, Technometrics 40 (1998) 89–99.
[5] T.D. Lin, C.W. Chen, J. Biopharm. Stat. 13 (2003) 334–354.
[6] M. Boulanger, L.A. Escobar, Technometrics 36 (1994) 260–272.
[7] A.S. Hedayat, X. Yan, L. Lin, J. Biopharm. Stat. 16 (2006) 35–39.
[8] Y. Zhang, W.Q. Meeker, Technometrics 48 (2006) 49–60.
[9] S. Chow, J. Shao, Biometrics 47 (1991) 1071–1079.
[10] R.T. Magari, J. Pharm. Sci. 91 (2002) 893–899.
[11] I.T. Some, P. Bogaerts, R. Hanus, M. Hanocq, J. Dubois, Int. J. Pharm. 184 (1999) 165–172.
[12] International Organization for Standardization, Guide to the Expression of Uncertainty in Measurements, ISO, Geneva, Switzerland, 1995.
[13] J. Kristiansen, J.M. Christensen, Ann. Clin. Biochem. 35 (1998) 371–379.
[14] J. Kristiansen, Clin. Chem. Lab. Med. 30 (2001) 920–931.
[15] J.S. Krouwer, Clin. Chem. 49 (2003) 1818–1821.
[16] J. Kristiansen, Clin. Chem. 49 (2003) 1822–1829.
[17] A. Konnert, C. Berding, S. Arends, Clin. Chem. Lab. Med. 44 (2006) 1175–1182.
[18] P. Billingsley, Probability and Measure, John Wiley & Sons Inc., New York, NY, 1986.
[19] SAS Institute, Stat User's Guide, Version 9.1.3 Service Pack 2, SAS Institute, Cary, NC, 2003.
[20] C. Agut, STP Pharma Pratiques 14 (2004) 311–324.