Drug Information Journal, Vol. 30, pp. 1105–1112, 1996
Printed in the USA.
0092-8615/96

STATISTICAL ASSESSMENT OF MEAN DIFFERENCES BETWEEN TWO DISSOLUTION DATA SETS*

YI TSONG, PHD, AND THOMAS HAMMERSTROM, PHD
Office of Epidemiology and Biometrics
PRADEEP SATHE, PHD, AND VINOD P. SHAH, PHD
Office of Pharmaceutical Sciences
Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland

When comparing the dissolution data of a postapproval change product and a reference approval product, the goal is to assess the similarity between the mean dissolution values at the observed sample time points. The decision on accepting or rejecting the hypothesis that the two batches have similar dissolution is based on the evidence regarding whether the difference in mean dissolution values between the test and reference products is no larger than the maximum expected difference between any two batches of the approval product. When dissolution value is measured at a single time point, the confidence interval of the true difference between the two batches is compared with the prespecified similarity limits. When dissolution values are measured at multiple time points, a multivariate statistical procedure for difference assessment can be a generalized form of the t-statistic procedure. The proposed procedure is a modification and generalization of the regular bioequivalence test concept. The application of the proposed multivariate analysis procedure is illustrated using an example.

Key Words: Dissolution test; Multivariate analysis of variance; Confidence region; Similarity test

INTRODUCTION

DISSOLUTION DATA OF the drug product are often used to assess the dosage form similarity when the drug manufacturer is making scale-up and postapproval changes, namely manufacturing site change, component and compositional changes, and equipment and process changes (1). Changes approved based on dissolution testing data are often accepted as adequate to assure dosage form similarity and hence *in vivo* performance. The similarity of the product with respect to dissolution means that the test (postapproval change) product has a dissolution performance no different than that expected from the reference (prechange) product except for the potential batch-to-batch or lot-to-lot variation.

Hence, in developing a statistical assessment of the difference between means of two dissolution data sets, the following need to be considered:

1. A well-defined similarity limit of the prechange product is established before com-

*The views expressed in this paper are professional opinions of the authors and may not necessarily represent the position of the U.S. Food and Drug Administration.

Reprint address: Yi Tsong, Division of Biometrics III, HFD-720, Office of Epidemiology and Biostatistics, Center for Drug Evaluation and Research, FDA, 5600 Fishers Lane, Rockville, MD 20857.

paring the dissolution data of the test and reference batches. The similarity limit is set either by the knowledge of the characteristics of the product or by the empirical experience on the batch-to-batch and the within-batch difference of the existing reference product. Similarity limits can be defined as global similarity or uniformly local similarity. Typically, a global similarity limit D_g is defined as a tolerable difference between the test batch and the reference batch over all time points. The limit D_g is often defined as a given percentage of dissolution, say, 10%. But with any measurement of difference, the similarity limit is defined by the measurement corresponding to D_g instead of D_g itself. On the other hand, uniform local similarity is defined as maximum tolerable difference D_t, at each sample time point. It is often defined as a given percentage of dissolution at time t, say, 12%. The similarity limit is then defined by the difference measurement at each time point and should reflect the similarity of the estimated batch dissolution profiles,

2. When the dissolution data of the reference product vary significantly from batch to batch, similarity limits become large and meaningless. In this case, dissolution data of test and reference batches may not be comparable for similarity,
3. The dissolution measurements of the test and the reference batches are made under identical conditions with an identical number of units. The reference batch used should be the most recently manufactured prechange product,
4. The conclusion of similarity is made with the consideration of the adequacy of the representativeness of the dissolution profiles (calculated with the sample tablets) of the test and reference batches, and
5. Dissolution values are sampled at the time points that properly and adequately represent the dissolution profile of the product.

This paper proposes a statistical assessment of the difference between the mean dissolutions of the test and reference products and proposes a statistical test for the hypothesis that the two batches are "globally similar."

SINGLE TIME POINT DISSOLUTION VALUE

For most immediate release drug products, the dissolution is rapid and the quality control dissolution requirement is only a single time point measurement. By considering the dissolution measurement as a variable, the difference between the mean dissolution values of the test and reference product can be easily measured and standardized with the regular t-statistic. There is an expected difference based on the empirical experience of the batch-to-batch variation of mean dissolution values of the reference product. Since the difference between mean dissolution values of the test and reference batches is an estimate based on the sampled tablets, a confidence interval of the true difference can be estimated and used when comparing with the expected difference.

For example, assume that the reference product dissolves rapidly and had one measurable dissolution value that is less than 100%. The average dissolution value based on 12 tablets of a standard batch is 78.5% and maximum batch-to-batch difference of the standard batches is no more than 7.5%. With the manufacturing site change, the manufacturer did dissolution testing of 12 tablets of the new batch after the site change (test batch) and 12 tablets of a reference batch manufactured recently at the old site. The dissolution values are given in the following example.

Example 1

Test batch			Reference batch		
76.5	79.6	82.1	85.3	83.5	82.5
78.9	81.5	77.6	81.4	84.4	79.0
79.8	83.4	80.2	80.5	78.5	85.3
75.8	81.2	80.5	81.5	79.9	83.2
Mean		79.76			82.08
Std		2.26			2.33

The difference, D, between test and reference

batch is 2.32%. The standard error of difference, stderr(D), is:

$$\sqrt{(2.26^2 + 2.33^2)/11} = 0.979$$

Then the 90% confidence interval of D would be the interval that contains all the differences $\mu_2 - \mu_1$ such that:

$$\left| [(\mu_2 - \mu_1) - D]/\text{stderr}(D) \right| \le t_{22,.95}. \quad (1)$$

Where $t_{22,.95}$ is the 95th percentile of t distribution with degrees of freedom = 22, that is, 2n – 2. The lower and upper limits of the interval L90 and U90 can be calculated by (2):

$$L90 = 2.32 - t_{22,.95} \cdot \text{stderr}(D) = 0.639$$
$$U90 = 2.32 + t_{22,.95} \cdot \text{stderr}(D) = 4.001.$$

Since both the lower and upper 90% confidence limits are within the inter-batch variation, ±7.5% (−7.5% ≤ 0.639% and 4.001% ≤ 7.5%), the product after change is accepted as a product with a dissolution value similar to that of the prechange product batch.

MULTIPLE TIME POINT DISSOLUTION

When the dissolution values are measured at multiple time points, dissolution measurement at each time point may be considered a variable. Variation in dissolution changes with time. These variables represent dissolution of one tablet at different time points, which are correlated. The use of the single time point approach then becomes erroneous, necessitating a new difference measurement.

A statistical distance often used to measure the difference between two multivariate means is the Mahalanobis distance (M-distance) (3):

$$D_M = \sqrt{[(x_2 - x_1)' S^{-1}_{pooled}(x_2 - x_1)]}, \quad (2)$$

where $S_{pooled} = (S_1 + S_2)/2$ is the sample variance-covariance matrix pooled across both batches, $x_1 = (x_{11}, x_{12}, \ldots, x_{1p})$ is the sample mean dissolution of the reference batch, and $x_{2.}$ is the sample mean dissolution of the test batch.

For the mean difference of two batches in dissolution measurements at multiple time points, the confidence region (CR), is defined as:

$$CR = K[(y - (x_2 - x_1))' S^{-1}_{pooled}(y - (x_2 - x_1)] \le F_{P,\,2n-P-1,\,.90}, \quad (3)$$

where $y = (\mu_{21} - \mu_{11}, \mu_{22} - \mu_{12}, \ldots, \mu_{2P} - \mu_{1P})$ is the difference within CR, $K = [(n^2)/(2n)]$

TABLE 1
Dissolution Data of a Reference Batch and a Test Batch

		% Dissolution							
Batch	Tablet	5-min	10-min	15-min	20-min	30-min	60-min	90-min	120-min
REF	1	42.06	59.91	65.58	71.81	77.77	85.67	93.14	94.23
REF	2	44.16	60.18	67.17	70.82	76.11	83.27	88.01	89.59
REF	3	45.63	55.77	65.56	70.50	76.92	83.91	86.83	90.12
REF	4	48.52	60.39	66.51	73.06	78.54	84.99	88.00	93.43
REF	5	50.49	61.82	69.06	72.85	78.99	86.86	89.70	90.79
REF	6	49.77	62.73	69.77	72.88	80.18	84.20	88.88	90.47
MEAN		46.77	60.21	67.28	71.97	78.05	84.82	89.09	91.43
TEST	1	19.99	36.70	47.77	55.08	65.69	81.37	92.39	97.10
TEST	2	22.08	39.29	49.46	56.79	67.22	82.42	89.93	95.62
TEST	3	21.93	38.54	47.76	55.14	65.25	83.49	90.19	95.62
TEST	4	22.44	39.46	49.72	58.67	69.21	84.93	94.12	95.51
TEST	5	25.67	42.35	52.68	59.71	71.51	86.61	93.80	96.70
TEST	6	26.37	41.34	51.01	57.75	69.44	85.90	94.45	98.07
MEAN		23.08	37.95	49.73	57.19	68.05	84.12	92.48	96.34

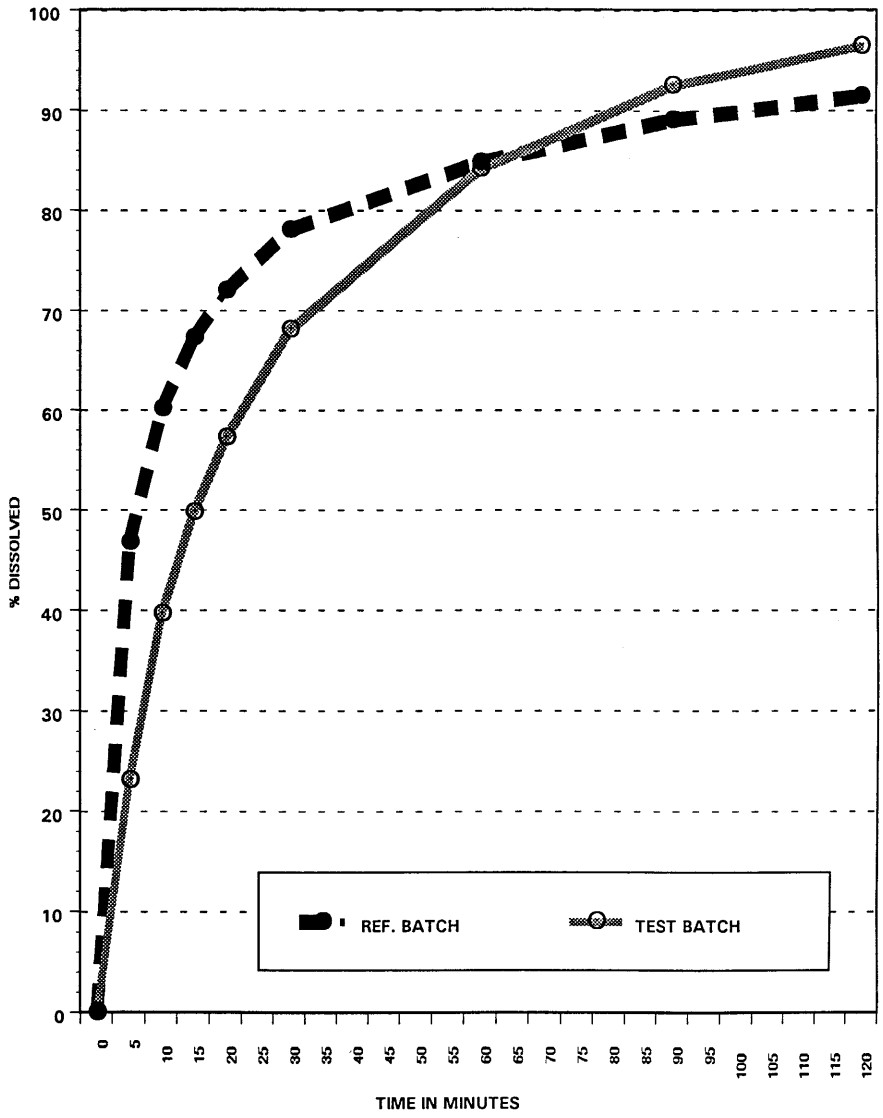

FIGURE 1. Mean dissolution of the test and reference batches.

$(2n - P - 1)/[(2n - 2)P]$, and $F_{P, 2n-P-1, .90}$ is the 90th percentile of F-distribution with degrees of freedoms P and $2n - P - 1$. The multivariate confidence region is compared with the overall similarity limits for overall similarity. It is to be noted that formula (1) is a special case of formula (3) when $P = 1$.

Formula (3) gives a p-variate 90% confidence region for the possible true differences. Let D_M^l be the lower 90% limit, and D_M^u be the upper 90% limit of the confidence interval of true M-distance D_M.. D_M^l and D_M^u represent, respectively, the two M-distances of the two values in CR that give the minimum and maximum M-distance to the original (the point of no dissolution difference). $D_M^l = 0$ if CR contains the origin. Otherwise, D_M^l and D_M^u can be calculated using Lagrange Multiplier method (4). The global similarity can be verified if the 90% confidence interval:

$$(D_M^l, D_M^u)$$

is imbedded in:

TABLE 2
Variance-Covariance Matrices of the Reference and Test Batches

		Time Point							
		5-min	10-min	15-min	20-min	30-min	60-min	90-min	120-min
REF	5-min	11.239	4.084	4.600	2.668	3.651	1.421	–2.790	–1.731
S1	10-min	4.084	5.749	3.430	2.105	2.424	1.117	1.822	0.389
	15-min	4.600	3.430	3.164	1.184	1.861	0.450	–0.295	–1.475
	20-min	2.668	2.105	1.184	1.260	1.427	0.881	0.615	0.867
	30-min	3.651	2.424	1.861	1.427	2.139	0.893	0.641	0.484
	60-min	1.421	1.117	0.449	0.881	0.893	1.706	1.673	1.211
	90-min	–2.790	1.822	–0.295	0.615	0.641	1.672	4.856	2.784
	120-min	–1.731	0.389	–1.475	0.867	0.484	1.211	2.784	3.655
TEST	5-min	5.962	4.685	4.090	3.340	4.827	4.542	2.862	1.271
S2	10-min	4.685	4.048	3.639	3.223	4.368	3.800	2.061	0.578
	15-min	4.090	3.639	3.627	3.284	4.468	3.341	2.355	0.681
	20-min	3.340	3.223	3.284	3.534	4.453	3.249	2.552	0.162
	30-min	4.827	4.368	4.468	4.453	5.870	4.339	3.590	0.701
	60-min	4.542	3.800	3.342	3.249	4.339	4.175	2.829	0.578
	90-min	2.862	2.061	2.355	2.552	3.590	2.829	4.014	1.188
	120-min	1.271	0.578	0.681	0.162	0.701	0.578	1.188	1.074

$$(-\sqrt{\{[D_g]'S^{-1}_{pooled}[D_g]\}},\quad \sqrt{\{[D_g]'S^{-1}_{pooled}[D_g]\}}),$$

where $[D_g]'$ is the $1 \times P$ vector with all entries equals to D_g, the difference specified as the global similar limit. The application of the procedure is illustrated in the following two examples (Table 1).

Data

The dissolution data and the corresponding mean dissolution of the test batch and the reference batch are given in Table 2 and Figure 1. Let S_1 and S_2 denote the sample covariance matrices of dissolutions at the eight time points for the test and reference batches, respectively, as shown in Table 3. The mean profiles of test and reference batches will be compared for similarity. Let D_g be the empirically defined limit, which is obtained from the standard batches of the reference product.

DATA1 data (comparing the 15- and 90-minute sample time points only): For comparing test and reference batches

$$P = 2,\ n = 6,\ S_{pooled} = (S_1 + S_2)/2,$$

$$(x_2 - x_1)' = (-17.54,\ 3.39)$$

$$K = [6 \cdot 6/(6 + 6)][(6 + 6 - 2 - 1)/$$

$$[(6 + 6 - 2)2] = 1.35$$

$$n_1 + n_2 - p - 1 = 9,\ F_{2,9,.90} = 3.01$$

$$CR = \{1.35 \begin{pmatrix} (\mu_{21} - \mu_{11}) + 17.54 \\ (\mu_{22} - \mu_{12}) - 3.39 \end{pmatrix}', \begin{pmatrix} 3.396 & 1.030 \\ 1.030 & 4.435 \end{pmatrix}^{-1}$$

$$((\mu_{21} - \mu_{11}) + 17.54,$$

$$(\mu_{22} - \mu_{12}) - 3.39) \le 3.01\}$$

where $(\mu_{21} - \mu_{11})$ and $(\mu_{22} - \mu_{12})$ are the possible value of difference at 15- and 90-minutes respectively. The confidence region is shown in Figure 2.

$$D_M = 10.44.$$

The points $((\mu_{21} - \mu_{11}), (\mu_{22} - \mu_{12}))$ on the boundary of CR with the minimum and the maximum M-distance from (0, 0), the origin, are (–15.03, 2.90) and (–20.05, 3.87), respectively. The 90% confidence interval of D (D^{l}_{M},

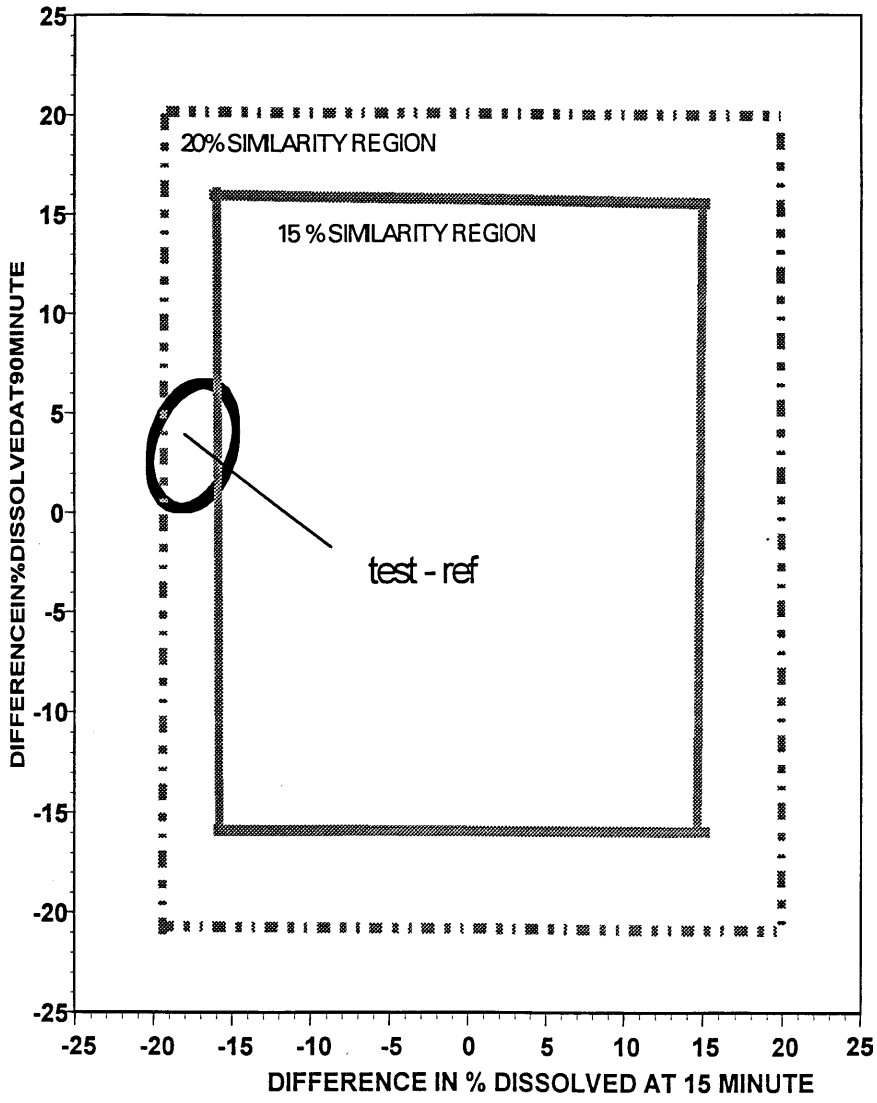

FIGURE 2. The 90% confidence region of difference in percentage dissolved between batches at 15- and 90-minutes.

D_M^u), is (8.95, 11.93). This is to be compared with:

$$RD = \sqrt{\{[15]'S_{pooled}^{-1}[15]\}} = 9.63.$$

Since 11.93 is larger than 9.63, it is concluded that the two batches are not globally similar.

DATA2 data (Comparing all eight time points):

$P = 8,\ n = 6,\ S_{pooled} = (S_1 + S_2)/2,$

$(x_2 - x_1)' = (-23.69, -20.52, -17.54, -14.80,$

$-10.02,\ 3.39,\ 5.00)$

$T^2 = 2104.46$

$K = [6 \cdot 6/(6 + 6)[(6 + 6 - 8 - 1)/$

$[(6 + 6 - 2)8] = .1125$

$D_M = 26.49$

$(D_M^l,\ D_M^u) = (19.65,\ 33.32)$

TABLE 3
Dissolution Data of Two Approved Standard Batches

		% Dissolution							
Batch	Tablet	5-min	10-min	15-min	20-min	30-min	60-min	90-min	120-min
STD 1	1	34.70	54.77	65.75	72.65	81.24	90.69	95.68	93.37
STD 1	2	39.43	59.40	67.24	75.23	83.52	93.52	96.14	96.46
STD 1	3	40.74	57.61	67.50	74.11	83.01	91.63	93.56	96.63
STD 1	4	40.95	57.53	69.05	77.18	85.02	92.71	95.56	96.38
STD 1	5	41.34	59.60	68.15	75.18	82.83	92.48	96.00	96.44
STD 1	6	41.93	57.06	67.33	76.44	83.39	93.60	97.91	97.61
MEAN		39.85	57.50	67.50	75.13	83.17	92.44	95.81	96.98
STD 2	1	44.35	59.22	64.90	68.13	73.10	77.32	82.38	85.51
STD 2	2	45.30	61.36	67.91	73.93	78.88	85.84	89.68	90.62
STD 2	3	47.35	59.77	64.79	68.72	72.25	77.06	81.00	84.09
STD 2	4	51.27	69.35	74.32	76.24	78.24	86.24	90.09	92.01
STD 2	5	52.32	64.86	69.20	74.43	79.81	86.12	89.96	91.85
STD 2	6	50.50	61.47	66.22	70.43	73.74	78.81	82.62	84.27
MEAN		48.52	62.67	67.89	71.98	76.00	81.90	85.96	88.06

Since D_M^u is greater than RD, it is concluded that the two batches are not globally similar.

DISCUSSION

When dissolution is measured at multiple time points, the statistic of Mahalanobis distance is used to assess the difference between the means of two sets of data with adjustment for difference in measurement variation at different time points and for the correlation among the measurements at multiple time points. Under the assumption that the dissolution measurement is multivariate normally distributed and that the two batches have the identical variance-covariance structure, the 90% confidence interval of the M-distance can be estimated. When the 90% confidence interval is imbedded within the similarity limits, the true difference between the test and reference batches is no more than the expected true difference between two batches of the same product as the reference batch. When the data are not normally distributed one may consider using a lognormal distribution assumption; similarity interpretation is on the means of the log transformed distribution values.

As shown in the formula, the confidence region is the ellipsoid restricted by the F percentile value. With a fixed number of tablets, more measurements (time points) may lead to a better representation of dissolution information. Since the second degrees of freedom of F is $2n - P - 1$, however, the maximum number of time points that can be used is $2n - 2$. The F percentile increases dramatically with the decrease of the second degrees of freedom. It leads to a large confidence region and a large confidence interval of M-distance. A model-dependent procedure may also be used in such cases.

When the data are neither normally distributed or lognormally distributed, one needs to consider the nonparametric or bootstrapping procedures. With within-batch dissolution variation being small, decisions made with point estimation are sometimes used. They should, however, be used cautiously and conservatively. For example, for 12 tablets in each batch and with a coefficient of variation of less than 0.10, instead of using 15% difference as similarity limit, one may use 10% instead.

Acknowledgment—The authors thank Drs. Roger Williams, Lawrence Lesko, and William Fairweather for their support and encouragement, and CDER's Immediate Release Dissolution Working Group for their suggestions and discussions.

REFERENCES

1. Guidance for Industry, Immediate Release Solid Oral Dosage Forms—Scale-Up and Post-Approval changes, Chemistry, Manufacturing and Controls. *In Vitro* Dissolution Testing, and *In Vivo* Bioequivalence Documentation. Rockville, MD; Center for Drug Evaluation and Research, FDA; November 1995.
2. Snedecor GW, Cochran WG. *Statistical Methods*. Seventh Edition. Ames, IO: The Iowa State University Press; 1980.
3. Johnson RA, Wichern DW. *Applied Multivariate Analysis*. Englewood Cliffs, NJ: Prentice-Hall, Inc.; 1989.
4. Freund JE. *Mathematical Statistics*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1962.