

Development of Statistical Computational Tools Through Pharmaceutical Drug Development and Manufacturing Life Cycle



Fasheng Li and Ke Wang

Abstract Statisticians at Pfizer who support Chemistry, Manufacturing, and Controls (CMC), and Regulatory Affairs (Reg CMC) have developed many statistical R-based computational tools to enable high efficiency, consistency, and fast turnaround in their routine statistical support to drug product and manufacturing process development. Most tools have evolved into web-based applications for convenient access by statisticians and colleagues across the company. These tools cover a wide range of areas, such as product stability and shelf life or clinical use period estimation, process parameter criticality assessment, and design space exploration through experimental design and parametric bootstrapping. In this article, the general components of these R-programmed web-based computational tools are introduced, and their successful applications are demonstrated through an application of estimating a drug product shelf life based on stability data.

Keywords Statistical computing · Web based tool · R programming · Regression analysis · Design of experiment

1 Introduction

Through the regulatory, chemistry, manufacturing and controls (Reg CMC) development lifecycle of a drug product, a series of compendial requirements, quality standards, and performance criteria must be well established and met. It usually takes years to perform data collection, analysis, and reporting on chemical process, formulation and manufacturing process, and analytical method development. Common practice in current pharmaceutical industry to “optimize” product compositions,

F. Li (✉) · K. Wang
Pharmaceutical Science and Manufacturing Statistics, Pfizer Inc., MS 8220-2356,
Eastern Point Road, Groton, CT 06340, USA
e-mail: fasheng.li@pfizer.com

K. Wang
e-mail: ke.wang2@pfizer.com

© Pfizer, Inc. 2019
R. Liu and Y. Tsong (eds.), *Pharmaceutical Statistics*,
Springer Proceedings in Mathematics & Statistics 218,
https://doi.org/10.1007/978-3-319-67386-8_8

manufacturing processes, and analytical methods is to apply designed experiments (DOEs), statistical models and statistical sampling techniques. Data generated in these procedures, which could be in a large amount, are usually analyzed or evaluated by statisticians or statistically trained professionals with commercial statistical software systems such as Design Expert, SAS, or SAS-JMP. With the increasing demand of statistical application and the challenge of limited number of trained statisticians, it is desirable to develop computational tools to conduct routine statistical analyses in more efficient and consistent ways. The computational tools promote consistency, efficiency, and reproducibility for routine statistical analysis. Version control, monitoring and regular maintenance are an integral part of developing the computational tools. The features of the computational tools align well with the requirements of Title 21 Code of Federal Regulations (CFR) Part 11 that the software systems should be readily available for and subject to FDA inspection (3) [1]. Working as statisticians at Pfizer supporting pharmaceutical development and Reg CMC, we have identified many opportunities and areas that benefit from statistical computation tools. Most tools are developed using a language such as R and have evolved into web-based applications for easy access by statisticians and colleagues at Pfizer. This article introduces the general requirements and structure of web-based statistical tools. The computational application is demonstrated through one tool which evaluates product stability and predicts shelf life or clinical use period.

2 Overview of Available Web-Based Statistical Tools

2.1 *Introduction of Components of Web-Based Statistical Applications*

Figure 1 illustrates three standard components of typical web-based applications: computer server, GUI platform server and application user. In practice, applet authors utilize the application servers to construct the computation script and the graphical user interface (GUI) of the application, and ensure successful communication between the application servers (usually a web browser) and the computer servers. General users only need to compile data into a required format by the applications. For a statistical computational application, additional software systems, such as R and SAS need to be installed onto the computer server for statistical analysis. The following section provides an overview of the web-based statistical applications developed by Pfizer pharmaceutical development and Reg CMC statisticians.

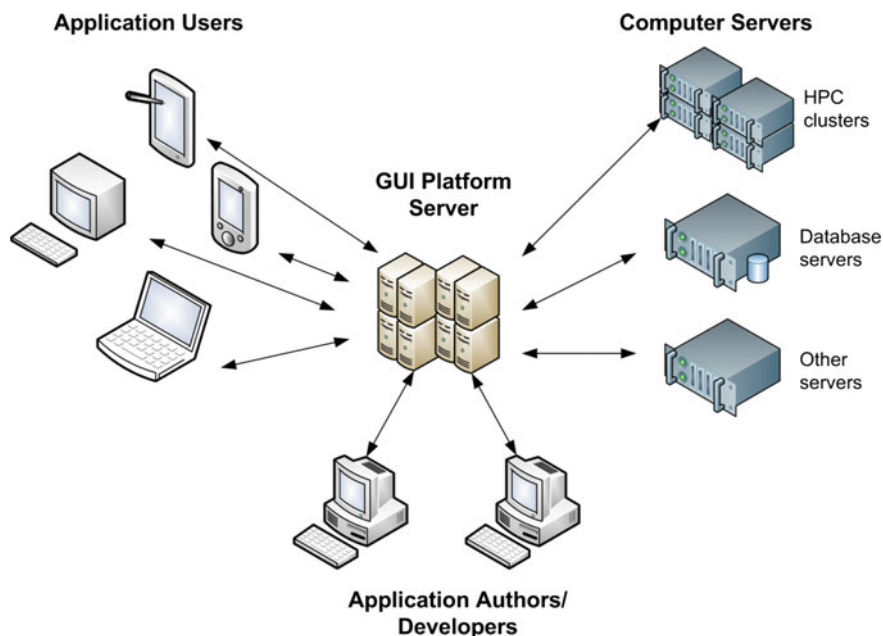


Fig. 1 Components of a typical web-based application

2.2 Overview of Web-Based CMC Development/Regulatory Statistics Applications

Most computational tools developed at Pfizer to support analytical method, product, and process development are written in script codes using R, SAS, MATLAB, JMP, Minitab, or MS Excel spread sheet templates. One example is drug product shelf life prediction. Long-term stability data are collected under various storage conditions, per ICH Q1A (2) and are evaluated per ICH Q1E (1) [2, 3]. The statistical analysis is coded in SAS and R to generate summary results and plots.

The commercial software packages, nevertheless are important tools for statisticians to carry out data analysis. However, individual usage of the software presents issues in portability, limited version control, and reproducibility. With support from Pfizer Information Technology group, statisticians have been able to turn the individual pieces of code into web-based applications. Figure 2 illustrates various web-based statistical applications developed by the CMC statisticians at Pfizer and the targeted areas throughout the life cycle of drug development and manufacturing. These applications are searchable and accessible to Global Pfizer colleagues.

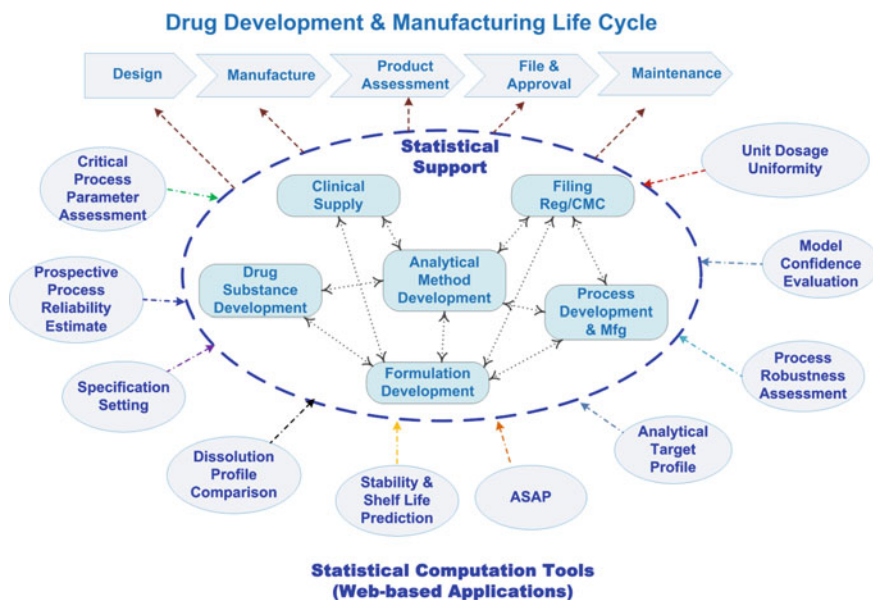


Fig. 2 Examples of statistical computational web-based applications throughout drug development and manufacturing life cycle

3 An Example Web-Based Statistical Computation Tool

Below, details are provided on the development and usage of one of the web-based applications listed in Fig. 2, Stability & Shelf Life Prediction.

For this application, assume that the stability data are collected from a registration stability program that follows ICH Q1A guidelines or a clinical stability program. Most stability programs have three registration batches per combination of strength, packaging configuration, and storage condition, whereas clinical stability program usually has only one batch. The online application of analyzing stability data is programmed in R, following ICH Q1E guidance for a specific combination of product, strength, package, and storage condition. The shelf life is determined by the decision criteria in the guidance. The clinical stability data is analyzed using a simple linear regression model, and the use period is determined, according to an internal criterion. For example, the use period of a clinical material is the shorter of the intersection of the 95% confidence interval and the specification limit or real stability time plus 12 months or longer if statistically supported. Therefore, the shelf life or clinical use period can be determined by a two-step procedure: model selection and projection of shelf life/use period.

3.1 Statistical Model Selection

For the statistical analysis of typical registration stability data, the following model selection procedure is performed based on the poolability of the data from the three batches. Assume $Y_b = y_{b1}, y_{b2}, \dots, y_{bT}$ are the stability data for an attribute at time period $t = 1, 2, \dots, T$ months for batch $b = 1, 2, \dots, B$ for a certain combination of strength, package type, and storage condition.

- (a) Fit a full model (the SSSI model—separate slopes and separate intercepts model):

$$y = \beta_0 + \beta_1 * time + \beta_{21} * batch + \beta_{12} * time * batch + \varepsilon \quad (1)$$

where the error ε is normally distributed with mean 0, and standard deviation σ . This model is referred to as the separate slopes and separate intercept model (SSSI), as it allows for different slopes and different intercepts for each batch.

Decision: If the p-value of the interaction of time and batch (time*batch) is <0.25 , STOP and use Eq. (1) for the shelf life projection; if the p-value of the interaction of time and batch (time*batch) is ≥ 0.25 , GOTO step (b).

- (b) Fit a reduced model (the CSSI model—common slope and separate intercepts model):

$$y = \beta_0 + \beta_1 * time + \beta_{21} * batch + \varepsilon \quad (2)$$

This model is referred to as a common slope and separate intercepts model (CSSI), as it permits the same slope estimate but different intercepts for all batches.

Decision: If the p-value of batch is <0.25 , STOP and use Eq. (2) for the shelf life projection; if the p-value of batch is ≥ 0.25 , GOTO step (c).

- (c) Fit a reduced model (the CSCI model—common slope and common intercept model):

$$y = \beta_0 + \beta_1 * time + \varepsilon \quad (3)$$

This model is referred to as a common slope and common intercept model (CSCI), since the same slope and intercept are used for all batches.

Decision: Eq. (3) is used for the shelf life projection.

The above described procedure for the statistical analysis of long-term registration stability data is summarized into a flow chart in Fig. 3. For typical one-batch clinical stability data, a simple linear regression model is used.

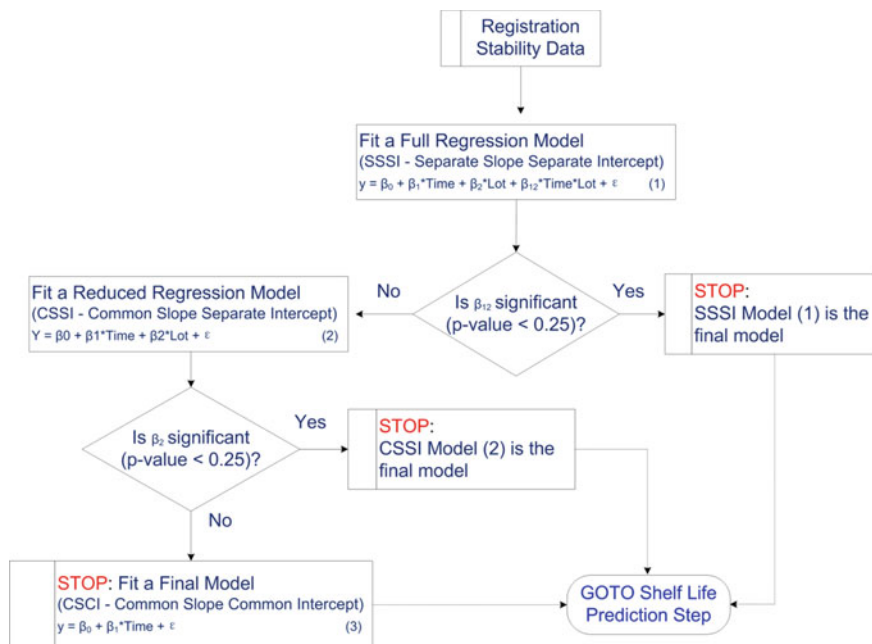


Fig. 3 Typical regression model selection per ICH Q1E stability data analysis

3.2 Shelf Life or Use-Period Projection

Once the regression model is determined, the 95% confidence interval (CI) can be calculated for any stability time point. The predicted shelf life/use period is determined as the shortest time point when the confidence limit intersects with the specification limit of the product. Notice that it is necessary to extrapolate the predictions and 95% CIs in order to determine the shelf life/use period beyond the maximum storage time of the stability data. Per ICH Q1E, the maximum extrapolation is two times of the maximum storage time (T_{\max}) when T_{\max} is < 12 months or an extrapolation of 12 months when T_{\max} is ≥ 12 months. Figure 4 illustrates how to establish the shelf life for an example data set. For this set of stability data, a separate slope and separate intercept model is selected and the shelf life is determined by the limiting lot (i.e. Lot 3). This shelf life limiting lot is determined, due to its fastest impurity A growth (largest slope) and thus its 95% CI intercepts with the specification limit of 1%, the earliest at 32.1 months. Therefore, 32.1 months (or 32 months) is the longest shelf life can be proposed. Practically, a shelf life of either 24 months or 30 months can be proposed for this product based on this set of data.

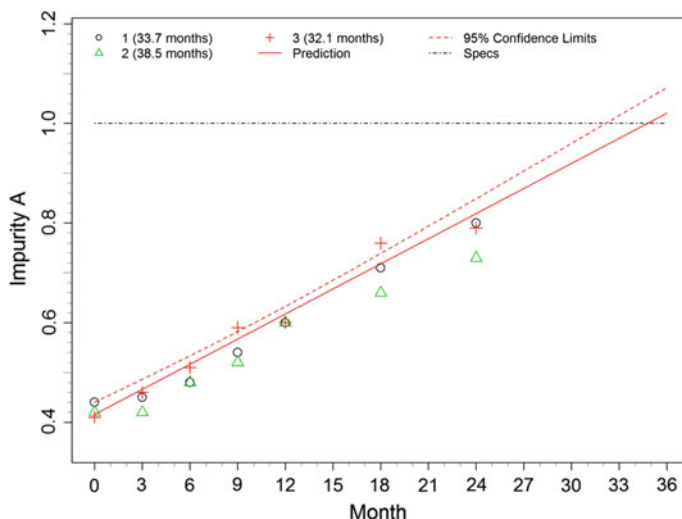


Fig. 4 Prediction of product shelf life based on regression model per ICH Q1E stability analysis: the predicted shelf life is the interception point (i.e., 32.1 months) of the upper 95% confidence limit with specification limit (i.e. the upper limit 1.0%)

3.3 The Internal Web-Based Online Application

Both long-term registration stability data and clinical stability data are collected routinely for all filed products. The repeated stability data analysis, including stability data plotting and drug product shelf life prediction, necessitated the development of a web-based application tool to standardize these statistical activities.

The web application for Registration and Clinical Stability Data Analysis and Shelflife/Use Period Prediction is programmed in R. A graphical user interface (GUI) is built to allow users to upload the relevant stability data to the program for analysis. The GUI of this application is displayed in Figs. 5 and 6 where the main interface contains links to various features, such as the user manual, example data sets in required formats, dialogues for uploading data, and choices of analyses.

Once stability data is uploaded and choices of statistical analyses and parameters are determined, the job is submitted and run in the background through the HPC computing cluster. As soon as the job is finished, users can view the results (including tables and graphical plots) through the web browser (e.g., Internet Explorer, Chrome). The application also provides the ability to download tables and graphs as well as consolidating the results in a .pdf formatted report. Figure 6a, b are snapshots of the output on a web browser.



Fig. 5 a Web-based application—registration and clinical stability data analysis and shelflife/use period prediction: GUI—main interface b Web-based application—registration and clinical stability data analysis and shelflife/use period prediction: GUI—further dialogues

(b)

Parameters Help Refine Analyses	
Pool by a Variable?	NA ▾
Confidence Level:	0.95
Type of Confidence Limit to Use:	Confidence ▾
P-value to Determine Poolability:	0.25
Minimum number of time points for a batch	3
SQRT Transformation of Stability Time?	None ▾
Unit of Stability Time: Month, Week, Day, Date (MM/DD/YYYY)	Month ▾
How many text lines per page for summary report?	30
Adjustment coefficients for y-axis:	0.9, 1.1
Color of the Regression Line:	Black ▾
Use Color for Plot Symbols	Black ▾
Symbol Size for Plot:	1.5
Size for Axis Label:	1.5
Tick Size for Axis:	1.3
Points Connected?	No ▾
Draw Specification on Plot?	Yes ▾
Plot similar to for filing?	No ▾

Miscellaneous Analyses and Plots	
Generate z-scores?	
Generate Z-Scores for Attributes?	No ▾
Adjust Initial Values at Time 0, Change Default Spec, or Change Expected Shelflife?	
Adjust initial values at time 0, change specs, and/or expected shelflife?	No ▾
Plot diagnostic plots and perform risk assessment for non-shelflife limiting CQAs	
Create diagnostic plots for linear regression models (not for Data Plot Only)?	No ▾
Perform risk assessment for non-shelflife limiting CQAs?	No ▾
Generate Summary of Stability Data and Slopes for Stability Batches	
Generate a summary table of stability data information	No ▾
Generate a summary table for slopes of stability batches:	No ▾

Fig. 5 (continued)

In summary, the implementation of the web-based statistical application of “registration and clinical stability data analysis and shelflife/use period prediction” is able to offer benefits and features such as,

- Align the statistical analyses of long term stability data
- Offer quick and convenient turnaround to analyze stability data, to generate shelf life plots and tables, and summary report
- Allow easy maintenance for feature updates due to the version controlled R program
- Run jobs in the background on HPC cluster or cloud computers.

4 Conclusions

The benefits and features of web-based statistical applications have been demonstrated through a selected program “registration and clinical stability data analysis and shelflife/use period prediction”. Statisticians and scientists supporting drug development and Reg CMC areas can offer their routine statistical activities with

(a)

1. Data Read-In

[Data_Analyzed.csv \(click to Open or Save\)](#)

1	PRODUCT	STRENGTH	STORAGE CONDITION	PACKAGE	LOT	METHOD	TIME	RESULT	SPECIFICATION	EXPIRY
1	Product B	ix mg	25C/60%RH	bottle	Lot 1	assay	0	100	95-105	24
2	Product B	ix mg	25C/60%RH	bottle	Lot 1	assay	3	98	95-105	24
3	Product B	ix mg	25C/60%RH	bottle	Lot 1	assay	6	98	95-105	24
4	Product B	ix mg	25C/60%RH	bottle	Lot 1	assay	9	100	95-105	24
5	Product B	ix mg	25C/60%RH	bottle	Lot 1	assay	12	99	95-105	24
6	Product B	ix mg	30C/65%RH	bottle	Lot 1	assay	0	100	95-105	24

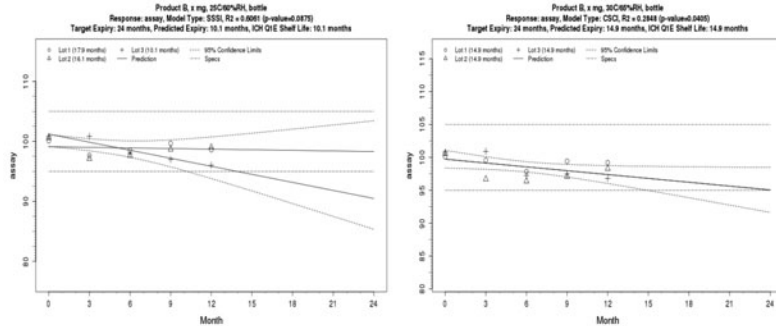
2. Summary of Shelflife Prediction

[ShelfLife_Estimate.csv \(click to Open or Save\)](#)

1	Product	Strength	Storage condition	Package	Response	Model	Lot	Pred Shelf Life	Q1E Shelf Life
1	Product B	ix mg	25C/60%RH	bottle	assay	Maximum Allowable	Expiry	10	10
2	Product B	ix mg	25C/60%RH	bottle	assay	SSSI	Lot 1	18	18
3	Product B	ix mg	25C/60%RH	bottle	assay	SSSI	Lot 2	16	16
4	Product B	ix mg	25C/60%RH	bottle	assay	SSSI	Lot 3	10	10

3. Plots of Shelf Life Prediction

[ShelfLifePlots.pdf \(click to Open or Save\)](#)



(b)

4. Summary of Stability Data Info

[Stability_Data_Information_Summary.csv \(click to Open or Save\)](#)

1	PRODUCT	STRENGTH	STORAGE CONDITION	PACKAGE	Lot	Time Point	Attribute	Specification
1	Product B	ix mg	25C/60%RH	bottle	Lot 1, Lot 2, Lot 3	0, 3, 6, 9, 12	assay	assay: 95-105
2	Product B	ix mg	25C/60%RH	bottle	Lot 4, Lot 5, Lot 6	0, 3, 6, 9, 12	-	assay: 95-105
3	Product B	ix mg	30C/65%RH	bottle	Lot 1, Lot 2, Lot 3	0, 3, 6, 9, 12	-	assay: 95-105
4	Product B	ix mg	30C/65%RH	bottle	Lot 4, Lot 5, Lot 6	0, 3, 6, 9, 12	-	assay: 95-105

5. Slopes of CQAs

[Stability_Data_Slope_Summary.csv \(click to Open or Save\)](#)

1	PRODUCT	STRENGTH	STORAGE CONDITION	PACKAGE	Response	Model	Lot	Pred Shelf Life	Intercept	Slope	P-Value
1	Product B	ix mg	25C/60%RH	bottle	assay	SSSI	Lot 1	18	99	-0.033	0.852
2	Product B	ix mg	25C/60%RH	bottle	assay	SSSI	Lot 2	16	99	-0.055	0.758
3	Product B	ix mg	25C/60%RH	bottle	assay	SSSI	Lot 3	10	101	-0.447	0.030
4	Product B	ix mg	30C/65%RH	bottle	assay	CSCI	Lot 1	15	100	-0.195	0.040

6. Summary Statistics of CQA Slopes

[Stability_Data_Slope_Summary_Statistics.csv \(click to Open or Save\)](#)

1	PRODUCT	STRENGTH	STORAGE CONDITION	PACKAGE	Response	N Batches	Mean Slope	Min Slope	Max Slope	sqrt(Total Variance)	sqrt(Between Batch Variance)	sqrt(Within Batch Variance)	Mean+3*StDev
1	Product B	ix mg	25C/60%RH	bottle	assay	3	-0.18	-0.45	-0.033	0.23	0.22	0.082	0.52
2	Product B	ix mg	25C/60%RH	bottle	assay	3	-0.15	-0.15	-0.150	0.00	0.00	0.000	-0.15
3	Product B	ix mg	30C/65%RH	bottle	assay	3	-0.19	-0.19	-0.195	0.00	0.00	0.000	-0.19
4	Product B	ix mg	30C/65%RH	bottle	assay	3	-0.10	-0.10	-0.105	0.00	0.00	0.000	-0.10

7. Report of Stability Data Analysis

[Stability_Data_Analysis_Report.pdf \(click to Open or Save\)](#)

Run by: Fasheng Li
Date run: 28-Aug-2017 15:16

Fig. 6 a Abbreviated Result—displayed in a browser of the web-based application—registration and clinical stability data analysis and shelflife/use period prediction: Data read-in, shelf life results and plots b Abbreviated Result—displayed in a browser of the web-based application—registration and clinical stability data analysis and shelflife/use period prediction: Summary of data, slopes, reports, etc.

increased consistency, improved efficiency, better alignment of statistical analyses, and easily retrievable results by deploying web-based statistical applications. These web-based statistical applications can standardize statistical approaches, centralize software pieces, validate and verify software pieces, and utilize high performance and cloud computer resources.

Acknowledgements The authors acknowledge Kimberly Vukovinsky, Robert J. Timpano, and colleagues in Pharmaceutical Science and Manufacturing Statistics group at Pfizer for their generous support of evaluating and commenting the web-based applications.

References

1. US Food and Drug Administration Department of Health and Human Services. 21 CFR 11: Electronic records; electronic signatures (2017)
2. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Q1E: Evaluation for Stability Data (2004)
3. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Q1A (R2): Stability Testing of New Drug Substances and Products (2003)