European Medicines Agency

August 2003
CPMP/ICH/420/02

ICH Topic Q 1 E
Evaluation of Stability Data

Step 5

NOTE FOR GUIDANCE ON EVALUATION OF STABILITY DATA

(CPMP/ICH/420/02)

TRANSMISSION TO CPMP	February 2002
RELEASE FOR CONSULTATION	February 2002
DEADLINE FOR COMMENTS	August 2002
FINAL APPROVAL BY CPMP	February 2003
DATE FOR COMING INTO OPERATION	August 2003

7 Westferry Circus, Canary Wharf, London, E14 4HB, UK
Tel. (44-20) 74 18 85 75 Fax (44-20) 75 23 70 40
E-mail: mail@emea.eu.int http://www.emea.eu.int

TABLE OF CONTENTS

EVALUATION OF STABILITY DATA

1. INTRODUCTION

1.1 Objectives of the Guideline

This guideline is intended to provide recommendations on how to use stability data generated in accordance with the principles detailed in the ICH guideline "Q1A(R) Stability Testing of New Drug Substances and Products" (hereafter referred to as the parent guideline) to propose a retest period or shelf life in a registration application. This guideline describes when and how extrapolation can be considered when proposing a retest period for a drug substance or a shelf life for a drug product that extends beyond the period covered by "available data from the stability study under the long-term storage condition" (hereafter referred to as long-term data).

1.2 Background

The guidance on the evaluation and statistical analysis of stability data provided in the parent guideline is brief in nature and limited in scope. The parent guideline states that regression analysis is an appropriate approach to analyzing quantitative stability data for retest period or shelf life estimation and recommends that a statistical test for batch poolability be performed using a level of significance of 0.25. However, the parent guideline includes few details and does not cover situations where multiple factors are involved in a full- or reduced-design study.

This guideline is an expansion of the guidance presented in the Evaluation sections of the parent guideline.

1.3 Scope of the Guideline

This guideline addresses the evaluation of stability data that should be submitted in registration applications for new molecular entities and associated drug products. The guideline provides recommendations on establishing retest periods and shelf lives for drug substances and drug products intended for storage at or below "room temperature"*. It covers stability studies using single- or multi-factor designs and full or reduced designs.

***Note:** The term "room temperature" refers to the general customary environment and should not be inferred to be the storage statement for labeling.

ICH Q6A and Q6B should be consulted for recommendations on the setting and justification of acceptance criteria, and ICH Q1D should be referenced for recommendations on the use of full- versus reduced-design studies.

2. GUIDELINES

2.1 General Principles

The design and execution of formal stability studies should follow the principles outlined in the parent guideline. The purpose of a stability study is to establish, based on testing a minimum of three batches of the drug substance or product, a retest period or shelf life and

label storage instructions applicable to all future batches manufactured and packaged under similar circumstances. The degree of variability of individual batches affects the confidence that a future production batch will remain within acceptance criteria throughout its retest period or shelf life.

Although normal manufacturing and analytical variations are to be expected, it is important that the drug product be formulated with the intent to provide 100 percent of the labeled amount of the drug substance at the time of batch release. If the assay values of the batches used to support the registration application are higher than 100 percent of label claim at the time of batch release, after taking into account manufacturing and analytical variations, the shelf life proposed in the application can be overestimated. On the other hand, if the assay value of a batch is lower than 100 percent of label claim at the time of batch release, it might fall below the lower acceptance criterion before the end of the proposed shelf life.

A systematic approach should be adopted in the presentation and evaluation of the stability information. The stability information should include, as appropriate, results from the physical, chemical, biological, and microbiological tests, including those related to particular attributes of the dosage form (for example, dissolution rate for solid oral dosage forms). The adequacy of the mass balance should be assessed. Factors that can cause an apparent lack of mass balance should be considered, including, for example, the mechanisms of degradation and the stability-indicating capability and inherent variability of the analytical procedures.

The basic concepts of stability data evaluation are the same for single- versus multi-factor studies and for full- versus reduced-design studies. Data from formal stability studies and, as appropriate, supporting data should be evaluated to determine the critical quality attributes likely to influence the quality and performance of the drug substance or product. Each attribute should be assessed separately, and an overall assessment should be made of the findings for the purpose of proposing a retest period or shelf life. The retest period or shelf life proposed should not exceed that predicted for any single attribute.

The decision tree in Appendix A outlines a stepwise approach to stability data evaluation and when and how much extrapolation can be considered for a proposed retest period or shelf life. Appendix B provides (1) information on how to analyze long-term data for appropriate quantitative test attributes from a study with a multi-factor, full or reduced design, (2) information on how to use regression analysis for retest period or shelf life estimation, and (3) examples of statistical procedures to determine poolability of data from different batches or other factors. Additional guidance can be found in the references listed; however, the examples and references do not cover all applicable statistical approaches.

In general, certain quantitative chemical attributes (e.g., assay, degradation products, preservative content) for a drug substance or product can be assumed to follow zero-order kinetics during long-term storage[1]. Data for these attributes are therefore amenable to the type of statistical analysis described in Appendix B, including linear regression and poolability testing. Although the kinetics of other quantitative attributes (e.g., pH, dissolution) is generally not known, the same statistical analysis can be applied, if appropriate. Qualitative attributes and microbiological attributes are not amenable to this kind of statistical analysis.

The recommendations on statistical approaches in this guideline are not intended to imply that use of statistical evaluation is preferred when it can be justified to be unnecessary. However, statistical analysis can be useful in supporting the extrapolation of retest periods or shelf lives in certain situations and can be called for to verify the proposed retest periods or shelf lives in other cases.

2.2 Data presentation

Data for all attributes should be presented in an appropriate format (e.g., tabular, graphical, narrative) and an evaluation of such data should be included in the application. The values of quantitative attributes at all time points should be reported as measured (e.g., assay as percent of label claim). If a statistical analysis is performed, the procedure used and the assumptions underlying the model should be stated and justified. A tabulated summary of the outcome of statistical analysis and/or graphical presentation of the long-term data should be included.

2.3 Extrapolation

Extrapolation is the practice of using a known data set to infer information about future data. Extrapolation to extend the retest period or shelf life beyond the period covered by long-term data can be proposed in the application, particularly if no significant change is observed at the accelerated condition. Whether extrapolation of stability data is appropriate depends on the extent of knowledge about the change pattern, the goodness of fit of any mathematical model, and the existence of relevant supporting data. Any extrapolation should be performed such that the extended retest period or shelf life will be valid for a future batch released with test results close to the release acceptance criteria.

An extrapolation of stability data assumes that the same change pattern will continue to apply beyond the period covered by long-term data. The correctness of the assumed change pattern is critical when extrapolation is considered. When estimating a regression line or curve to fit the long-term data, the data themselves provide a check on the correctness of the assumed change pattern, and statistical methods can be applied to test the goodness of fit of the data to the assumed line or curve. No such internal check is possible beyond the period covered by long-term data. Thus, a retest period or shelf life granted on the basis of extrapolation should always be verified by additional long-term stability data as soon as these data become available. Care should be taken to include in the protocol for commitment batches a time point that corresponds to the end of the extrapolated retest period or shelf life.

2.4 Data Evaluation for Retest Period or Shelf Life Estimation for Drug Substances or Products Intended for Room Temperature Storage

A systematic evaluation of the data from formal stability studies should be performed as illustrated in this section. Stability data for each attribute should be assessed sequentially. For drug substances or products intended for storage at room temperature, the assessment should begin with any significant change at the accelerated condition and, if appropriate, at the intermediate condition, and progress through the trends and variability of the long-term data. The circumstances are delineated under which extrapolation of retest period or shelf life beyond the period covered by long-term data can be appropriate. A decision tree is provided in Appendix A as an aid.

2.4.1 No significant change at accelerated condition

Where no significant change occurs at the accelerated condition, the retest period or shelf life would depend on the nature of the long-term and accelerated data.

2.4.1.1 Long-term and accelerated data showing little or no change over time and little or no variability

Where the long-term data and accelerated data for an attribute show little or no change over time and little or no variability, it might be apparent that the drug substance or product will remain well within the acceptance criteria for that attribute during the proposed retest period or shelf life. In these circumstances, a statistical analysis is normally considered unnecessary but justification for the omission should be provided. Justification can include a discussion of the change pattern or lack of change, relevance of the accelerated data, mass balance, and/or other supporting data as described in the parent guideline. Extrapolation of the retest period or shelf life beyond the period covered by long-term data can be proposed. The proposed retest period or shelf life can be up to twice, but should not be more than 12 months beyond, the period covered by long-term data.

2.4.1.2 Long-term or accelerated data showing change over time and/or variability

If the long-term or accelerated data for an attribute show change over time and/or variability within a factor or among factors, statistical analysis of the long-term data can be useful in establishing a retest period or shelf life. Where there are differences in stability observed among batches or among other factors (e.g., strength, container size and/or fill) or factor combinations (e.g., strength-by-container size and/or fill) that preclude the combining of data, the proposed retest period or shelf life should not exceed the shortest period supported by any batch, other factor, or factor combination. Alternatively, where the differences are readily attributed to a particular factor (e.g., strength), different shelf lives can be assigned to different levels within the factor (e.g., different strengths). A discussion should be provided to address the cause for the differences and the overall significance of such differences on the product. Extrapolation beyond the period covered by long-term data can be proposed; however, the extent of extrapolation would depend on whether long-term data for the attribute are amenable to statistical analysis.

- Data not amenable to statistical analysis

Where long-term data are not amenable to statistical analysis, but relevant supporting data are provided, the proposed retest period or shelf life can be up to one-and-a-half times, but should not be more than 6 months beyond, the period covered by long-term data. Relevant supporting data include satisfactory long-term data from development batches that are (1) made with a closely related formulation to, (2) manufactured on a smaller scale than, or (3) packaged in a container closure system similar to, that of the primary stability batches.

- Data amenable to statistical analysis

If long-term data are amenable to statistical analysis but no analysis is performed, the extent of extrapolation should be the same as when data are not amenable to statistical analysis. However, if a statistical analysis is performed, it can be appropriate to propose a retest period or shelf life of up to twice, but not more than 12 months beyond, the period covered by long-term data, when the proposal is backed by the result of the analysis and relevant supporting data.

2.4.2 Significant change at accelerated condition

Where significant change* occurs at the accelerated condition, the retest period or shelf life would depend on the outcome of stability testing at the intermediate condition, as well as at the long-term condition.

***Note:** The following physical changes can be expected to occur at the accelerated condition and would not be considered significant change that calls for intermediate testing if there is no other significant change:
softening of a suppository that is designed to melt at 37ºC, if the melting point is clearly demonstrated,
failure to meet acceptance criteria for dissolution for 12 units of a gelatin capsule or gel-coated tablet if the failure can be unequivocally attributed to cross-linking.

However, if phase separation of a semi-solid dosage form occurs at the accelerated condition, testing at the intermediate condition should be performed. Potential interaction effects should also be considered in establishing that there is no other significant change.

2.4.2.1 No significant change at intermediate condition

If there is no significant change at the intermediate condition, extrapolation beyond the period covered by long-term data can be proposed; however, the extent of extrapolation would depend on whether long-term data for the attribute are amenable to statistical analysis.

- Data not amenable to statistical analysis

When the long-term data for an attribute are not amenable to statistical analysis, the proposed retest period or shelf life can be up to 3 months beyond the period covered by long-term data, if backed by relevant supporting data.

- Data amenable to statistical analysis

When the long-term data for an attribute are amenable to statistical analysis but no analysis is performed, the extent of extrapolation should be the same as when data are not amenable to statistical analysis. However, if a statistical analysis is performed, the proposed retest period or shelf life can be up to one-and-half times, but should not be more than 6 months beyond, the period covered by long-term data, when backed by statistical analysis and relevant supporting data.

2.4.2.2 Significant change at intermediate condition

Where significant change occurs at the intermediate condition, the proposed retest period or shelf life should not exceed the period covered by long-term data. In addition, a retest period or shelf life shorter than the period covered by long-term data could be called for.

2.5 Data Evaluation for Retest Period or Shelf Life Estimation for Drug Substances or Products Intended for Storage Below Room Temperature

2.5.1 Drug substances or products intended for storage in a refrigerator

Data from drug substances or products intended to be stored in a refrigerator should be assessed according to the same principles as described in Section 2.4 for drug substances or

products intended for room temperature storage, except where explicitly noted in the section below. The decision tree in Appendix A can be used as an aid.

2.5.1.1 No significant change at accelerated condition

Where no significant change occurs at the accelerated condition, extrapolation of retest period or shelf life beyond the period covered by long-term data can be proposed based on the principles outlined in Section 2.4.1, except that the extent of extrapolation should be more limited.

If the long-term and accelerated data show little change over time and little variability, the proposed retest period or shelf life can be up to one-and-a-half times, but should not be more than 6 months beyond, the period covered by long-term data normally without the support of statistical analysis.

Where the long-term or accelerated data show change over time and/or variability, the proposed retest period or shelf life can be up to 3 months beyond the period covered by long-term data if (1) the long-term data are amenable to statistical analysis but a statistical analysis is not performed, or (2) the long-term data are not amenable to statistical analysis but relevant supporting data are provided.

Where the long-term or accelerated data show change over time and/or variability, the proposed retest period or shelf life can be up to one-and-a-half times, but should not be more than 6 months beyond, the period covered by long-term data if (1) the long-term data are amenable to statistical analysis and a statistical analysis is performed, and (2) the proposal is backed by the result of the analysis and relevant supporting data.

2.5.1.2 Significant change at accelerated condition

If significant change occurs between 3 and 6 months' testing at the accelerated storage condition, the proposed retest period or shelf life should be based on the long-term data. Extrapolation is not considered appropriate. In addition, a retest period or shelf life shorter than the period covered by long-term data could be called for. If the long-term data show variability, verification of the proposed retest period or shelf life by statistical analysis can be appropriate.

If significant change occurs within the first 3 months' testing at the accelerated storage condition, the proposed retest period or shelf life should be based on long-term data. Extrapolation is not considered appropriate. A retest period or shelf life shorter than the period covered by long-term data could be called for. If the long-term data show variability, verification of the proposed retest period or shelf life by statistical analysis can be appropriate. In addition, a discussion should be provided to address the effect of short-term excursions outside the label storage condition (e.g., during shipping or handling). This discussion can be supported, if appropriate, by further testing on a single batch of the drug substance or product at the accelerated condition for a period shorter than 3 months.

2.5.2 Drug substances or products intended for storage in a freezer

For drug substances or products intended for storage in a freezer, the retest period or shelf life should be based on long-term data. In the absence of an accelerated storage condition for drug substances or products intended to be stored in a freezer, testing on a single batch at an elevated temperature (e.g., 5°C ± 3°C or 25°C ± 2°C) for an appropriate time period should

be conducted to address the effect of short-term excursions outside the proposed label storage condition (e.g., during shipping or handling).

2.5.3 Drug substances or products intended for storage below -20°C

For drug substances or products intended for storage below -20°C, the retest period or shelf life should be based on long-term data and should be assessed on a case-by-case basis.

2.6 General Statistical Approaches

Where applicable, an appropriate statistical method should be employed to analyze the long-term primary stability data in an original application. The purpose of this analysis is to establish, with a high degree of confidence, a retest period or shelf life during which a quantitative attribute will remain within acceptance criteria for all future batches manufactured, packaged, and stored under similar circumstances.

In cases where a statistical analysis was employed to evaluate long-term data due to a change over time and/or variability, the same statistical method should also be used to analyse data from commitment batches to verify or extend the originally approved retest period or shelf life.

Regression analysis is considered an appropriate approach to evaluating the stability data for a quantitative attribute and establishing a retest period or shelf life. The nature of the relationship between an attribute and time will determine whether data should be transformed for linear regression analysis. The relationship can be represented by a linear or non-linear function on an arithmetic or logarithmic scale. In some cases, a non-linear regression can better reflect the true relationship.

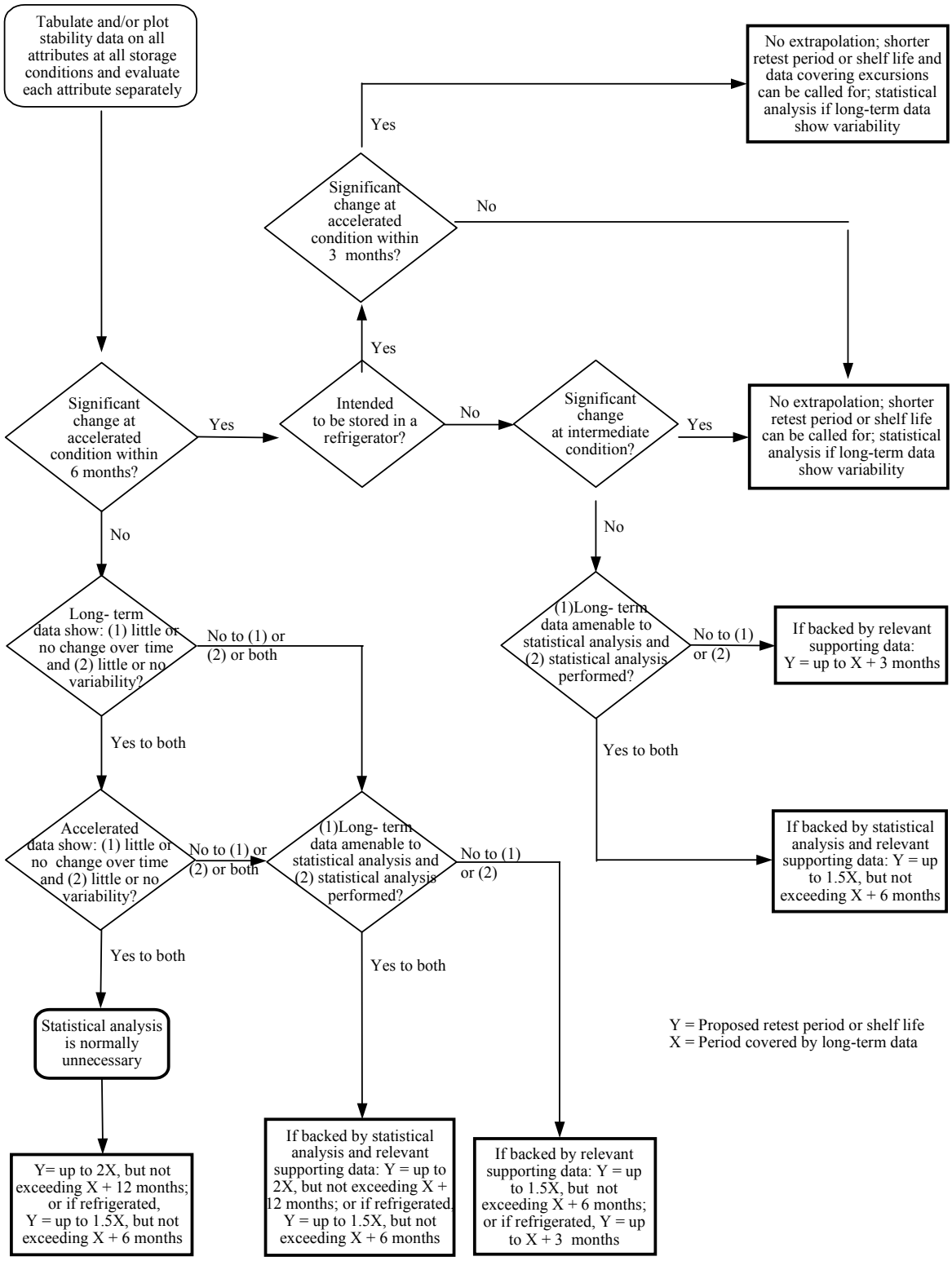
An appropriate approach to retest period or shelf life estimation is to analyze a quantitative attribute (e.g., assay, degradation products) by determining the earliest time at which the 95 percent confidence limit for the mean intersects the proposed acceptance criterion.

For an attribute known to decrease with time, the lower one-sided 95 percent confidence limit should be compared to the acceptance criterion. For an attribute known to increase with time, the upper one-sided 95 percent confidence limit should be compared to the acceptance criterion. For an attribute that can either increase or decrease, or whose direction of change is not known, two-sided 95 percent confidence limits should be calculated and compared to the upper and lower acceptance criteria.

The statistical method used for data analysis should take into account the stability study design to provide a valid statistical inference for the estimated retest period or shelf life. The approach described above can be used to estimate the retest period or shelf life for a single batch or for multiple batches when the data are combined after an appropriate statistical test. Examples of statistical approaches to the analysis of stability data from single or multi-factor, full- or reduced-design studies are included in Appendix B. References to current literature sources can be found in Appendix B.6.

3. APPENDICES

Appendix A: Decision Tree for Data Evaluation for Retest Period or Shelf Life Estimation for Drug Substances or Products (excluding Frozen Products)



Appendix B: Examples of Statistical Approaches to Stability Data Analysis

Linear regression, poolability tests, and statistical modeling, described below, are examples of statistical methods and procedures that can be used in the analysis of stability data that are amenable to statistical analysis for a quantitative attribute for which there is a proposed acceptance criterion.

B.1 Data Analysis for a Single Batch

In general, the relationship between certain quantitative attributes and time is assumed to be linear[1]. Figure 1 shows the regression line for assay of a drug product with upper and lower acceptance criteria of 105 percent and 95 percent of label claim, respectively, with 12 months of long-term data and a proposed shelf life of 24 months. In this example, two-sided 95 percent confidence limits for the mean are applied because it is not known ahead of time whether the assay would increase or decrease with time (e.g., in the case of an aqueous-based product packaged in a semi-permeable container). The lower confidence limit intersects the lower acceptance criterion at 30 months, while the upper confidence limit does not intersect with the upper acceptance criterion until later. Therefore, the proposed shelf life of 24 months can be supported by the statistical analysis of the assay, provided the recommendations in Sections 2.4 and 2.5 are followed.

When data for an attribute with only an upper or a lower acceptance criterion are analyzed, the corresponding one-sided 95 percent confidence limit for the mean is recommended. Figure 2 shows the regression line for a degradation product in a drug product with 12 months of long-term data and a proposed shelf life of 24 months, where the acceptance criterion is not more than 1.4 percent. The upper one-sided 95 percent confidence limit for the mean intersects the acceptance criterion at 31 months. Therefore, the proposed shelf life of 24 months can be supported by statistical analysis of the degradation product data, provided the recommendations in Sections 2.4 and 2.5 are followed.

If the above approach is used, the mean value of the quantitative attribute (e.g., assay, degradation products) can be expected to remain within the acceptance criteria through the end of the retest period or shelf life at a confidence level of 95 percent.

The approach described above can be used to estimate the retest period or shelf life for a single batch, individual batches, or multiple batches when combined after appropriate statistical tests described in Sections B.2 through B.5.

B.2 Data Analysis for One-Factor, Full-Design Studies

For a drug substance or for a drug product available in a single strength and a single container size and/or fill, the retest period or shelf life is generally estimated based on the stability data from a minimum of three batches. When analyzing data from such one-factor, batch-only, full-design studies, two statistical approaches can be considered.
The objective of the first approach is to determine whether the data from all batches support the proposed retest period or shelf life.
The objective of the second approach, testing for poolability, is to determine whether the data from different batches can be combined for an overall estimate of a single retest period or shelf life.

B.2.1 Evaluating whether all batches support the proposed retest period or shelf life

The objective of this approach is to evaluate whether the estimated retest periods or shelf lives from all batches are longer than the one proposed. Retest periods or shelf lives for individual batches should first be estimated using the procedure described in Section B.1 with individual intercepts, individual slopes, and the pooled mean square error calculated from all batches. If each batch has an estimated retest period or shelf life longer than that proposed, the proposed retest period or shelf life will generally be considered appropriate, as long as the guidance for extrapolation in Sections 2.4 and 2.5 is followed. There is generally no need to perform poolability tests or identify the most reduced model. If, however, one or more of the estimated retest periods or shelf lives are shorter than that proposed, poolability tests can be performed to determine whether the batches can be combined to estimate a longer retest period or shelf life.

Alternatively, the above approach can be taken during the pooling process described in Section B.2.2. If the regression lines for the batches are found to have a common slope and the estimated retest periods or shelf lives based on the common slope and individual intercepts are all longer than the proposed retest period or shelf life, there is generally no need to continue to test the intercepts for poolability.

B.2.2 Testing for poolability of batches

B.2.2.1 Analysis of covariance

Before pooling the data from several batches to estimate a retest period or shelf life, a preliminary statistical test should be performed to determine whether the regression lines from different batches have a common slope and a common time-zero intercept. Analysis of covariance (ANCOVA) can be employed, where time is considered the covariate, to test the differences in slopes and intercepts of the regression lines among batches. Each of these tests should be conducted using a significance level of 0.25 to compensate for the expected low power of the design due to the relatively limited sample size in a typical formal stability study.

If the test rejects the hypothesis of equality of slopes (i.e., if there is a significant difference in slopes among batches), it is not considered appropriate to combine the data from all batches. The retest periods or shelf lives for individual batches in the stability study can be estimated by applying the approach described in Section B.1 using individual intercepts and individual slopes and the pooled mean square error calculated from all batches. The shortest estimate among the batches should be chosen as the retest period or shelf life for all batches.

If the test rejects the hypothesis of equality of intercepts but fails to reject that the slopes are equal (i.e., if there is a significant difference in intercepts but no significant difference in slopes among the batches), the data can be combined for the purpose of estimating the common slope. The retest periods or shelf lives for individual batches in the stability study should be estimated by applying the approach described in Section B.1, using the common slope and individual intercepts. The shortest estimate among the batches should be chosen as the retest period or shelf life for all batches.

If the tests for equality of slopes and equality of intercepts do not result in rejection at a level of significance of 0.25 (i.e., if there is no significant difference in slope and intercepts among the batches), the data from all batches can be combined. A single retest period or shelf life can be estimated from the combined data by using the approach described in Section B.1 and

applied to all batches. The estimated retest period or shelf life from the combined data is usually longer than that from individual batches because the width of the confidence limit(s) for the mean will become narrower as the amount of data increases when batches are combined.

The pooling tests described above should be performed in a proper order such that the slope terms are tested before the intercept terms. The most reduced model (i.e., individual slopes, common slope with individual intercepts, or common slope with common intercept, as appropriate) can be selected for retest period or shelf life estimation.

B.2.2.2 Other methods

Statistical procedures[2-6] other than those described above can be used in retest period or shelf life estimation. For example, if it is possible to decide in advance the acceptable difference in slope or in mean retest period or shelf life among batches, an appropriate procedure for assessing the equivalence in slope or in mean retest period or shelf life can be used to determine the data poolability. However, such a procedure should be prospectively defined, evaluated, and justified and, where appropriate, discussed with the regulatory authority. A simulation study can be useful, if applicable, to demonstrate that the statistical properties of the alternative procedure selected are appropriate[7].

B.3 Data Analysis for Multi-Factor, Full-Design Studies

The stability of the drug product could differ to a certain degree among different factor combinations in a multi-factor, full-design study. Two approaches can be considered when analyzing such data.
The objective of the first approach is to determine whether the data from all factor combinations support the proposed shelf life.
The objective of the second approach, testing for poolability, is to determine whether the data from different factor combinations can be combined for an overall estimate of a single shelf life.

B.3.1 Evaluating whether all factor combinations support the proposed shelf life

The objective of this approach is to evaluate whether the estimated shelf lives from all factor combinations are longer than the one proposed. A statistical model that includes all appropriate factors and factor combinations should be constructed as described in Section B.3.2.2.1, and the shelf life should be estimated for each level of each factor and factor combination.

If all shelf lives estimated by the original model are longer than the proposed shelf life, further model building is considered unnecessary and the proposed shelf life will generally be appropriate as long as the guidance in Sections 2.4 and 2.5 is followed. If one or more of the estimated shelf lives fall short of the proposed shelf life, model building as described in Section B.3.2.2.1 can be employed. However, it is considered unnecessary to identify the final model before evaluating whether the data support the proposed shelf life. Shelf lives can be estimated at each stage of the model building process, and if all shelf lives at any stage are longer than the one proposed, further attempts to reduce the model are considered unnecessary.

This approach can simplify the data analysis of a complicated multi-factor stability study compared to the data analysis described in Section B.3.2.2.1.

B.3.2 Testing for poolability

The stability data from different combinations of factors should not be combined unless supported by statistical tests for poolability.

B.3.2.1 Testing for poolability of batch factor only

If each factor combination is considered separately, the stability data can be tested for poolability of batches only, and the shelf life for each non-batch factor combination can be estimated separately by applying the procedure described in Section B.2. For example, for a drug product available in two strengths and four container sizes, eight sets of data from the 2x4 strength-size combinations can be analyzed and eight separate shelf lives should be estimated accordingly. If a single shelf life is desired, the shortest estimated shelf life among all factor combinations should become the shelf life for the product. However, this approach does not take advantage of the available data from all factor combinations, thus generally resulting in shorter shelf lives than does the approach in Section B.3.2.2.

B.3.2.2 Testing for poolability of all factors and factor combinations

If the stability data are tested for poolability of all factors and factor combinations and the results show that the data can be combined, a single shelf life longer than that estimated based on individual factor combinations is generally obtainable. The shelf life is longer because the width of the confidence limit(s) for the mean will become narrower as the amount of data increases when batches, strengths, container sizes and/or fills, etc. are combined.

B.3.2.2.1 Analysis of covariance

Analysis of covariance can be employed to test the difference in slopes and intercepts of the regression lines among factors and factor combinations[7, 8]. The purpose of the procedure is to determine whether data from multiple factor combinations can be combined for the estimation of a single shelf life.

The full statistical model should include the intercept and slope terms of all main effects and interaction effects and a term reflecting the random error of measurement. If it can be justified that the higher order interactions are very small, there is generally no need to include these terms in the model. In cases where the analytical results at the initial time point are obtained from the finished dosage form prior to its packaging, the container intercept term can be excluded from the full model because the results are common among the different container sizes and/or fills.

The tests for poolability should be specified to determine whether there are statistically significant differences among factors and factor combinations. Generally, the pooling tests should be performed in a proper order such that the slope terms are tested before the intercept terms and the interaction effects are tested before the main effects. For example, the tests can start with the slope and then the intercept terms of the highest order interaction, and proceed to the slope and then the intercept terms of the simple main effects. The most reduced model, obtained when all remaining terms are found to be statistically significant, can be used to estimate the shelf lives.

All tests should be conducted using appropriate levels of significance. It is recommended that a significance level of 0.25 be used for batch-related terms, and a significance level of 0.05 be

used for non-batch-related terms. If the tests for poolability show that the data from different factor combinations can be combined, the shelf life can be estimated according to the procedure described in Section B.1 using the combined data.

If the tests for poolability show that the data from certain factors or factor combinations should not be combined, either of two alternatives can be applied: (1) a separate shelf life can be estimated for each level of the factors and of the factor combinations remaining in the model; or (2) a single shelf life can be estimated based on the shortest estimated shelf life among all levels of factors and factor combinations remaining in the model.

B.3.2.2.2 Other methods

Alternative statistical procedures[2-6] to those described above can be applied. For example, an appropriate procedure for assessing the equivalence in slope or in mean shelf life can be used to determine the data poolability. However, such a procedure should be prospectively defined, evaluated, properly justified, and, where appropriate, discussed with the regulatory authority. A simulation study can be useful, if applicable, to demonstrate that the statistical properties of the alternative procedure selected are appropriate[7].

B.4 Data Analysis For Bracketing Design Studies

The statistical procedures described in Section B.3 can be applied to the analysis of stability data obtained from a bracketing design study. For example, for a drug product available in three strengths (S1, S2, and S3) and three container sizes (P1, P2, and P3) and studied according to a bracketing design where only the two extremes of the container sizes (P1 and P3) are tested, six sets of data from the 3x2 strength-size combinations will be obtained. The data can be analyzed separately for each of the six combinations for shelf life estimation according to Section B.3.2.1, or tested for poolability prior to shelf life estimation according to Section B.3.2.2.

The bracketing design assumes that the stability of the intermediate strengths or sizes is represented by the stability at the extremes. If the statistical analysis indicates that the stability of the extreme strengths or sizes is different, the intermediate strengths or sizes should be considered no more stable than the least stable extreme. For example, if P1 from the above bracketing design is found to be less stable than P3, the shelf life for P2 should not exceed that for P1. No interpolation between P1 and P3 should be considered.

B.5 Data Analysis For Matrixing Design Studies

A matrixing design has only a fraction of the total number of samples tested at any specified time point. Therefore, it is important to ascertain that all factors and factor combinations that can have an impact on shelf life estimation have been appropriately tested. For a meaningful interpretation of the study results and shelf life estimation, certain assumptions should be made and justified. For instance, the assumption that the stability of the samples tested represents the stability of all samples should be valid. In addition, if the design is not balanced, some factors or factor interactions might not be estimable. Furthermore, for different levels of factor combinations to be poolable, it might have to be assumed that the higher order factor interactions are negligible. Because it is usually impossible to statistically test the assumption that the higher order terms are negligible, a matrixing design should be used only when it is reasonable to assume that these interactions are indeed very small, based on supporting data.

The statistical procedure described in Section B.3 can be applied to the analysis of stability data obtained from a matrixing design study. The statistical analysis should clearly identify the procedure and assumptions used. For instance, the assumptions underlying the model in which interaction terms are negligible should be stated. If a preliminary test is performed for the purpose of eliminating factor interactions from the model, the procedure used should be provided and justified. The final model on which the estimation of shelf life will be based should be stated. The estimation of shelf life should be performed for each of the terms remaining in the model. The use of a matrixing design can result in an estimated shelf life shorter than that resulting from a full design.

Where bracketing and matrixing are combined in one design, the statistical procedure described in Section B.3 can be applied.

B.6 References

1. Carstensen, J.T., "Stability and Dating of Solid Dosage Forms" *Pharmaceutics of Solids and Solid Dosage Forms*, Wiley-Interscience, 182-185, 1977

2. Ruberg, S.J. and Stegeman, J.W., "Pooling Data for Stability Studies: Testing the Equality of Batch Degradation Slopes" *Biometrics*, 47:1059-1069, 1991

3. Ruberg, S.J. and Hsu, J.C., "Multiple Comparison Procedures for Pooling Batches in Stability Studies" *Technometrics*, 34:465-472, 1992

4. Shao, J. and Chow, S.C., "Statistical Inference in Stability Analysis" *Biometrics*, 50:753-763, 1994

5. Murphy, J.R. and Weisman, D., "Using Random Slopes for Estimating Shelf-life" *Proceedings of American Statistical Association of the Biopharmaceutical Section*, 196-200, 1990

6. Yoshioka, S., Aso, Y, and Kojima, S., "Assessment of Shelf-life Equivalence of Pharmaceutical Products" *Chem. Pharm. Bull.*, 45:1482-1484, 1997

7. Chen, J.J., Ahn, H., and Tsong, Y., "Shelf-life Estimation for Multifactor Stability Studies" *Drug Inf. Journal*, 31:573-587, 1997

8. Fairweather, W., Lin, T.D., and Kelly, R., "Regulatory, Design, and Analysis Aspects of Complex Stability Studies" *J. Pharm. Sci*., 84:1322-1326, 1995

B.7 Figures

Figure 1

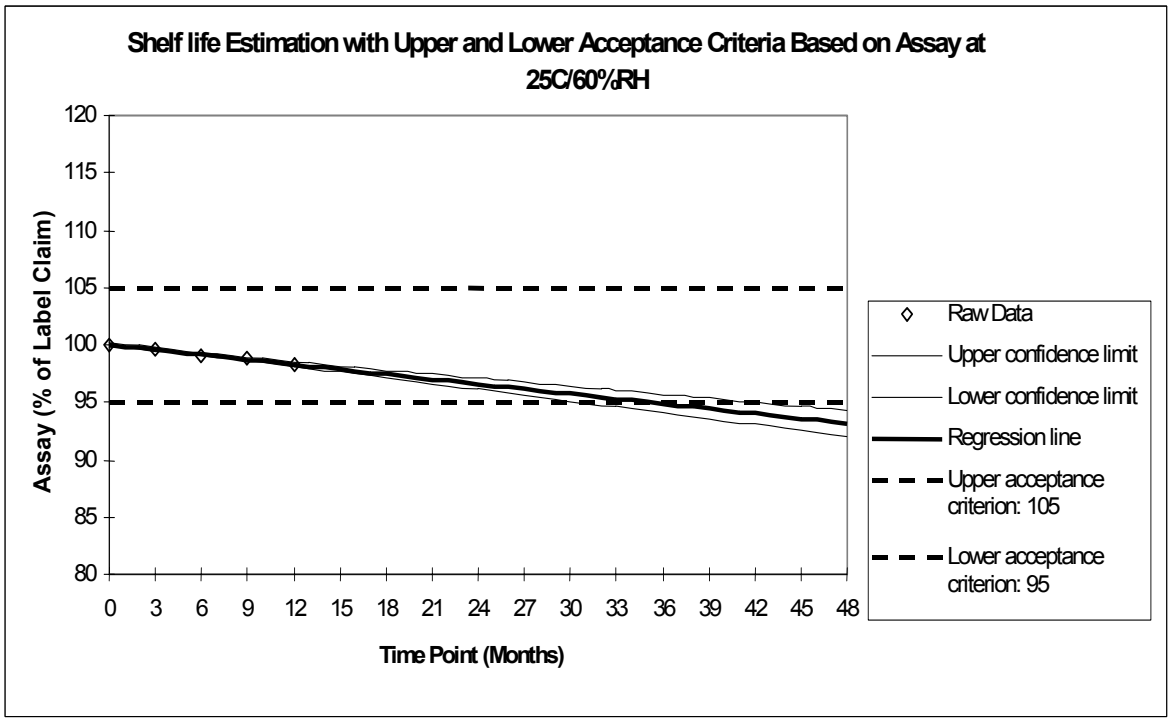


Figure 2