Table 4. Analysis of variance

Factor	Square sum (S)		Degree of Freedom (ψ)		Variance (V)	
Concentration	S_{conc}	3096.35	ψ_{conc}	4	Vconc	774.09
Laboratory	S_R	9953.01	ψ_R	10	V_R	995.30
Experiment	S_{RW}	1225.1	ψ_{RW}	30	V_{RW}	40.84
Residual	S_r	642.26	ψ_r	90	V_r	7.14
Total	S_T	14916.72	ψ_T	134		

Various parameters were calculated from the results shown in Table 3. Similarly, intra-sample pecision, inter-day precision, and inter-laboratory precision (CV) were calculated to be 1.30%, 1.63% and 5.02%, respectively, from the variance values. The confidence interval of σ_{IM}, intermediate precision (standard deviation) with 95% confidence coefficient, was calculated to be $8.0 \leqq \sigma_{IM} \leqq 18.4$.

安定性試験における品質確保基準に関する研究

研究協力者　吉岡澄江　国立医薬品食品衛生研究所　薬品部第二室室長

安定性試験データの解析におけるロット間変動の検定法として、FDA が推奨する ANCOVA 法と新たに開発したレンジに基づく方法の有用性を比較検討した。モンテカルロ法によって発生させた安定性試験データを用いて、両方法よってロット間変動の検定を行い、ロット間変動に対する検出力を比較した結果、0.5%以下の誤差の分析法を用いて安定性データが得られた場合に、レンジに基づく方法は ANCOVA 法と同等の確率でロット間の分解曲線の傾きの差を検出できることが明らかになり、ANCOVA 法に代わる方法として活用できる可能性が示された。

協力研究者
阿曽幸男　国立医薬品食品衛生研究所　主任研究官

A．研究目的

医薬品の有効期間は、経時的な含量低下のように定量的な品質が経時的に変化する医薬品の場合には、通常、回帰曲線の 95%信頼限界の下限値が規格値に等しくなる時点までの期間として設定される。有効期間の設定に用いるデータ数が増大するほど信頼限界の幅が小さくなり、したがって長い有効期間が設定できる。新有効成分含有医薬品の安定性試験に関する ICH 調和ガイドラインには、ロット間での安定性の変動が小さい場合には、試験を行ったすべてのロットから得られたデータを一括し、それに基づいて一つの有効期間の推定値を求め、その値をすべてのロットの有効期間として適用することが認められている。ロット間の変動は、各ロットについて得られる回帰曲線の傾きおよび切片に基づいて、共分散分析(ANCOVA)（資料１）をおこなうことによって検定することができる。しかし、ANCOVA の検出力は定量誤差の大きさによって大きく影響され、定量誤差が大きくなるほど検出力は著しく低下する。その結果、大きい定量誤差を含むデータほど、ロット間の安定性の差を見逃す傾向が大きくなる。

本研究ではこれまでに、ANCOVA 法に代わる方法として、個々のロットで得られる有効期間の推定値についてその同等性を推定値のレンジに基づいて検定する方法（レンジに基づく方法）を提案してきた。このレンジに基づく方法は、ANCOVA 法に比較して、検出力が定量誤差の大きさによって受ける影響が著しく小さいことを前年度までに明らかにした。しかし、この結果は、回帰曲線の傾きと切片の変動を回帰の一様性の検定によって同時に評価するようにデザインされた ANCOVA 法をモデルとして、それに対してレンジに基づく方法を比較した研究によって得られたものであり、現在 FDA が安定性試験ガイダンスに採用している ANCOVA 法、すなわち、回帰曲線の傾きと切片の変動を別々に検定する ANCOVA 法との比較検討は行われていない。

本年度は、レンジに基づく方法を FDA が推奨する ANCOVA 法と比較することを目的とし、モンテカルロ法によって発生させた安定性試験データを用いて両方法にしたがってロット間変動の検定を行い、ロット間変動に対する検出力を比較検討した。

B．研究方法

安定性試験データの発生

医薬品の経時的な分解がゼロ次速度式によって表されると仮定し、３ロットの安定性試験データをモンテカルロ法によって 500 セット発生させた。分解曲線の切片は 100%と仮定した。また、分解曲線の傾きは、３ロット中２ロットでは 0.1%/month、残りの１ロットでは 0.12%/month あるいは 0.13%/month と仮定した。すなわち、１ロットが他のロットより 20%あるいは 30%大きい傾きを持つと仮定した。正規分布（平均０および 0.02～2.0%の標準偏差）から選んだ乱数を分解曲線の理論値に加えて、０、３、６、９、12 および 18 ヶ月時点における測定データを得た。測定の繰り返し回数は２回とした。理論値に加えた定量誤差の分布は、Fig.1 に示すように、正規分布を示した。

ANCOVA 法によるロット間変動の検定

FDA SAS Formulation Stability Program を基に開発された PASG Excel routine (http://pasg.org.uk/excel.htm) を用いて 500 セットの安定性データのロット間変動を評価した。含量規格値は 95-105%とした。ロット間の安定性の差を

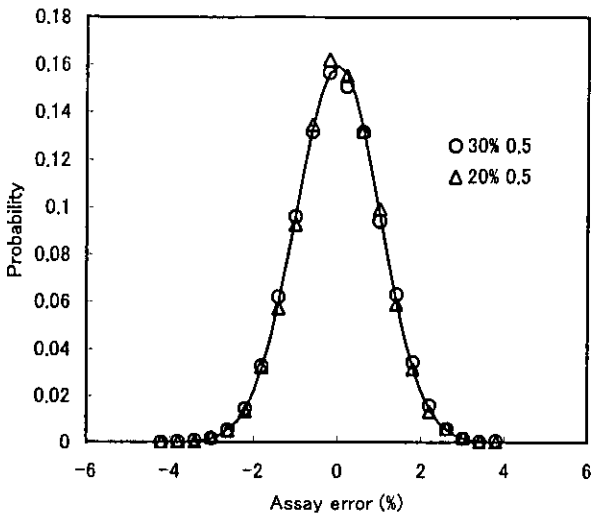


Fig. 1. Distributions of assay errors used for generating the data of 30% stability difference among batches and 0.5% assay error (○), and the data of 20% stability difference and 0.5% assay error (△).
Solid line represents theoretical normal distribution.

見逃す確率（β誤差）として、各ロットの回帰曲線が共通の傾きおよび切片をもつと判断された確率を求めた。

<u>レンジに基づく方法によるロット間変動の検定</u>

上記の 500 セットの安定性データに基づいて、各ロットの有効期間を PASG Excel routine を用いて推定した。個々のロットについて得られた有効期間の推定値のレンジ（3つの推定値の中での最大値と最小値の差）を計算し、レンジが最大値の 15%より小さい場合には、個々のロットからの有効期間の推定値は同等であると判断した。β誤差として、個々のロットからの有効期間が同等であると誤って判断された確率を求めた。

C．研究結果

新有効成分を含有する医薬品の有効期間を設定する際には、18 ヶ月までの長期安定性試験データに基づいて、その医薬品が 3 年以上の有効期間を有するかどうかを判断することが、しばしば求められる。そこで、本研究では 0、3、6、9、12 および 18 ヶ月時点に 2 回の繰り返し測定を行うと仮定して測定データをシミュレートした。3 ロット中 2 ロットについては、規格値の下限が 95%、定量誤差が 0.5%の時に約 3 年の有効期間になるように、回帰曲線の傾きは 0.1%/month とした。残りの 1 ロットは、2 つのロットより約 6 ヶ月短い有効期間を有するように、2 つのロットより 20 あるいは 30%大きい

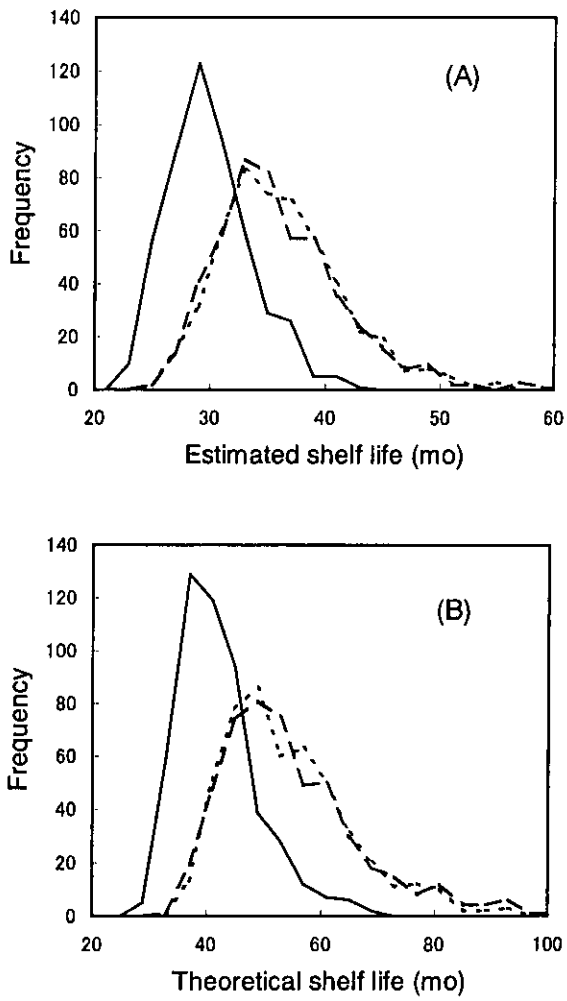


Fig. 2. Distributions of estimated shelf lives (A) and theoretical shelf lives (B) from three batches.
The slope of degradation curve was 0.1%/month for batches A and B, and 0.13%/month for batch C.
Assay error: 0.5%.
Theoretical shelf life is the time at which mean degradation curve intersects the acceptance criterion (95%).

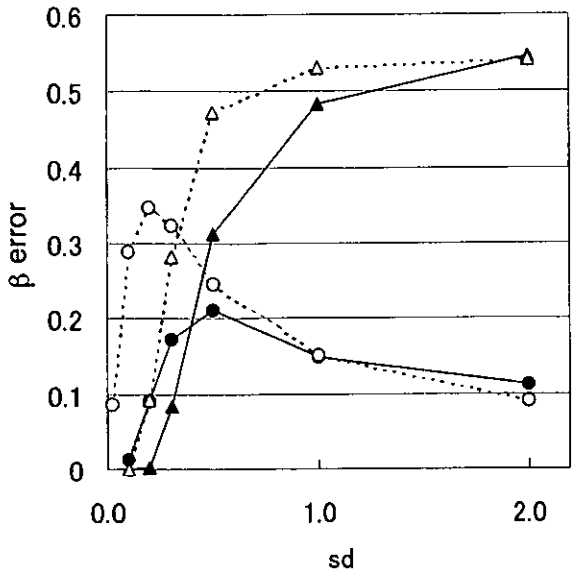


Fig. 3. Effect of assay error on the b error of the ANCOVA approach (△▲) and the Equivalence approach (○●).
Stability difference among batches : 20% (△○) and 30% (▲●)

回帰曲線の傾きを有すると仮定した。Fig.2 に、回帰曲線が規格の下限値を切る時点までの期間として計算した有効期間の理論値および回帰曲線の95％信頼限界の下限が規格値の下限値を切る時点までの期間として計算した有効期間の推定値の分布を示す。一般的に、500 セットのデータはシミュレーションとして少ないと考えられているが、Fig.2 に示すように、同じ傾きの回帰曲線をもつ 2 つのロットについて同様の分布が得られたことから、ANCOVA 法とレンジに基づく方法の比較検討には、500 セットで十分であると考えられる。

ANCOVA 法とレンジに基づく方法のそれぞれで観察されたβ誤差を Fig.3 に示す。ANCOVA 法のβ誤差は、定量誤差の増大とともに大きくなった。一方、レンジに基づく方法のβ誤差は、定量誤差がある値で極大値を示した。分解曲線の傾きがロット間で 30％異なる場合には、定量誤差が 0.3％までの範囲では ANCOVA 法の方がレンジに基づく方法よりも小さなβ誤差を示したが、定量誤差が 0.5％になると、ANCOVA 法のβ誤差の方が大きくなり、さらに定量誤差が大きくなるにしたがって、両方法によるβ誤差の差が大きくなった。これらの結果から、定量誤差が 0.5％以下の場合には ANCOVA 法およびレンジに基づく方法はロット間の安定性の差に対して同等の検出力を持つと考えることができる。また、定量誤差が 0.5％より大きい場合には、レンジに基づく方法の方が、安定性の差を見逃す確立は小さいことが明らかになった。

ロット間での分解曲線の傾きの差が 20％に減少すると、Fig.3 に示すように、ANCOVA 法およびレンジに基づく方法のいずれもβ誤差は 30％の差の場合より増大した。定量誤差が 0.3％では、両方法のβ誤差はほとんど同じであり、それ以下の定量誤差では ANCOVA 法が、またそれ以上の定量誤差ではレンジに基づく方法の方が小さなβ誤差を示した。

0.5％以下の誤差の分析法を用いて安定性データが得られた場合に、レンジに基づく方法はANCOVA 法と同等の確率でロット間の分解曲線の傾きの差を検出できることが明らかになり、ANCOVA 法に代わる方法として活用できる可能性が示された。

D．考察

ANCOVA 法は回帰曲線の平均の傾きからの偏差の総和に基づいてロット間の傾きの差を検定するのに対して、本研究で開発したレンジに基づく方法は、有効期間の推定値の最大値と最高値から計算した単一の値に基づいて有効期間の推定値の差を検定するため、推定される有効期間の変動は考慮されない。したがって、レンジに基づく方法の感度は試験に用いるロットの数に大きく影響される。この問題を避けるためには、推定値の変動を考慮して有効期間の推定値の差を検定する方法を確立することが必要であり、そのための基礎検討が必要と考えられる。

また、本研究において、レンジに基づく方法で用いられた 15％の基準値は前年度までに報告されたように、β誤差が 20％を超えないように設定したものである。この基準値の妥当性を判断するためには、β誤差に加えて、ロット間で安定性の差がない場合に差があると判断してしまうα誤差を考察することが必要である。本研究では ANCOVA 法とレンジに基づく方法をβ誤差に基づいて比較したが、さらにα誤差に基づいて両方法を比較することが必要であり、今後の検討課題であると考えられる。

E．結論

モンテカルロ法によって発生させた安定性試験データを用いてFDAが推奨するANCOVA法およびレンジに基づく方法によってロット間変動の検定を行い、ロット間変動に対する検出力を比較した結果、誤差が0.5％以下の分析法によって安定性データが得られる場合には、レンジに基づく方法はANCOVA法と同等のβ

誤差においてロット間の分解曲線の傾きの差を検出できることが明らかになり、レンジに基づく方法はANCOVA法に代わる方法として活用できる可能性が示された。

F．研究発表

1．論文発表

S.Yoshioka, Y.Aso, S.Kojima: A Comparison of the ANCOVA and range –based approaches for assessing batch-to-batch variability of the stability of pharmaceutical products. Chem. Pharm.Bull., Submitted.

2．学会発表

なし

G．知的所有権の取得状況

1．特許取得

なし

2．実用新案登録

なし

3．その他

ICH Q1E Step-2
EVALUATION OF STABILITY DATA
6 February, 2002

Table of Contents

ICH Q1E Step-2
EVALUATION OF STABILITY DATA

1. INTRODUCTION

1.1 Objectives of the Guidelines

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) on "Stability Testing of New Drug Substances and Products" (hereafter referred to as the parent guideline) to propose a retest period or shelf life. This guideline describes when and how limited extrapolation can be undertaken to propose a retest period for a drug substance or shelf life for a drug product beyond the observed range of data from the long-term storage condition.

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. Although the parent guideline states that regression analysis is an acceptable 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, it includes few details. In addition, the parent guideline does not cover situations where multiple factors are involved in a full or reduced-design study.

1.3 Scope of the Guideline

This guideline, an annex to the parent guideline, is intended to provide a clear explanation of expectations when proposing a retest period or shelf life and storage conditions based on the evaluation of stability data for both quantitative and qualitative test attributes. This guideline outlines recommendations for establishing a retest period or shelf life based on stability data from single or multi-factor and full or reduced-design studies. ICH Q6A and Q6B provide guidance on the setting and justification of acceptance criteria.

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.

A systematic approach should be adopted in the presentation and evaluation of the stability information, which 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).

Where appropriate, attention should be paid to reviewing the adequacy of the mass balance. Factors that can cause an apparent lack of mass balance should be considered, for example, the mechanisms of degradation and the stability-indicating capability and inherent variability of the analytical procedures. 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.

The recommendations in this guideline on statistical approaches 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 the extrapolation of retest periods or shelf lives in certain situations and may be called for to verify the retest periods or shelf lives in other cases.

The basic concepts of stability data evaluation are the same for single- versus multi-factor studies and for full- versus reduced- design studies. Data evaluation from the formal stability studies, and as appropriate, supporting data should be used 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 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.

A flow diagram is provided in Appendix A and some statistical approaches are provided in Appendix B on how to analyze and evaluate long-term stability data for appropriate quantitative test attributes from a study with a multi-factor full or reduced design. The statistical method used for data analysis should consider the stability study design to provide a valid statistical inference for the estimated retest period or shelf life. Appendix B also provides information on how to use regression analysis for retest period or shelf life estimation and examples of statistical procedures to determine poolability of data from different batches or other factors. Additional guidance is provided by the list of references; however, the examples and references do not attempt to cover all other 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. Data for these attributes are therefore amenable to linear regression and poolability testing, as illustrated in Appendix B. Qualitative attributes are not amenable to statistical analysis, and microbiological attributes and certain quantitative attributes (e.g., pH, dissolution) are generally not amenable to the type of statistical analysis described in Appendix B.

2.2 Data presentation

Data for all attributes should be presented in an appropriate format (e.g., tabular, graphical, narrative) and an evaluation of those data should be included in the application. 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

Limited extrapolation to extend the retest period or shelf life beyond the observed range of available long-term data can be proposed in the application, particularly if no significant change is observed at the accelerated condition. Any extrapolation should take into consideration the possible worst-case situation at the time of batch release.

Extrapolation is the practice of using a known data set to infer information about future data sets. An extrapolation of stability data assumes that the same change pattern will continue to apply beyond the observed range of available long-term data. Hence, the use of extrapolation should be justified in terms of, for example, what is known about the mechanisms of degradation, the goodness of fit of any mathematical model, and the existence of relevant supporting data.

The correctness of the assumed change pattern is crucial if extrapolation beyond the available long-term data is contemplated. For example, when estimating a regression line or curve within the available 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 available beyond the length of observed 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 batchesa time point that corresponds to the extrapolated retest period or shelf life.

2.3 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. In general, stability data for each attribute should be assessed sequentially, beginning with significant change, if any, at the accelerated condition and, if appropriate, the intermediate condition, and progressing through the trends and variability of long-term data. The circumstances are delineated under which extrapolation of retest period or shelf life beyond the observed length of long-term data can be appropriate.

2.4.1 No significant change at accelerated condition

Where no significant change occurs at the accelerated condition, the retest period or shelf life setting 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 may be apparent that the drug substance or product will remain well within its acceptance criterion for that attribute during the proposed retest period or shelf life. Under these circumstances, it is normally considered unnecessary to go through a statistical analysis, but justification for the omission should be provided. Justification can include a discussion of the mechanisms of degradation or lack of degradation, relevance of the accelerated data, mass balance, and/or other supporting data as defined in the parent guideline.

Extrapolation of the retest period or shelf life beyond the length of available long-term data can be proposed. A proposed retest period or shelf life up to twice the length of available long-term data can be proposed, but it should not exceed the length of available long-term data by more than 12 months.

2.4.1.2 Long-term or accelerated data showing change over time and 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 considerable differences in stability observed among batches or other factors (e.g., container size and/or fill, strength) or factor combinations (e.g., strength-by-container size and/or fill), the proposed retest period or shelf life should be based on the shortest period supported by the worst batch, 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 a difference on the product. Extrapolation beyond the length of available 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 (for qualitative attributes or certain quantitative attributes)*

When relevant supporting data are provided, a retest period or shelf life up to one and a half times the length of available long-term data can be proposed, but should not exceed the length of available long-term data by more than 6 months. Relevant supporting data include satisfactory long-term data from development batches that are made with a closely related formulation to, manufactured on a smaller scale than, or packaged in a container closure system similar to that of the primary stability batches.

- *Data amenable to statistical analysis*

If a statistical analysis is not performed, the extent of extrapolation should be the same as above (i.e., when relevant supporting data are provided, a retest period or shelf life up to one-and-a-half times the length of available long-term data can be proposed, but should not exceed the length of available long-term data by more than 6 months.) However, if a statistical analysis is performed, it can be appropriate to propose a retest period or shelf life of up to twice the length of available long-term data, when supported by the statistical analysis and supporting data, although this proposed retest period or shelf life should not exceed the length of available long-term data by more than 12 months.

2.4.2 Significant change at accelerated condition

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

*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 (potential interaction effects should also be

considered in establishing that there is no other significant change): (1) softening of a suppository that is designed to melt at 37°C, if the melting point is clearly demonstrated; and (2) failure to meet acceptance criteria for dissolution for 12 units of a gelatin capsule or gel-coated tablet if it can be unequivocally attributed to cross-linking. However, phase separation of semisolid dosage forms at the accelerated condition could call for testing at the intermediate condition.

2.4.2.1 No significant change at intermediate condition

If there is no significant change at the intermediate condition, extrapolation beyond the length of available 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*

Based on an attribute that is not amenable to statistical analysis, a retest period or shelf life can be proposed, when relevant supporting data are provided, but the proposed retest period or shelf life should not exceed the length of available long-term data by more than 3 months.

- *Data amenable to statistical analysis*

If the long-term data for an attribute are amenable to statistical analysis but such an analysis is not performed, the extent of extrapolation would be the same as above. However, if a statistical analysis is performed, it can be appropriate to propose a retest period or shelf life of up to one-and-half times the length of available long-term data, when supported by the statistical analysis and relevant supporting data, but not exceeding the length of available long-term data by more than 6 months.

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 extent of available long-term data. In addition, a shorter retest period or shelf life could be called for. If the long-term data show variability, verification of the retest period or shelf life by statistical analysis can be appropriate.

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 products intended to be stored in a refrigerator should be assessed according to the same principles described throughout this document for the general case pertaining to products intended for "room temperature" storage, except where explicitly noted in the section below. A decision tree is provided in Appendix A as an aid to the guidance below.

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 length of available long-term data can be proposed. The proposed retest period or shelf life can be up to one and a half times the length of available long-term data, but should not exceed the length of available long-term data by more than 6 months.

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 real time data available at the long-term storage condition. No extrapolation can be considered.

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 the real time data available at the long-term storage condition. No extrapolation should be performed. 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 for a period shorter than 3 months.

2.5.2 Drug substances or products intended for storage in a freezer

For drug substances and products intended for storage in a freezer, the retest period or shelf life should be based on the real time data obtained at the long-term storage condition. 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 and products intended for storage below -20°C, the retest period or shelf life should be based on the real time data obtained at the proposed long-term storage conditio. and should be assessed on a case by case basis.

2.3 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. This same method could also be applied to 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. Usually, the relationship can be represented by a linear or non-linear function on an arithmetic or logarithmic scale. Sometimes a non-linear regression can be expected to better reflect the true relationship.

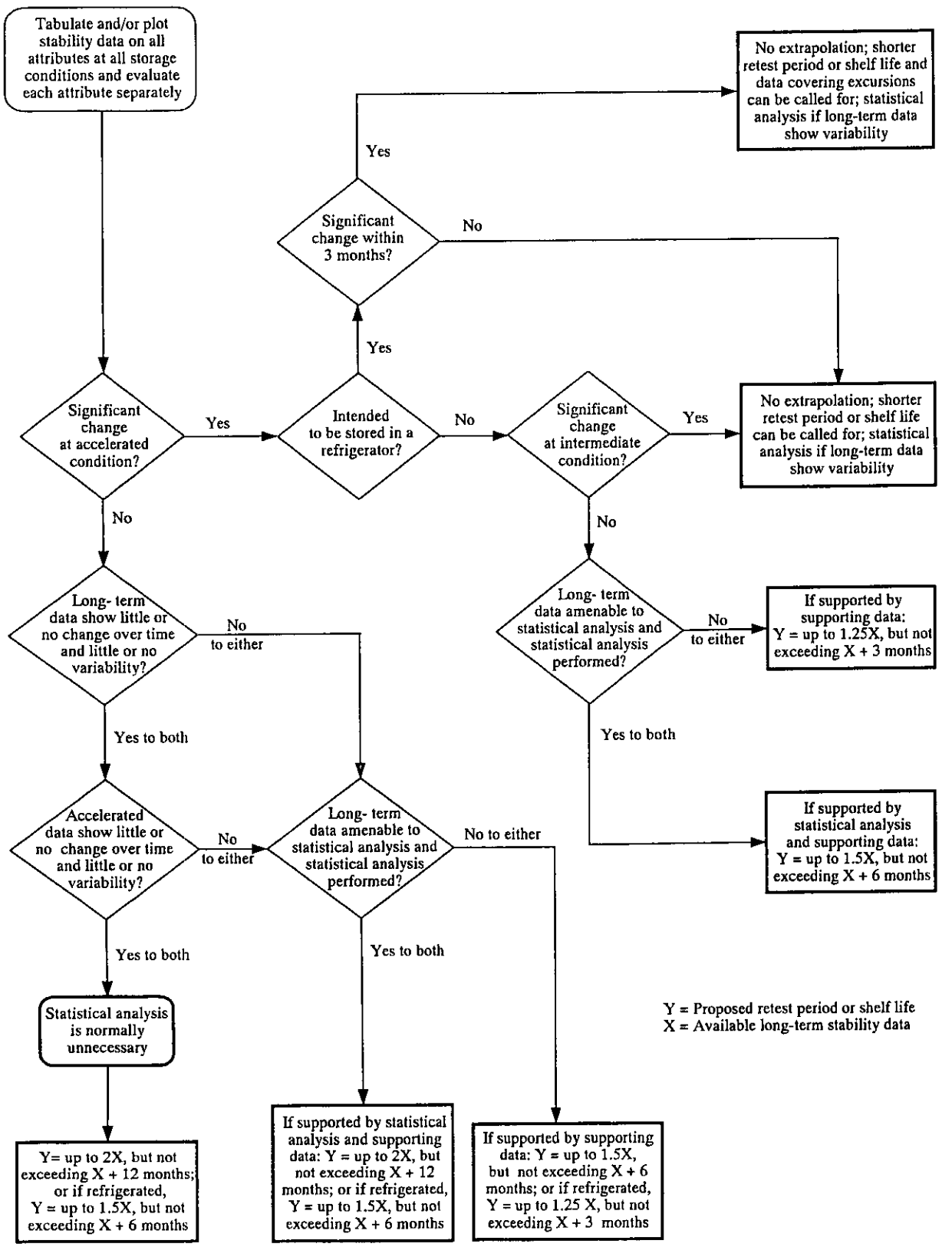
An appropriate approach to retest period or shelf life estimation is to analyze a quantitative attribute by determining the earliest time at which the 95 percent confidence limit for the mean around the regression curve intersects the proposed acceptance criterion.

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 criterion. For an attribute which 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 combined after an appropriate statistical test. Examples of statistical approaches to the analysis of stability data from full, bracketing, and matrixing designs 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 for a quantitative attribute that is amenable to linear regression and for which there is a proposed acceptance criterion.

B.1. Data Analysis for a Single Batch

In certain cases, the relationship between an attribute and time is assumed to be linear.[1] Figure 1a shows the regression line for assay of a product with upper and lower acceptance criteria of 105 percent and 95 percent of label claim, respectively, with 12 months of long-term data. 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. The lower 95 percent confidence limit intersects the lower acceptance criterion at 30 months. Therefore, a proposed shelf life of up to 24 months can be supported by the statistical analysis as long as 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 regression line is recommended. Figure 1b shows the regression line for a degradation product with 12 months of long-term data, where the acceptance criterion is not more than 1.4 percent. The one-sided 95 percent confidence limit for the mean intersects the acceptance criterion at 31 months. Therefore, a proposed retest period or shelf life of 24 months can be supported by the statistical analysis of the degradation data as long as the recommendations in Sections 2.4 and 2.5 are followed.

If the above approach is used, the values of the quantitative attribute (e.g., assay, degradation products) can be expected to remain within acceptance criteria through the end of the retest period or shelf life at a confidence level of 95 percent. If, however, the acceptance criterion for the quantitative attribute calls for individual values, confidence limits for the individual values should be used (e.g., content uniformity for some complex dosage forms).

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. Two approaches can be considered when analyzing such data. The objective of the first approach, testing for poolability, is to determine whether the data from different batches can be combined for an overall estimate of a single shelf life. The objective of the second approach is to determine whether the data from all batches support the proposed shelf life.

B.2.1 Testing for poolability of batches

B.2.1.1 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 formal stability study.

If the test rejects the hypothesis of equality of slopes (i.e., there is a significant difference in slopes among batches), it is considered inappropriate to combine the data from all batches. The retest periods or shelf lives for individual batches in the stability study can then be estimated by applying the approach as described in B.1 using individual intercepts and individual slopes and the pooled mean square error calculated from all data. 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., 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 then be estimated by applying the approach as described in 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., there is no significant difference in slope and intercepts among the batches), the data from all batches can be combined. A single shelf life can be estimated from the combined data by using the approach as described in B.1 and applied to all batches. The estimated shelf life from the combined data is usually longer than that from individual batches because the confidence limit(s) about the regression line will become narrower as the amount of data increases when batches are combined.

The above pooling tests 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 shelf life estimation.

B.2.1.2 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 shelf life among batches, 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, and justified and, where appropriate, discussed with the regulatory authority. A simulation study can be useful, if applicable, to demonstrate the appropriate statistical properties of the alternative procedure selected.[7]

B.2.2 Evaluating whether all batches support proposed petest period or shelf life

The objective of this approach is to evaluate whether some of the batches have retest periods or shelf lives shorter than those proposed. Retest periods or shelf lives for individual batches should first be estimated using the procedure described in B.1 with individual intercepts, individual slopes, and the pooled mean square error calculated from all data. 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 Section 2.4-2.5 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, this approach can be taken during the pooling process described in B.2.1.1. If the regression lines for the batches are found to have a common slope and the estimated shelf lives based on the common slope and individual intercepts are all longer than the proposed shelf life, there is generally no need to continue to test the intercepts for poolability.

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, testing for poolability, is to determine whether the data from different factor combinations can be combined for an overall estimate of a single retest period or shelf life. The objective of the second approach is to determine whether the data from all factor combinations support the proposed retest period or shelf life.

B.3.1 Testing for poolability

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

B.3.1.1 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 retest period or shelf life for each non-batch factor combination can be estimated separately by applying the procedure described in 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 will 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 B.3.1.2.

B.3.1.2 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 retest period or shelf life longer than that estimated based on individual factor combinations is generally obtainable.

The retest period or shelf life is longer because the confidence limit(s) about the estimated regression line will become narrower as the amount of data increases when batches, strengths, container sizes and/or fills, etc. are combined.

B.3.1.2.1 Analysis of covariance

Analysis of covariance can be employed to test the difference in slope and intercept 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 retest period or 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 test for poolability should be specified to determine whether there are statistically significant differences among factors and factor combinations. Generally, the pooling test 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 test 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 significant level of 0.25 be used for any terms involving batch and a significant level of 0.05 be used for terms not involving batch. 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 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 of the factor combinations remaining in the model.

B.3.1.2.2 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 the appropriate statistical properties of the alternative procedure selected.[7]

B.3.2 Evaluating whether all factor combinations support proposed retest period or shelf life

The objective of this approach is to evaluate whether some of the factor combinations have shelf lives shorter than the proposed shelf life. The statistical model should be constructed as described in B.3.1.2.1, and the shelf life can be estimated for each level of each factor and factor combination. If all estimated shelf lives are longer than the proposed shelf life, no further model building is considered necessary and the proposed shelf life will generally be considered 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 B.3.1.2.1 can be employed. However, it is generally 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, and if all shelf lives are longer than the proposed, further modeling is considered unnecessary. This approach can simplify the data analysis of a complicated multi-factor stability study compared to that described in B.3.1.2.1.

B.4. Data Analysis For Bracketing Design Studies

The same statistical procedures as described in B.3 can be applied to the analysis of stability data obtained from a bracketing design. 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 B.3.1.1, or tested for poolability prior to shelf life estimation according to B.3.1.2.

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-by-factor interactions 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-by-factor interactions could not be estimable. Furthermore, for different levels of factor combinations to be poolable, it might have to be assumed that the higher order factor-by-factor interactions are negligible. Because it is impossible to statistically test the assumption that the higher order terms are negligible, this type of matrixing design should be used only when it is reasonable to assume that these interactions are indeed very small, based on supporting data.

606 The statistical procedure described in B.3 can be applied to the analysis of stability data
607 obtained from a matrixing design. The same procedure for pooling the data from
608 different batches, strengths, and/or container sizes and/or fill should be applied.
609 However, since not every combination of factors will be tested at all time points, the
610 statistical analysis should clearly identify the procedure and assumptions used. For
611 instance, the assumptions underlying the model in which interaction terms are
612 negligible should be stated. If a preliminary test is performed for the purpose of
613 eliminating factor combinations from the model, the procedure used should be provided
614 and justified. The final model on which the estimation of shelf life will be based should
615 be stated. The estimation of shelf life should be performed for each of the terms
616 remaining in the model. The use of a matrixing design can result in a shorter estimated
617 shelf life than use of a full design.
618
619 Where bracketing and matrixing are combined in one design, the statistical procedure
620 described in B.3 can be applied.
621
622
623