DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY, 17(1), 39-54 (1991)

SHORT-TERM STABILITY DETERMINATION USING SAS™

N. R. Bohidar

Philadelphia College of Pharmacy and Science
and
Villanova University

ABSTRACT

Appropriate statistical procedures are provided for determining the shelf-life of a drug product. Inverse regression analysis is proposed here. However, the formulas associated with the confidence limits are very cumbersome. Complete SAS computer program statements are presented so that they may be used in a computer to accomplish the task. It has been shown that there are situations in which the confidence limits may not exist. The conditions under which such situations arise are discussed and alternate procedures are proposed.

INTRODUCTION

The primary purpose of this paper is to provide the developmental pharmacist with adequate statistical and computational means to determine, for any time-interval, the

shelf-life of a drug-product which is currently in an early phase (phase I or II) of its clinical trial. During this period, stability characteristics of the drug product are routinely monitored at specified intervals to gain knowledge of its current stability status. For this purpose, routine analysis of the available stability data is an absolute necessity. Either for the intention of expediency or for lack of proper computational resources, the data analysis is generally carried out by resorting to informal graphing procedures or to subjective guess-estimate techniques which invariably produce inconclusive results. Since the conclusions derived from the statistical results have far-reaching consequences, it is advisable that appropriate data analysis techniques be used with a well-tested computer software easily implemented in any available computer system.

The paper presents (a) the various statistical formulations associated with the formal analysis of stability data, based on the suggested statistical approaches outlined in (1), (b) a complete set of SAS™ computer program statements, and (c) the salient results of the analysis of several stability studies obtained by using the computer program and their appropriate interpretation.

THEORY

Short-term thermal stability property of a pharmaceutical preparation of drug content has been shown to follow, in

general, a zero-order or a first-order kinetics (2). The primary purpose of a stability study is to experimentally determine the length of time that would be required to reach a Q (fixed) proportion of label claim using a linear least squares statistical model. Note that a first order kinetics model can be modified to a zero-order kinetics model by expressing the dependent variable in logarithmic scale.

Let X be the vector of time observation points, $X=(X_1X_2---X_n)$, which are fixed and are measured without error. Let Y be the vector of assay responses corresponding to each of the time points in X. The least squares estimate of the rate of degradation is $B^*=(X'X)^{-1}X'Y$ based on the model $Y=XB+E$, where E is the vector of experimental errors associated with assay responses assumed to follow a Gaussion distribution. To test the hypothesis of no-degradation over time, one uses,

$$F_{1,\ n-2}=[B^{*\prime}X'Y-R(bo)]/[(Y'Y-B^{*\prime}X'Y)/n-2]$$

where $R(bo)=(\Sigma Y)^2/n$. If the no-degradation hypothesis is rejected ($p < 0.05$), then one concludes that there is statistically significant loss of potency over the time period considered. The point estimate of the shelf-life is obtained by employing the inverse regression method (3), as follows

$$X^*=(Yo-A)/B^*$$

where Yo denotes the Q% of label claim.

Since the term shelf-life has been defined by the guidelines as the time required for a stability sample to reach a level of potency equivalent to 90% of label claim within the experimental variation encountered, one has to estimate not only the point estimate but also the 95% one-sided lower confidence limit (95% OSLC limit) for the mean time (1). To accomplish this and simplifying the notation, let the estimated regression line be denoted by the equation,

$$Y^*=A+BX$$

where, Y^* is the predicted value of the assay response for a given value of X and, A and B are respectively the intercept and slope of the line. The point estimate of the shelf-life is then (by solving for X from the above equation),

$$X^*=(Yo-A)/B$$

where Yo is the 90% of label claim.

To construct the 95% OSLC limit of X^*, one has to derive the variance formula for X^*, $V(X^*)$. Since X^* is a ratio and not a linear function of the regression parameters, the derivation becomes complicated. However, there are two ways to accomplish this.

(1) Fieller Approach (4),

The expression above may be rewritten as

$$X^*=(Yo-\bar{Y})/B + \bar{X} \qquad (\text{since } A=\bar{Y}-B\bar{X})$$

where $\bar{Y}$ and $\bar{X}$ are the means of the Y's and X's.

Upon rearrangement, we have,

$$Yo-\bar{Y}-B(X^*-\bar{X})=0$$

The variance formula for the left hand side of the above equation is equal to:

$$S^2((1/n) + (X^*-\bar{X})^2/[X^2])$$

where $[X^2]$ stands for $[(X_1-\bar{X})^2 + (X_2-\bar{X})^2 + \cdots + (X_n-\bar{X})^2]$

It is well known that, the following ratio has a t-distribution,

$$t=(Yo-\bar{Y}-B(X^*-\bar{X}))/S((1/n)+(X^*-\bar{X})^2/[X^2])^{1/2})$$

Squaring both sides of the equation, and rearranging, we have,

$$[Yo-\bar{Y}-B(X^*-\bar{X})]^2-t_a^2s^2((1/n) + (X^*-\bar{X})^2/[X^2])=0$$

This is a quadratic equation in X* and has two solutions. The smaller root is the 95% OSLC limit, with ta representing the tabular t-value for 0.10 level of significance obtained from a t-table (3).

The computational form of the 95% OSLC limit derived as a solution of the equation above is as follows:

$$X_L^* = X^* + \frac{G}{1-G}(X^*-\bar{X}) - \frac{t_{.10}S}{B(1-G)}\left(\frac{1-G}{n} + \frac{(X^*-\bar{X})^2}{[X^2]}\right)^{1/2}$$

where $G=t_{.10}^2S^2/B^2[X^2]$, L = Lower limit, G <<1.0,

S^2 = measure of assay variability (apart from the residual regression variation)(5),

and B = Absolute value of B (ignoring the sign)

This is the most appropriate statistical formula to be used for the calculation of the 95% OSLC limit.

(2) Statistical Differential Method (4):

Let, as usual, $X^*=(Yo-A)/B$.

According to the statistical differential method (sometimes called the propagation of error principle), the variance formula for X^*, $V(X^*)$, is derived from the following expression:

$$V(X^*) = \left(\frac{dX^*}{dA}\right)^2 V(A) + 2\,\frac{dX^*}{dA}\cdot\frac{dX^*}{dB}\,Cov(A,B) + \left(\frac{dX^*}{dB}\right)^2 V(B)$$

where d represents the partial derivative operator, V(A)= variance formula for A, V(B)=variance formula for B and Cov(A,B)= covariance between A and B.

Now by substituting the various derivatives, and the variances and covariance formulas, we have,

$$V(X^*)=(S^2/B^2)((1/n) + (X^*-\bar{X})^2/[X^2])$$

The 95% one-sided lower confidence limit is as follows,

$$X_L^*=X^*-(t_{.10}S/B)((1/n) + (X^*-\bar{X})^2/[X^2])^{1/2}$$

The formula is generally not used because it assumes the G-value to be zero. This is not necessarily true in laboratory experiments.

The 95% OSLC limit of Y^* for a given value of X is as follows:

$$Y_L^*=Y^*-t_{.10}S\sqrt{((1/n) + (Xo-\bar{X})^2/[X^2])}$$

where $Y^*=A+BXo$, Xo=a given value of X.

Composition of the G-value: To test the hypothesis of no degradation over time, the t-statistic is

$$t=(B)([X^2])^{1/2}/S.$$

For the slope to be significant, the t-statistic must exceed the tabular $t_{0.10}$ value. Symbolically we have,

$$(B^2[X^2]/S^2) > t^2_{0.01}$$

(squaring both sides of the inequality).

Now, consider the reciprocal of the above expression

$$S^2/(B^2[X^2]) < 1/t^2_{0.10}$$

Multiplying both sides of the inequality by t^2, we have

$$(t^2_{0.10}S^2/B^2[X^2]) < 1.0$$

(Note: The right hand side=G-value)

This inequality shows that the G-value should be less than 1.0.

The usual distributional assumptions are implicit in these derivations. The algorithm incorporated into the computer program for computing the confidence limit utilizes an iterative procedure constrained by a stated convergence criterion.

STABILITY COMPUTER PROGRAM

The program statements are presented in Addendum-A. They are written in the language of the SAS Statistical Computer Package distributed by the SAS Institute of North Carolina (6). The hardware used here is the VAX/VMS Version 4.5 terminal of the Digital Equipment Corporation. This information is

sufficient for the computer department of the user's facility to provide instructions of implementation and execution.

Note that, on the body of the program, five statements are numerically marked. These statements need attention prior to implementation. The modifications to these statements are as follows: (a) The changes needed for the statement in (1) should be provided by the computer department; (b) There is no change needed for the statement in (2) if the name of the data file (See Addendum-B, Table B-1) remains the same as "CAP466". However, should one prefer to use another name, that name must be typed in instead; (c) The statement in (3) pertains to the projected shelf-life in months and if the latest time period sampled was at month 12, then one may change the statement as,

```
Do I=12 To 60 BY 1;
```

implying that the confidence limit lies somewhere between 12 and 60 months; (d) If 5 mg was label claim, then insert into statement (4) a low and a high value which include 90% of label claim, such as,

```
Do Yo=4.2 To 4.7 BY 0.05;
```

(e) Insert the same statement above into the statement in (5). For other situations, pertinent values should be inserted in (3), (4) and (5) statements.

The names of the data and program files, in this case, were CAP466.DAT and STAB.SAS, respectively. While typing in the program, one should be careful about possible omissions of any

syntexes and spaces in the statements. Note that, for the convenience of the user, the two addendums (A and B) are presented at the end of the paper, independent of the text material.

Explanation of Computer Printout:

The program generates seven pages of statistical output. However, the most relevant statistical results are presented here in Addendum-B in Tables B-2, B-3 and B-4. The stability data used for the analysis are given in Table B-1. In this study the drug potency was 1 mg, and the interest here is to find the 95% OSLC limit of the shelf-life from the computer output. For this purpose, one enters into Table B-2 and locates 0.9 in column Yo and reads off from the next column on the left (XHAT LOW) the value of 38.62. This is the required 95% OSLC limit. For the point estimate, go to the next column on the left and read off 50.435 from the XHAT column. Now go to Table B-3 and locate 0.90 (or a closer value) in column entitled YLOW and read off 39 from the next column on the right. The results from Table B-2 and Table B-3 must be the same (except for rounding), even though they are calculated differently. Table B-4 provides the 95% one-sided lower confidence limit (YLOW) for each of the assay responses observed. It may be noted here that the 95% OSLC limit (YLOW) for the 12-month time point was 0.97830, which is significantly higher ($P < 0.05$) than 0.90, 90% label claim.

TABLE 1
SUMMARY OF SELECTED STATISTICS OF SEVEN STABILITY STUDIES

DP	N	T	B	SSS	G	P	L
DP-1	14	5	-0.00217	SIG	0.11	50.6	38.7
DP-2	17	4	-0.01275	NS	2.66	34.7	-
DP-3	15	3	-0.03890	NS	2.34	12.1	-
DP-4	16	5	-0.01536	SIG	0.14	43.9	33.8
DP-5	15	5	-0.01113	SIG	0.02	59.6	52.9
DP-6	15	5	-0.03780	SIG	0.63	38.5	23.8
DP-7	14	5	-0.01020	NS	0.84	135.6	-

Legend:

DP=Drug product, N=Total number of observations, T=Number of time points considered, B=Rate of degradation, SSS=Statistical Significance of Slope, G=Index of precision, P=Point estimate of shelf-life, L=95% one-sided lower confidence limit, "-" denotes that L is non-estimable (non-existent), SIG=Slope is significantly different from zero ($p<0.05$) and NS=Slope is not significantly ($p>0.05$) different from zero.

STABILITY STUDIES: RESULTS AND DISCUSSION

Table 1 presents a summary of the relevant statistical results of the seven stability studies considered for discussion. Each study was conducted based on an in-house rigid experimental protocol and on regulatory guidelines. The drug product was stored at 30°C, and random composite samples of 10 units each were analyzed independently at each of the five selected time periods, 0, 3, 6, 12 and 24 months for the drug products 1, 4, 5, 6 and 7. However, the length of the study periods for DP-2 and DP-3 were 12 and 6 months, respectively.

Table B-1 (Addendum-B) provides the total data set for the DP-1 study. Note the number of repeats at each time points.

A cursory examination of the contents of the table reveals that DP-2, DP-3 and DP-7 do not have a value for their 95% OSLC limits (column L). It is also seen that the slopes of the same drug products are not statistically significant (column SSS) and that the G-values associated with the three products either exceed or are close to the value of 1.0. This finding indicates that it is possible to obtain the 95% OSLC limit only if the rate of degradation is statistically significant and the G-value is far less than the value of 1.0. Indeed, as a general rule, if the G-value is in the neighborhood of 0.05 and the slope is significant ($p<0.05$), one has a high probability of estimating the 95% one-sided lower confidence limit.

The crucial quantity here is the G-value which is a non-linear function of the sample estimates of the slope, assay variability, and spread measure of the independent variable. The tabular t-value also plays an important role. It should be noted here that the size of the G-value entirely depends upon the relative magnitudes of the estimated values of the three quantities, noted above, obtained in the study being considered.

As far as the computer program is concerned, a non-existent 95% OSLC limit is indicated by a negative value, implying that it is not a meaningful estimate of the shelf-life (Table B-2). For the situation in which the 95% OSLC limit is

not estimable, the point estimate and the 95% OSLC limit of the mean response of the last time period would be of considerable value in understanding the current stability status of the drug product.

Note: The user may contact the author (DR. N. R. BOHIDAR, 1530 BRIDAL PATH, LANSDALE, PA, 19446) for any clarification with regard to the contents of Addendum-A and Addendum-B.

ACKNOWLEDGEMENT

Grateful acknowledgement is due to Dr. F. C. Masilungan for the valuable comments and discussions.

REFERENCES

1. Food and Drug Administration, "Guideline for Submitting Documentation for the Stability of Human Drugs and Biologics." Center for Drugs and Biologics, Rockville, MD (1987).
2. N.R. Bohidar and K.E. Peace, Pharmaceutical Formulation Development. Chapter IV. "Biopharmaceutical Statistics For Drug Development." Marcel Dekker, Inc., New York, NY (1988).
3. G.W. Snedecor and W.G. Cochran, "Statistical Methods", 6th Ed. Iowa State University Press, Ames, IA (1980).

4. J.J. Hubert, N.R. Bohidar and K.E. Peace, Assessment of Pharmacological Activity. Chapter III. "Biopharmaceutical Statistics For Drug Development." Marcel Dekker, Inc., New York, NY (1988).

5. D.J. Finney, "Statistical Methods in Biological Assay" Third Edition. Charles Griffin and Co., Ltd., London (1978).

6. SAS Institute Inc., "SAS User's Guide: Statistics," Version 5 Edition. SAS Institute Inc., Cary, NC (1985).

ADDENDUM-A
SAS COMPUTER PROGRAM

```
options ftype;  1
data one;
infile CAP466; 2
input x y;

PROC SORT DATA=ONE; BY X;

PROC MEANS NOPRINT DATA=ONE;
BY X ;
VAR Y;
OUTPUT OUT=SP N=N VAR=VARY ;

PROC MEANS NOPRINT DATA=ONE;
VAR X;
OUTPUT OUT=CSSQ CSS=SSXC;


DATA SP;
SET SP;
N=N-1;
VARY=VARY*N;

PROC MEANS DATA=SP;
VAR N VARY;
OUTPUT OUT=S2P SUM= SUMN SUMVARY;

DATA S2P;
SET S2P;

SPOOLED=SQRT(SUMVARY/SUMN);
KEEP SUMN SPOOLED;

PROC MEANS DATA=ONE NOPRINT;
VAR X;
OUTPUT OUT=XSTAT N=N MEAN=MEAN ;

PROC REG DATA=ONE NOPRINT outest=BETAS;
MODEL Y=X/ P ;
OUTPUT OUT=REG P=YHAT;
```

ADDENDUM-A (CONT.)

```
DATA TINVER (KEEP=SUMN T);
SET S2P;
T=TINV(.10,SUMN);
T=ABS(T);


DATA FINAL (KEEP=T SPOOLED B INTERCEP MEAN SSXC N MERGER);
MERGE S2P TINVER CSSQ XSTAT BETAS;
MERGER=1;
B=X;

DATA REG ;
SET REG;
MERGER=1;

DATA ALL;
MERGE REG FINAL; BY MERGER;

PROC PRINT;

DATA CI;
SET ALL;

YLOW=YHAT- (T*SPOOLED*SQRT(1/N + ((X-MEAN)**2/SSXC)) ) ;

PROC PRINT;
VAR Y YHAT YLOW X;
TITLE3 ' ONE SIDED CONFIDENCE INTERVALS (ALPHA=0.10) - ACTUAL DATA';

DATA XZEROCI;
SET FINAL;

DO I=30 TO 60 BY 3;  3

        XVAL=I;
        YHAT=INTERCEP + B*XVAL;
        YLOW=YHAT- (T*SPOOLED*SQRT(1/N + ((XVAL-MEAN)**2/SSXC)) ) ;
        OUTPUT;
END;

PROC PRINT;
VAR  YHAT YLOW XVAL;
TITLE2 'ONE SIDED CONFIDENCE INTERVALS (ALPHA=0.10) - X0 = 30 TO 60 BY 3';


DATA INV;
SET FINAL;


G=T**2*SPOOLED**2/B**2*SSXC;

IF G<1 THEN DO;

        DO YO=4.2 TO 4.7 BY .1 ; 4    /* Y0 VALUES. INSERT ENVALUES AND IN - */
                                      /* CREMENTS IN THE "DO LOOP"  */

                XHAT=(YO-INTERCEP)/B;
                XHATLOW= XHAT + G/(1-G)*(XHAT-MEAN) - (T*SPOOLED/B*(1-G)*(SQRT(
                         (1-G)/N + (XHAT-MEAN)**2/SSXC)));
                OUTPUT;
        END;
END;
```

ADDENDUM-A (CONT.)

```
ELSE DO;
        DO YO =4.2 TO 4.7 BY .1 ; 5   /* YØ VALUES. INSERT ENVALUES AND IN - */
                                      /* CREMENTSIN THE "DO LOOP"            */
                XHAT=(YO-INTERCEP)/B;
                XHATLOW=XHAT-(T*SQRT((SPOOLED**2/B**2)*(1/N + (XHAT-MEAN)**2
                        /SSXC)));
                OUTPUT;
        END;
END;

PROC PRINT;
VAR XHAT XHATLOW YO;
TITLE1 'XHAT AND LOWER 1Ø PERCENT C.I. ';

                                              /* OLD STAB1 */
proc reg data=one;
model y = x / p r cli clm;
output out=c p=pred l95=l95 u95=u95 r=resid cookd=cookd;

proc plot data=c;
plot y*x='A' pred*x='p' u95*x='u' l95*x='l'/
overlay vpos=32 hpos=8Ø;
plot resid*x /vref=Ø vpos=18 hpos=6Ø;
plot cookd*x /vref=Ø vpos=18 hpos=6Ø;
```

ADDENDUM-B

COMPUTER OUTPUT

TABLE B-1
DP-1 DATA

X	Y
0	1.01
0	1.01
0	1.02
0	1.00
0	1.00
0	1.01
1	1.01
1	1.01
3	1.02
3	1.02
6	0.98
6	0.97
12	0.99
12	0.99

TABLE B-2
XHAT AND LOWER 10 PERCENT C.I.

OBS	XHAT	XHATLOW	YO
1	96.413	73.17	0.8
2	50.435	38.62	0.9
3	4.457	3.37	1.0
4	-41.522	-52.69	1.1
5	-87.500	-110.09	1.2

ADDENDUM-B (CONT.)

TABLE B-3
SAS
ONE SIDED CONFIDENCE INTERVALS (ALPHA=0.10) - XO=30 TO 60 BY 3

OBS	YHAT	YLOW	XVAL
1	0.944444	0.929732	30
2	0.937920	0.921600	33
3	0.931395	0.913465	36
4	0.924870	0.905328	39
5	0.918345	0.897189	42
6	0.911820	0.889049	45
7	0.905296	0.880907	48
8	0.898771	0.872765	51
9	0.892246	0.864621	54
10	0.885721	0.856477	57
11	0.879196	0.848333	60

TABLE B-4
SAS
ONE SIDED CONFIDENCE INTERVALS (ALPHA=0.10) - ACTUAL DATA

OBS	Y	YHAT	YLOW	X
1	1.01	1.00969	1.00687	0
2	1.01	1.00969	1.00687	0
3	1.02	1.00969	1.00687	0
4	1.00	1.00969	1.00687	0
5	1.00	1.00969	1.00687	0
6	1.01	1.00969	1.00687	0
7	1.01	1.00752	1.00499	1
8	1.01	1.00752	1.00499	1
9	1.02	1.00317	1.00092	3
10	1.02	1.00317	1.00092	3
11	0.98	0.99664	0.99391	6
12	0.97	0.99664	0.99391	6
13	0.99	0.98359	0.97830	12
14	0.99	0.98359	0.97830	12