Drug Development and Industrial Pharmacy, 22(11), 1055-1061 (1996)

A Statistical Approach to Blend Uniformity Acceptance Criteria

J. Guentensberger, P. Lameiro, A. Nyhuis, B. O'Connell, and S. Tigner

Geneva Pharmaceuticals, Inc., 2555 W. Midway Blvd. Broomfield, Colorado 80020

ABSTRACT

Recent Food and Drug Administration (FDA) validation guidelines and comments indicate that applying finished product content uniformity specifications to blend testing is unacceptable. The scenario the FDA has presented is one in which disorder increases as the process progresses so that blend test specifications should be more restrictive (tighter) than finished product testing specifications. In other publications, it has been suggested that finished product assay limits be applied as a blend specification along with a lower relative standard deviation value than the current USP content uniformity limit (6.0%). This approach is questionable since assay results are applied to an aggregate finished product sample rather than individual doses. A method is presented in this paper for applying statistical tolerance limits (STLs) to blend data. This procedure provides a 95% confidence level that at least 90% of the values for the entire population are within the calculated limits. These statistical tolerance limits provide an acceptable criterion that is statistically tighter than the application of USP XXIII finished product content uniformity specifications. In addition, this method involves a decision process or multiple-level evaluation based on a statistical comparison of the variance and mean for the blend and finished product. In cases where the calculated STLs are unacceptable, the decision process allows for determining if the out-of-specification values from the first level of testing are due to a true blend failure or if the cause of the aberration is due to other phenomena, which could include sampling technique, thief design, and analytical testing problems.

INTRODUCTION

Current Food and Drug Administration (FDA) mandates for process validation (Dietrick, 1993) call for blend analyses as part of the validation campaign and application of blend specifications which are more stringent than specifications applied to the finished product. Considering blend testing in a wider sense, the goal of performing this type of evaluation during validation is to confirm the adequacy of the blending operation. In addition, the blend testing and acceptance criteria are intended to improve the likelihood that the finished product will meet the current USP Content Uniformity specifications.

Geneva Pharmaceuticals is a pharmaceutical company currently manufacturing over 250 products as solid dosage forms. Considering the wide variety of approved products and aggressive validation projects, it has become extremely important for the technical staff to develop a standard, scientifically sound, and usable program with which to evaluate blend data.

A multidepartmental team was formed to devise a standardized unit blend sampling and acceptance criteria. The team considered the following items in developing the system:

1. The criteria developed should be based in statistics to give a high level of assurance that the subsequent compression or encapsulation unit operations would yield product passing USP Content Uniformity specifications.
2. The system developed should be in general agreement with the current FDA guidelines on blend uniformity acceptance criteria.
3. The criteria developed should have a "two-tiered" format consisting of a first and second level of acceptance.
4. The blend sample size should approximate the size of one dosage unit and the number of blend samples taken from a given blend should not be less than ten.
5. Based on the Wolin decision, the system developed should not utilize the "outlier" concept (i.e., the elimination of widely varying individual results without scientific or valid exclusion criteria).

FDA publications suggest potential guidelines for blend uniformity criteria be set at 90–110% of label claim. Rather than utilizing a criteria that is based on and designed for an aggregate sample, a method using statistical tolerance limits (STLs) was chosen to evaluate blend uniformity. Statistical tolerance limits are calculated values that bound, with a specified confidence, a percentage of a population. The method of statistical tolerance limits meets all of the attributes listed above including the FDA requirement for a blend acceptance criteria more stringent than the USP finished product content uniformity specification.

METHODS

Unit blend data from validation and development lots from a wide range of products were analyzed to determine, based on history, the most appropriate specifications for statistical tolerance limits applied to unit blend sampling. Data collected from all lots were evaluated since implementing unit sampling at Geneva Pharmaceuticals, Inc. Prior to review, the data were sorted by relative standard deviation (RSD) and mean using a PC spreadsheet. Review of the sorted data indicated that statistical tolerance limits based on a 95% confidence that at least 90% of the blend lies between 85.0% and 115.0%, empirically correspond to an RSD of not more than 4.0%.

Statistical tolerance limits for blend testing are calculated which indicate that, with a 95% confidence level, 90% of the estimated population is within the lower and upper statistical tolerance limits. The statistical tolerance limits are constrained within 85.0% and 115.0% of blend uniformity, meaning that there is a 95% chance that a minimum of 90% of the blend is within 85.0% and 115.0% (see Fig. 1). Equations one and two are used to calculate the lower statistical tolerance limit (LSTL) and the upper statistical tolerance limit (USTL), respectively.

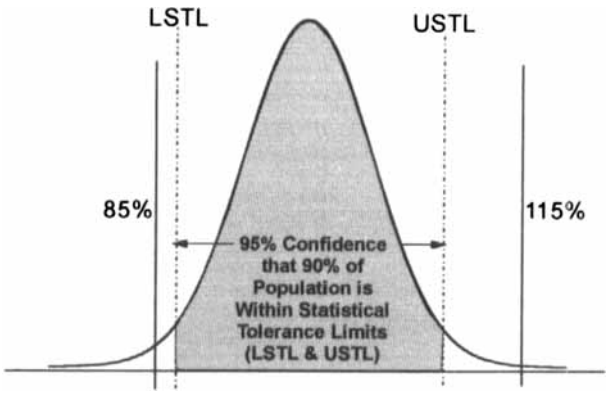

Figure 1. Illustration of statistical tolerance limits relative to USP specifications.

$$LSTL = x_b - ks_b \quad (1)$$

$$USTL = x_b - ks_b \quad (2)$$

where x_b = mean of the unit blend samples; s_b = standard deviation of the unit blend samples; k = constant: for 10 samples, 95% confidence, 90% of population, k = 2.839; see Appendix for more information about the origin of k.

When the unit blend statistical tolerance limits fall outside the 85.0–115.0% range, the blend data are compared to finished product content uniformity data of 30 units to determine how accurately the unit blend samples predict finished product variability. First, the variation of the unit blend data is statistically compared to the variation of the final dosage unit data using an F test [Eq. (3)].

$$F = \frac{s_b^2}{s_f^2} \quad (3)$$

where s_b = standard deviation of the unit blend samples; s_f = standard deviation of the finished product.

The sample statistic F is applied to the test interval:

$$0.3858 \le F \le 2.592 \quad (4)$$

If the variations are not statistically different at the 95% confidence level, the means are compared using a t statistic [Eq. (5)]. (The F test is performed first because homogeneity of variance is an assumption for this t test.)

$$t = \frac{\bar{x}_b - \bar{x}_f}{\sqrt{\frac{1}{n_b} - \frac{1}{n_f}}\sqrt{\frac{[(n_b - 1)s_b^2 + (n_f - 1)s_f^2]}{n_b + n_f - 2}}} \quad (5)$$

where x_b = mean of the unit blend samples; x_f = mean of the finished product content uniformity testing; s_b = standard deviation of the unit blend samples; s_f = standard deviation of the finished product content uniformity testing; n_b = integer number of blend samples; n_f = integer number of finished product content uniformity testing.

Solving this for a blend sample of 10 and a final unit sample of 30, yields:

$$t = \frac{\bar{x}_b - \bar{x}_f}{\sqrt{\frac{(9s_b^2 + 29s_f^2)}{285}}} \quad (6)$$

The sample statistic t is next applied to the test interval:

$$-2.024 \le t \le 2.024 \quad (7)$$

If this condition is false, it is concluded with a 95% confidence that the means of the populations are different. If this condition is true, the null hypothesis is accepted: the means of the populations are equal. If both the variation and the means of the unit blend samples and the finished product are statistically the same at a 95% confidence level, it is concluded that the blend results are indicative of finished product variability. If either the F test or t test fails, the statistical tolerance limits are calculated for the 30 finished product content uniformity values:

$$LSTL = x_f - ks_f \quad (8)$$

$$USTL = x_f + ks_f \quad (9)$$

where x_f = mean of the 30 finished product content uniformity values; s_f = standard deviation of the 30 finished product content uniformity values; k = constant: for 30 samples, 95% confidence, 90% of population, k = 2.140; see Appendix for more information about the origin of k.

If the statistical tolerance limits for 30 finished product content uniformity are constrained within 85.0% and 115.0%, there is a 95% chance that a minimum of 90% of the blend is within 85.0% and 115.0%.

COMPARISON TO USP

Attributes Portion of USP XXIII Content Uniformity

The use of statistical tolerance limits (STLs) described here for the control of blend uniformity is more stringent than both the attribute and variables portions of USP XXIII finished product content uniformity sampling plan.

The STLs are calculated so that 90% of the estimated population is within the lower and upper STLs, with a 95% confidence level. The STLs are constrained within 85% and 115% of blend uniformity. There is 95% chance that a *minimum* of 90% of the blend falls within 85% and 115%. See Fig. 1.

Using the USP sampling plan, consider the chance of finding one or more defective samples when the batch contains 10% defectives. This can be calculated precisely using a hypergeometric distribution, if the popu-

lation size is known; a more practical estimate uses the binomial distribution:

$$P(r) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \quad (10)$$

This function yields the probability P for finding exactly r defectives in a sample of n units from a large population containing proportion p (lower case) of defectives.

The USP finished product first-level attribute sampling plan will fail on one defective found outside 85% and 115%. Solving $P(r)$ for $r = 0$, $n = 10$ and $p = 0.1$, the chance of finding exactly zero defectives in a sample of $n = 10$ from a 10% defective lot is about 0.35, or 35%. The chance of finding one or more defectives is $1-P(0) = 0.65$, or 65%. Therefore, this sampling plan provides only a 65% confidence in detecting a 10% defective lot.

The second level of USP sampling plan provides a higher confidence by increasing the sample size, *but only after failing the first level criterion.* Therefore, it could be argued that USP is only as useful as its first, less stringent level. Nevertheless, the USP second-level effectiveness is also calculated here.

These are the two calculations for the second level:

$P(r)$ for $r = 0$, $n = 30$, $p = 0.1$,
which equals .0424 (11)

$P(r)$ for $r = 1$, $n = 30$, $p = 0.1$,
which equals .1413 (12)

Second-level USP fails when any one sample from $n = 30$ is outside 75% and 125%. The chance of detecting one or more defectives is $1-P(0) = 1-0.0424 = 0.9576$, or about 96%. This part of USP provides a 96% chance of detection, but only for units outside 75% and 125%.

Second-level USP also fails when two or more samples are found outside 85% and 115%. The chance of detecting two ore more defectives is 1-P(0)-P(1) 1-0.0424-0.1413 0.8163, or about 82%. Therefore, second-level USP provides only an 82% chance of finding two or more defectives outside 85% and 115%.

In contrast, Geneva's method *always* provides a 95% confidence that less than 10% of the population is outside 85% and 115%, whereas the USP attribute sampling plan detects samples outside this region at only 65% and 82% confidence levels.

Variables Portion of USP XXIII

Note that the previous discussion does not take into account the USP XXIII variables criteria, namely RSD < 6% for first level and RSD < 7.8%, for second level. Geneva's empirical data shows that RSD stays below 4% when applying these STL criteria. The reader may find similar though more thorough discussion of statistical power of the USP XXIII Content Uniformity Test in the listed references: Cowdery and Michaels (1980) and Nicolette (1981).

The variables portion of USP may be examined in the following way: First, consider that the USP variables criteria are based on the need for a 95% confidence of finding a lot standard deviation of $\sigma \leq 10\%$ according to Cowdery and Michaels (1980). Next, assuming a lot perfectly centered at 100% and a standard deviation of $\sigma = 10\%$, note that the Z scores corresponding to 115% and 125% are $Z = 1.5$ and $Z = 2.5$, respectively. This implies that the areas under the normal distribution beyond the 85–115% and 75–125% specifications are 13.4% and 1.24%, respectively.

Therefore, the RSD specifications of USP, applied to a perfectly centered lot, only ensure 86.6% of the lot within 85–115%, and 98.8% of the lot within 75–125%. Note that these defect levels are for worst-case lot standard deviation and best-case lot averages. When lot averages that are less than ideal, which is usually the case, the defect levels will be greater.

The proposed use of STLs provides a 95% confidence that 90% of the population is within 85–115%, in all situations, even noncentered lots. This also implies a lot standard deviation of $\sigma \leq 9.1\%$, for a centered lot. Therefore, a centered lot would be ensured of having 99.4% of the lot within 75–125%.

DISCUSSION

Based on the concept of statistical tolerance limits as discussed previously, the following criteria are proposed for application to unit blend sampling. This is a two-tiered system (analogous to USP content uniformity criteria) in which the first level of acceptance is based on an evaluation of blend data, and the second level of acceptance, when needed, is based on a comparison of blend data to finished product data. The second level of acceptance provides a mechanism for performing an investigation of a first-level failure (Fig. I).

Level 1: Calculated statistical tolerance limits for 10 unit blend samples are between 85.0% and 115.0%. If not, conduct an investigation to assess the sampling and testing process, then proceed to Level 2.

Level 2:

- The variation of the unit blend data from 10 unit samples *is not* statistically the same as the variation of the final dosage form data from 30 units

AND

- The calculated statistical tolerance limits of the final dosage form data from 30 units are between 85.0% and 115.0%.

OR

- The variation of the unit blend data from 10 unit samples *is* statistically the same as the variation of the final dosage form data from 30 units

AND

- The mean of the unit blend data from 10 unit samples *is not* statistically the same as the mean of the final dosage form data from 30 units

AND

- The calculated statistical tolerance limits of the final dosage data from 30 units are between 85.0% and 115.0%.

The statistical comparison between the variation of the unit blend results and the finished dosage form results is accomplished via the calculated F statistic, where $0.3858 \leq F \leq 2.592$ indicates insufficient evidence to conclude a difference between the variation of the blend and finished product populations.

The statistical comparison between the means of the unit blend results and the finished dosage form results is accomplished via the calculated t statistic, where $-2.024 \leq t \leq 2.024$ indicates insufficient evidence to conclude a difference between the means of the blend and finished product populations.

Thus, a product which both fails first-level blend criteria and for which the F or t statistic for the finished dosage form falls within the ranges noted above, would fail the criteria.

These principles were illustrated in a recent application of these STL-based blend acceptance criteria to an actual product. Unit blend results for Product A were as follows:

$\bar{x}_{n=10} = 94.8\%$
$s_{n=10} = 4.6\%$
Range = 85.1–98.8%
RSD = 4.9%
LSTL = 81%
USTL = 107.0%

This blend fails first-level criteria with a LSTL of 81.6%. Proceeding to second-level testing, 30 finished product units obtained from a composite sample were analyzed for content uniformity, with the following results:

$\bar{x}_{n=30} = 99.9\%$
$s_{n=30} = 1.7\%$
Range = 96.9–103.4%
RSD = 1.7%
LSTL = 96.3%
USTL = 103.5%

Comparing the variation of the blend and finished product via the calculated F statistic resulted in a value of 7.5. This value indicates that the variation of the blend and finished product populations are not the same, and the blend results are not indicative of product uniformity. Thus we conclude with a 95% confidence that at least 90% of the finished product falls within the finished product STLs of 96.3–103.5%. Since these finished products STLs are well within the limits of 85.0–115.0%, the product is deemed acceptable.

As further confirmation of the acceptability of this batch, extensive additional content uniformity testing was performed using samples obtained throughout the tableting run, with the following results:

$\bar{x}_{n=150} = 98.9\%$
$s_{n=150} = 1.8\%$
Range = 90.1–105.5%
RSD = 1.8%
$95.3\% < \text{LSTL} < 97.1\%$
$100.7\% < \text{USTL} < 102.5\%$

An important consideration for basing unit blend criteria on statistical tolerance limits is the provision of more stringent acceptance limits than required for finished product content uniformity by USP (all individuals 85.0–115.0%, with an RSD of not more than 6.0%). The foregoing discussion clearly illustrates the more stringent nature of the STL-based criteria with respect to USP finished product content uniformity criteria.

CONCLUSION

The advent of unit blend sampling in the pharmaceutical industry has precipitated a need for a rational ac-

ceptance criteria for the analysis of unit blend samples. The criteria developed satisfies, at a minimum, the following requirements:

- The criteria should be based on a statistical evaluation that characterizes the entire blend.
- The criteria should be more stringent than finished product testing, for example, USP content uniformity testing.
- The criteria should have provision for a second level of acceptance.

The system described herein satisfies all of the listed objectives. STLs are calculated using the mean and standard deviation of the data set as well as a constant which reflects a given confidence of detecting a given level of defects. Since the constant may be obtained directly from a table, and since mean and standard deviation are statistical parameters that are widely known and understood, the calculation of STLs is straightforward. The application of STLs outlined in this article was shown both theoretically and empirically to provide statistically tighter control than provided by first- and second-level content uniformity specifications listed in the USP XXIII.

Finally, this procedure contains a second level of acceptance by which the unit blend data are compared statistically to 30-unit finished product content uniformity data. STLs may then be applied to the finished product content uniformity data to estimate the uniformity of the finished product population. This second level of acceptance also provides a means by which a bias in the unit blend sampling process may be identified, addressed, and overcome without an unduly burdensome investigation process.

APPENDIX: BASIS FOR STATISTICAL TOLERANCE LIMIT FACTORS

Given a normal population characterized by $\bar{x}$ and s^2, a sample of n units is taken from the population to determine the interval $\bar{x} \pm ks$ which contains at least the proportion P of the population with the probability g.

The work of Wald and Wolfowitz (1946) and Odeh and Owen (1980) take the following approach to the derivation of k. The equation

$$\gamma = \frac{2\sqrt{n}}{\sqrt{2\pi}} \int_0^\infty \Pr\left\{\chi_f^2 \geq \frac{fr^2}{k^2}\right\} e^{\frac{n\bar{x}^2}{2}} d\bar{x} \tag{13}$$

is iteratively evaluated using numerical methods for given values of n, P, and k, until the desired value of g is reached. In this equation f denotes the degrees of freedom for the c^2 function and r is the unique root of the equation

$$G(\bar{x} + r) - G(\bar{x} - r) = P \tag{14}$$

where $G(x)$ is the cumulative normal up to x.

Where P is the proportion of the distribution within the tolerance limits, γ is the confidence, and χ_γ^2 is defined by:

$$P(\chi^2 > \chi_\gamma^2) = \gamma \tag{15}$$

Other, slightly different, but apparently more common values of k are given by Eisenhart, Hastay, and Wallis (1947). In this case, Newton's method is applied to the following for the computation of r:

$$g(r) = \frac{1}{\sqrt{2\pi}} \int_{\frac{1}{\sqrt{N}}r}^{\frac{1}{\sqrt{N}}+r} e^{\frac{\bar{x}^2}{2}} d\bar{x} - P = 0 \tag{16}$$

Finally, k is evaluated with:

$$k = r\sqrt{\frac{n}{\chi_\gamma^2}} \tag{17}$$

REFERENCES

Berman, J., and Planchard, J. A. (1995), Blend uniformity and unit dose sampling, Drug Dev. Ind. Pharm., 21(11), 1257–1283.

Bowker, A. H. (1946), Computation of factors for tolerance limits on a normal distribution when the sample is large, Ann. Math. Statistics, 17, 238.

Bowker, A. H. (1947), In *Selected Techniques of Statistical Analysis for Scientific and Industrial Research and Production and Management Engineering* (Eisenhart, C., Hastay, M. W., and Wallis, W. A., eds.), McGraw-Hill, New York.

Carstensen, J. T., and Rhodes, C. T. (1993), Sampling in blending validation, Drug Dev. Ind. Pharm., 19(20), 2699–2708.

Chow, Shein-Chung, and Liu, Jen-Pei (1995), *Statistical Design and Analysis in Pharmaceutical Science—Validation, Process Controls, and Stability*, Marcel Dekker, New York.

Cowdery, S., and Michaels, T. (1980). An evaluation of the proposed dose uniformity acceptance sampling plan for tablets and capsules, *Pharmacopeial Forum*, The United States Pharmacopeial Convention, Inc., pp. 614–618.

Danziger, L., and Davis, S. A. (1964), Tables of distribution-free tolerance limits, Ann. Math. Statistics, 35, 1361–1365.

Deming, W. E. (1976), On the use of judgment samples, Rep. Statistical Appl. Jpn. Union Sci. Eng., 23, 25–31.

Dietrick, J. M. (1993), Food and drug administration recommendations on validation issues, Regulatory Affairs, 5, 363–366.

Dudewicz, E. J. (1988), Basic statistical methods, in *Juran's Quality Control Handbook*, 4th ed., (Juran, J. M., and Gryna, F. M., eds.), McGraw-Hill, New York.

Eisenhart, C., Hastay, M. W., and Wallis, W. A. (1947), *Techniques of Statistical Analysis*, McGraw-Hill, New York.

Ellison, B. E. (1946), On two-sided tolerance intervals for a normal distribution, Ann. Math. Statistics, 35, 762–772.

Gardiner, D. A., and Hull, N. C. (1966), An approximation of two-sided tolerance limits for normal populations, Technometrics, 8, 115–122.

Grant, E. L., and Leavenworth, R. S. (1988), *Statistical Quality Control,* 6th ed., McGraw-Hill, New York.

Hahn, G. (1970), Statistical intervals for a normal population, Part II: Formulas, assumptions, some derivations, J. Quality Technol., 2, 195–206.

Hahn, G. J., and Meeker, W. Q. (1991), *Statistical Intervals—A Guide for Practitioners*, John Wiley & Sons, New York.

Howe, W. G. (1969), Two-sided tolerance limits for normal populations—Some improvements, J. Amer. Statist. Assoc., 64, 610–620.

Lieberman, A. (1957), *Tables for the Determination Two-Sided Tolerance Limits for the Normal Distribution*, Report No. 373-17(55), Bureau of Ships, Navy Department, Washington, DC.

Lord, A. G. (1987), Content uniformity in tablets: A literature review, Pharm. Technol., September, 64–70.

Mouradian, G. (1966), Tolerance limits for assemblies and engineering relationships, *ASQC Technical Conference Trans.*, Milwaukee, pp. 598–606.

Nicolette, B. P. (1981), Statistical implications of the USP XXIII Content Uniformity Test, *Pharmacopeial Forum*, The United States Pharmacopeial Convention, Inc., pp. 878–887.

Odeh, R. E. (1978), Tables of two-sided tolerance factors for a normal distribution, Commun. Statist. Simul. Comp. B, 7, 183–201.

Odeh, R. E., and Owen, D. B. (1980). *Tables for Normal Tolerance Limits, Sampling Plans, and Screening*, Marcel Dekker, New York.

Patel, J. K. (1986), Tolerance intervals—a review, Commun. Statist. Theory Methods, 15, 2719–2762.

Tietjen, G. L., and Johnson, M. E. (1979), Exact statistical tolerance limits for sample variances, Technometrics, 21, 107–110.

Wald, A., and Wolfowitz, J. (1946), Tolerance limits for a normal distribution, Ann. Math. Statistics, 17, 208–215.

Weissberg, A., and Beatty, G. H. (1960), Tables of tolerance limit factors for normal distributions, Technometrics, 2, 483–500; Errata 3, 576–577.