Drug Development and Industrial Pharmacy, 25(10), 1083–1091 (1999)

RESEARCH PAPER

Calculating the Risk of Batch Failure in the Manufacture of Drug Products

John Stafford

Novartis Pharma AG, CH-4002 Basel, Switzerland

ABSTRACT

A statistical tool has been developed to estimate the risk of batch failure in commercial production from the results of the development scale-up campaigns. This tool serves to indicate the robustness of the manufacturing procedure and its ability to meet the specifications set. Other uses include assistance in the resetting of specifications in the light of actual scale-up batch manufacturing experience and assessing the impact of changes in specifications proposed by the registration authorities. A template is given that simplifies the calculations required.

INTRODUCTION

Batches of pharmaceutical products are rejected because of equipment failure, human error, or inability to meet specifications. Equipment qualification and maintenance schedules reduce the possibility of equipment failure, and training and standard operating procedures guard against human error.

The failure to meet specifications is product specific and is governed by the robustness of the product in terms of the formulation, manufacturing procedure, sampling, and analytical methodology. It is of obvious value to have some means of estimating this robustness when a product in development becomes a candidate for commercialization, with registration and transfer to production.

One method often used to estimate the robustness is the process capability index C_pK, defined as:

$$C_pK = \frac{\text{Process Mean} - \text{Specification Limit}}{3 \times \text{Process Standard Deviation}} \quad (1)$$

where decreasing values below 1.0 indicate decreasing degrees of process acceptability, and values above 1.0 signify increasing degrees of process robustness.

The risk of batch failure represents another method of calculating robustness. It uses the same three parameters of process mean, process standard deviation, and specification limit to calculate the probability of batch rejection or risk of batch failure. In addition, a reliability index is estimated to indicate the value of the result.

The risk of batch failure is preferred over the process capability index because the risk of batch failure has an unambiguous meaning that is readily understood at all levels of the organization, both technical and nontechnical. In addition, the risk of batch failure has immediate

impact if discussions with management are necessary to obtain resources for robustness improvement.

STATISTICAL BACKGROUND

The incidence of occurrence of out-of-limit results can be calculated on the basis of a Gaussian distribution on the assumption that results obtained are Gaussian distributed. Present experience with such results indicates they are indeed Gaussian distributed. For example, the content uniformity data (200 data points) on a batch of capsules from a recently registered Novartis low-dosage product very closely approximates a Gaussian distribution (see Fig. 1).

The normalized differential Gaussian distribution function for any individual value of i is given for a mean value μ and standard deviation value σ by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{2}$$

where z, the normal standard deviate, is given by

$$z = \frac{\mu - i}{\sigma} \quad \text{or} \quad z = \frac{i - \mu}{\sigma} \tag{3}$$

(see Fig. 1a).

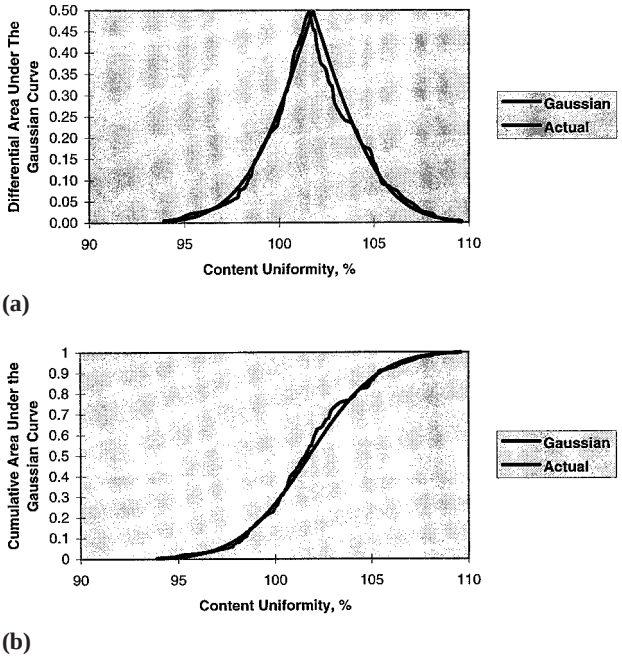


Figure 1. Low-dosage hard gelatin capsule product.

The area under the curve above or below any value of z is given by

$$Area = \frac{1}{\sqrt{2\pi}} \int_z^{\infty} e^{-z^2/2} \cdot dx \quad \text{or}$$

$$Area = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} \cdot dx \tag{4}$$

The frequency of failing to meet specification limits can then be estimated if z is assigned the values of these limits:

$$z = \frac{Limit_{Upper} - \mu}{\sigma} \quad \text{or} \quad z = \frac{\mu - Limit_{Lower}}{\sigma} \tag{5}$$

(see Fig. 2).

During scale-up validation, the number of data points is not large, so μ and σ, the means and standard deviations for an infinite number of data points, are not definitely known.

Nevertheless, the probability of occurrence of out-of-limit results can still be estimated on the above basis if the effect of using a smaller number n of data points is taken into account. This can be done by using the mean and standard deviation of the smaller number of data points in place of μ and σ in the above equations. The error involved in making this substitution can then be estimated on the basis of the Student distribution (see Fig. 3).

This error is calculated as a reliability index. The reliability index is defined by

$$\text{Reliability Index} = \frac{[\text{probability for the value } n - 1]}{[\text{probability for the value } n = \infty]} \tag{6}$$

where n is the actual number of data points.

Batch Failure Risk and Reliability Index Acceptance Criteria

The acceptable level of the risk of batch failure is given for the various scale-up phases as follows: For pilot batch phase, the risk of batch failure is less than 5%; for production scale-up phase (process validation and confirmation), the risk of batch failure is less than 1%. The higher risk of batch failure of 5% for the pilot batch phase is accepted because of the small number of data points available at this stage. However, if the risk of batch failure exceeds 1%, then the relevant experts should consider

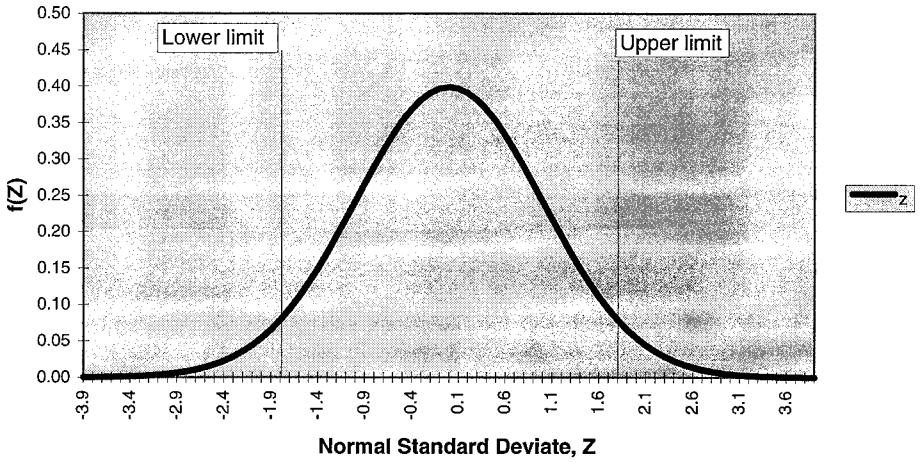


Figure 2. Gaussian distribution of individual results obtained.

whether the cause of the higher value is analytical or process related and initiate appropriate corrective measures.

Lower levels of batch failure risk are required for later stages of scale-up, particularly since the calculation of the risk of batch failure will be made on pooled data from all the scale-up phases (permissible when no significant process or analytical procedure change has occurred).

A consequence of the mathematical treatment is that the reliability index increases rapidly with decreasing batch failure risk. In assessing this index, the batch failure risk is first checked. If this risk is very low (much less than 1), then the reliability index is ignored. In the region of borderline batch failure risk (0.5% to 5%), an acceptable reliability index would be in the region of about 3%.

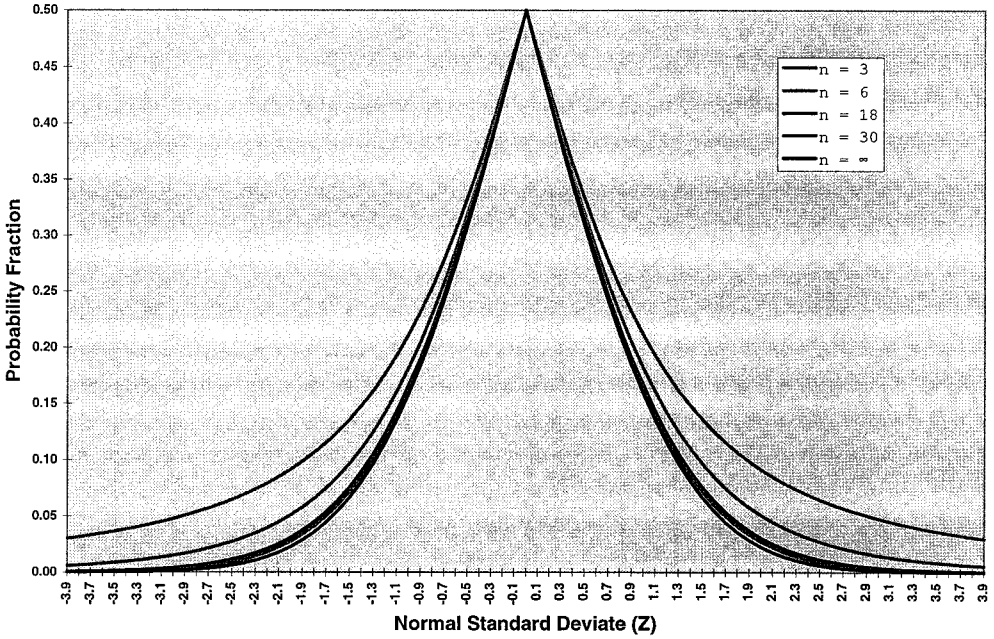


Figure 3. Differential Student distribution.

Method Description

The risk of batch failure method is primarily designed to be used prospectively in that it forecasts the risk of not meeting specifications. This is done on the basis of the mean and standard deviation of a limited amount of data generated during the scale-up phase of development of new products and redevelopment of already marketed products. It is thus an evaluation tool to help establish whether the manufacturing, analytical, and sampling procedures for a development product can meet the set specifications and thus are sufficiently mature for registration and transfer into production. If not, then the tool identifies those specification limits that are endangered.

The method is different from and not related to the consumer risk/producer risk calculation, which estimates, on the basis of the sampling plan and a defined probability, the risk to the consumer of a batch being released when it should have been rejected or to the producer of a batch being rejected when it should have been released.

Uses for the Method

The method of determining the risk of batch failure provides the following:

- an unambiguous assessment of the robustness of the manufacturing procedure (and includes the variability associated with the analytical method and sampling procedure)
- an additional tool in setting specifications for the scale-up phase
- an additional tool for the justification of specifications during development, scale-up, and registration
- a method for estimating the impact of any changes proposed by the registration authorities on the basis of submitted data

Examples of some of these uses are given below.

USE OF THE METHOD: CASE STUDIES ON ASSAY

The European Pharmacopeial assay limit of 95–105% drug content is one of the toughest specifications in the pharmaceutical manufacture of solid forms.

Medium-Dosage Product

A recently introduced product contains a very hygroscopic drug substance. To avoid problems with weigh-up, the drug substance is predispensed to meet batch quantity requirements when bagging into aluminum foil bags and sealing at the point of drug substance manufacture. Thus, the contents of each bag are used in total on drug product manufacture.

To ensure full potency, the predispensed quantity of drug substance contained a fixed overage (initially 3%) for loss on drying (LOD) for a tight LOD drug substance specification (of 2.5% to 3.5%) as determined by thermogravimetry.

The risk of batch failure for the pilot batch phase is estimated as 14.6%, primarily on the basis of a very high assay standard deviation of 2.9%. The reliability index is reasonable at 1.8% (see Table 1).

The rate of batch failure for the production scale-up batches is about 1%, mainly because of a much reduced assay standard deviation of about 1%, even though the mean assay is still disturbingly high at 102.5%. Here, the reliability index, at 7.1%, is quite high for this risk of batch failure.

When the data for the pilot and production scale-up phases are pooled, the expectancy of rejection is 7.9%. The reliability index is a satisfactory 1.47 for the 6 pooled data points.

Thus, the data available at this point in time predicts about 8% of the batches of this product will fail in serial production because of assay; this prediction could be made before a single batch failure has occurred.

The first 32 batches of this product were produced with a fixed LOD compensation of 3.0%. Of these 32 batches, 6.25% were actually rejected because of assay (see Table 2). The precision of the 7.9% failure frequency on the basis of the 6 pilot and scale-up batches (see above) is very encouraging. It also compares reasonably well with a failure frequency of 12.0% calculated on the basis of the mean and standard deviation of the 32 batches.

The assay problem with this medium-dosage product was removed by a series of measures, the most important of which was the change in method for determining the LOD for the fixed LOD compensation (change from thermogravimetric analysis [fixed LOD compensation of 3% LOD] to the Karl Fischer method [fixed LOD compensation for 2% LOD]). The results can be seen in Table 3. The calculated batch failure frequency dropped from 7.45% to 0.01%, and no batch failures were observed with the 39 batches listed (or in any batches of this prod-

Table 1

Risk of Batch Failure Calculation for a Medium-Dosage Product: Pilot Batch and Scale-Up Phase

	Pilot Batches	Production Scale-Up Batches	Pilot and Production Scale-Up Batches
Assay	Batch 1, 98.9	Batch 4, 103.6	Batch 1, 103.6
	Batch 2, 101.94	Batch 5, 102.58	Batch 2, 102.58
	Batch 3, 104.68	Batch 6, 101.42	Batch 3, 101.42
			Batch 4, 98.9
			Batch 5, 101.94
			Batch 6, 104.68
Mean (%)	101.84	102.53	102.19
Standard deviation (%)	2.89	1.09	1.99
Expected rejects (%)	14.62	1.19	7.90
Reliability index	1.81	7.27	1.47
Actual rejects (%)	0	0	0

uct produced since then). This clearly demonstrates the success of the corrective measures taken.

Thus, the method of calculating the rate of batch failure can also serve to show that corrective measures to eliminate a problem have been successful, that is, it serves as an all-clear signal.

Assay: ''What If'' Calculations, Assay Standard Deviation

''What if'' calculations allow the consequences to be explored in the sense of what would happen to the batch failure rate if one of the variables is changed; the variable changed might be the standard deviation of individual values, the mean value, the limits, and so on. Figure 4 shows how the rate of batch failure varies for an assumed assay mean of 100% and varying standard deviation, as well as for an assumed standard deviation of 1.9 and a mean varying from 100%.

It is seen that, with a mean of 100%, the expected rate of batch failure begins to rise above a threshold value of 1% when the standard deviation exceeds about 2.0%. In addition, for an assumed standard deviation of 1.9%, the expected rate of batch failure increases rapidly above 1% when the assay mean falls below 100%. This mirrors the batch rejection behavior for a fixed standard deviation of

Table 2

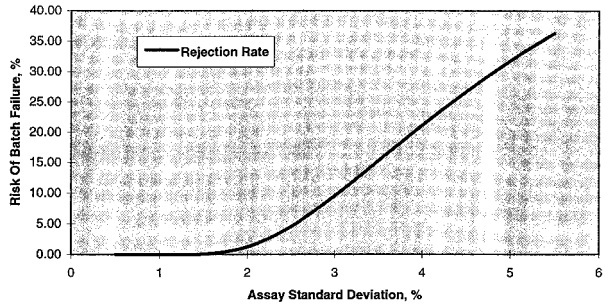
Rate of Batch Failure for Assay of a Medium-Dosage Drug Product: Commercial Production with Fixed Loss on Drying Overage Compensation of 3.0%

Upper/Lower Limit Drug Product Assay	Assay (%)
Upper limit	105
Lower limit	95
Number of data points − 1 (n − 1)	31
Number of data points	32
Mean	102.69
Standard deviation	1.97
Expected rejects (%)	12.01
Reliability index	1.04
Actual rejects (%)	6.25

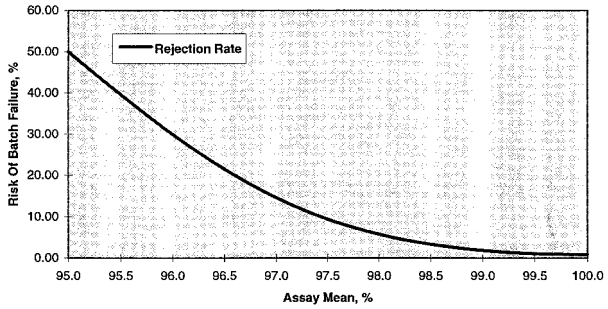
Table 3

Rate of Batch Failure for Assay of a Medium-Dosage Drug Product: Commercial Production with Fixed Loss on Drying Overage Compensation of 2.0%

Upper/Lower Limit Drug Product Assay	Assay (%)
Upper limit	105
Lower limit	95
Number of data points − 1 (n − 1)	38
Number of data points	39
Mean	100.80
Standard deviation	1.13
Expected rejects (%)	0.01
Reliability index	3.21
Actual rejects (%)	0



(a)



(b)

Figure 4. Variation of risk of batch failure with assay standard deviation for a mean of 100% and limits of 95–105%.

1.9% when the assay mean increases above 100%. Analytical experts regarded setting a 2.0% standard deviation limit on the variability of the assay results as tough, so the assay was considered in more detail.

An examination of production experience with the assay results of 20 batches of each of the 10 larger volume solid dosage form products produced at Novartis St. Johann manufacturing site showed that only 2 of these products had standard deviations close to or above 2.0% (see Table 4). One was drug product 1, the recently registered medium-dosage product mentioned above, with a maximum value of 106.9%, which exceeds the upper permitted limit of 105%. The other was drug product 2, a direct tableting mixture with a minimum value uncomfortably close to the lower permitted limit of 95%. In both these cases, corrective measures to remove an assay problem were undertaken successfully. One of the larger volume products, drug product 10, had a mean in the region of 98%, indicating an endangered assay potential. But, the standard deviation of 0.67 was so small that the batch failure risk was negligible.

Thus, working on a threshold batch failure rate limit of less than 1% for the serial production of new introductions, the assay standard deviation would have to be less than 2.0% if the assay mean is close to 100%. An even lower assay standard deviation may be required if the assay mean departs significantly from 100%. If these requirements cannot be met and corrective measures prove unsuccessful, then an assay specification differing from the European Pharmacopeial value of 95–105% should be considered when registering the product.

SETTING SPECIFICATIONS: WHAT IF CALCULATIONS

In a recent development with a high-dose and moderate-dose combination preparation, a statistical evaluation of the risk of batch failure was made on the results for

Table 4

Rate of Batch Failure for Commercial Production of Larger Volume Products: Assay

Product	Number of Data Points − 1 (n − 1)	Mean	Standard	Minimum	Maximum	Expected Rejects (%)	Reliability Index
Drug product 1, moderate dosage	19	102.85	2.01	98.20	106.90	14.244	1.050
Drug product 2, low dosage	19	99.15	1.99	95.90	103.50	2.016	1.468
Drug product 3, low dosage	19	99.74	1.64	96.90	103.00	0.259	2.701
Drug product 4, low dosage	19	98.25	1.11	96.10	101.35	0.171	2.532
Drug product 5, low dosage	19	100.36	1.47	97.80	103.80	0.093	3.711
Drug product 6, low dosage	19	101.51	1.09	99.10	103.90	0.068	3.445
Drug product 7, low dosage	19	100.32	1.31	97.20	103.30	0.020	6.704
Drug product 8, high dosage	19	99.65	1.18	97.20	101.60	0.004	12.688
Drug product 9, low dosage	19	100.36	1.05	97.70	102.10	0.001	34.907
Drug product 10, high dosage	19	97.89	0.67	96.20	99.10	0.001	23.313

Table 5

High-Dose Combination Preparation: Technology Transfer from Development to Production Summary of Risk of Batch Failure Calculations

Property	Specification Validation Protocol	Specification Change Required	Risk of Batch Failure with New Specification	Number of Batches Considered	Status
Granulate particle size					
Fraction < 90 μm	≤15%	≤15%	0.00%	6	Unchanged
Fraction > 710 μm	≤15%	≤30%	0.01%	6	Widened
Fraction > 90 μm < 710 μm	70% ± 20%	≥50%	0.00%	6	Widened
Pour and tapped volumes, tableting mixture					
Pour volume	200 ± 30 ml/100 g	180 ± 30 ml/100 g	0.00%	6	Changed
Tap volume, 2500 taps	165 ± 15 ml/100 g	150 ± 20 ml/100 g	0.38%	6	Widened
Tabletting mixture blend uniformity	100% ± 10%	100% ± 10%	Drug substance 1, 0.00%	3	Unchanged
			Drug substance 2, 0.01%	3	Unchanged
Tablet disintegration time	≤15 min	≤15 min	0.00%	6	Unchanged
Tablet hardness	80 ± 40 N	80 ± 50 N	0.00%	3	Widened
Tablet thickness	3.15 ± 0.15 mm	3.2 ± 0.2 mm	0.00%	6	Widened
Tablet friability (100 revolutions)	≤2%	≤1%	0.00%	6	Tightened
Tablet content uniformity	USP 23	USP 23	Drug substance 1, 0.00%	3	Unchanged
			Drug substance 2, 0.00%	3	Unchanged
Tablet assay	100% ± 5%		Drug substance 1, 0.00%	4	Unchanged
			Drug substance 2, 0.00%	4	Unchanged
Dissolution rate from tablet (Rotating paddle, USP)					
Drug substance 1	*Q*(60 min) 35.0%	*Q*(60 min) 45.0%	0.69%, level 2	4	Tightened
	Q(180 min) 55.0%	*Q*(180 min) 65.0%	0.32%, level 2	4	Tightened
Drug substance 2	*Q*(45 min) 55.0%	*Q*(45 min) 60.0%	0.45%, level 2	4	Tightened

the scale-up batches using the specifications set (Table 5). For some of these specifications, the risk of batch failure calculations indicated that the preset limits could be safely tightened for commercial batches. These specifications included tablet friability and tablet dissolution rate.

Other specifications required widening since actual production-scale batch size manufacturing experience showed the preset limits were too tight. Risk of batch failure calculations was made for different set limits until acceptable failure rates were achieved. The specifications requiring widening included granulate particle size and granulate and tableting mixture poured and tapped volumes.

The risk of batch failure calculation showed that other specifications for this combination product could safely be left unchanged or that no corrective measures were necessary. These properties included assay, content uniformity, tableting mixture blend uniformity, and tablet disintegration time.

ASSESSING THE IMPACT OF CHANGES IN SPECIFICATIONS: WHAT IF CALCULATIONS

The impact of a change in specifications can be readily estimated, as in the case of an impurity found in the medium-dose product mentioned above (see Table 6). Table 6 also includes data generated after submission, but the exercise illustrates the principle.

Table 6

Medium-Dosage Drug Product: Concentrations of a Drug Substance Impurity in the Drug Product

Upper impurity limit (%)	1.0	0.8
Number of data points − 1 (n − 1)	9	9
Number of data points	10	10
Mean	0.51	0.51
Standard deviation	0.21	0.21
Expected rejects (%)	0.87	7.83
Reliability index	2.37	1.21
Actual rejects (%)	0	10

The expected risk of batch failure because of this impurity is 0.87% for an upper impurity limit of 1.0%, and none of the batches manufactured would be rejected with this limit. One registration authority proposed impurity limit tightening to 0.8%, and this new limit causes the expected rate of failure to jump to a value of 7.8% and gives an actual failure rate of 10%.

Thus, this proposed impurity limit tightening from 1.0% to 0.8% would have a negative effect on the ability to meet specifications and significantly increases the risk of batch failure.

OTHER EXPERIENCE WITH THE USE OF RISK OF BATCH FAILURE CALCULATIONS

The risk of batch failure has been calculated for several other solid dosage form products, as well as for various other dosage forms, including semisolid products, parenterals (including a depot injectable product) and liquid forms (including a nasal spray product). A list of the product parameters investigated is shown in Table 7. In each case, the results have been equally encouraging.

The method can be used whenever there is numerical data available and there is a target value to be reached, even if not meeting the target does not actually lead to batch failure.

CALCULATION

Using the number, mean, standard deviation, and Student distribution functions in the Microsoft Excel spreadsheet, a template could be constructed that fulfills all the calculation needs in an easy-to-use manner. This template has been circulated within the Novartis Company development and production centers.

Table 7

Product Parameters Assessed for Risk of Batch Failure or Failure to Meet Set Requirements

Parameter	Dosage Form
Assay	Solid dosage forms Liquid dosage forms Semisolid dosage forms Parenterals
Drug substance impurities	Solid dosage forms Liquid dosage forms Semisolid dosage forms Parenterals
Dissolution rate: frequency of level 2 determinations needed	Solid dosage forms
Content uniformity	Solid dosage forms Liquid dosage forms Semisolid dosage forms Parenterals
Drug substance impurities	Solid dosage forms Liquid dosage forms Semisolid dosage forms Parenterals
Liquid characteristics: viscosity, refractive index, etc.	Liquid dosage forms Semisolid dosage forms
Leachables, residues	Liquid dosage forms Parenterals
Air/oxygen in purged systems	Liquid dosage forms
Fill volume, fill weight	Liquid dosage forms Semisolid dosage forms
Drug content per application	Nasal spray
Microbiological count	Liquid dosage forms Vials
External phase compensation for granulate yield	Solid dosage forms
Powder characteristics: particle size, bulk volume, etc.	Solid dosage forms Intermediates
Tablet/capsule characteristics: weight, disintegration time, etc.	Solid dosage forms
Incidence of necessity for level 2 dissolution rate testing	Solid dosage forms

To construct this template, start a new Excel spreadsheet.

Data in, say, cells F5 to F7	F5:F7
Enter value upper limit: in cell F12	F12
Enter value lower limit: in cell F13	F13
Value, sample number 1, in cell F15:	= F17-1
Excel count function in cell F17	= COUNT(F5:F7)
Excel mean in cell F18:	= AVERAGE(F5:F7)
Excel standard deviation in cell F19	= STDEV(F5:F7)
Value Z(upper) in cell F21:	= (F12-F18)/F19
Value Z(lower) in cell F22:	= (F18-F13)/F19

Excel TDIST function in cell F24	= TDIST(F21;10000000000;1)
Excel TDIST function in cell F25	= TDIST(F22;10000000000;1)
Excel TDIST function in cell F27	= TDIST(F21;F15;1)
Excel TDIST function in cell F28	= TDIST(F22;F15;1)
Expected rate of failure in cell F31	= (F24+F25)*100
Reliability index in cell F32	= (F27+F28)/(F24+F25)

Save this template in an appropriate place.

To use this template, it is assumed that the data to be treated have been entered into an Excel spreadsheet. The template can be added to this data spreadsheet using cut and paste.

Enter the appropriate limits in cells F12 and F13 of the template. Copy/paste the template to data for evaluation.

Place the cursor on cell F17. With the mouse, activate the function button f_x on the menu bar. The count function box appears on the screen. Use the mouse to mark the data points and then press the finish button on the count function box. The actual number of data points appears in cell F17.

Place the cursor on cell F18. With the mouse, activate the function button f_x on the menu bar. The average function box appears on the screen. Use the mouse to mark the data points and then press the finish button on the average function box. The actual mean of the data points appears in cell F18.

Place the cursor on cell F19. With the mouse, activate the function button f_x on the menu bar. The standard deviation function box appears on the screen. Use the mouse to mark the data points and then press the finish button on the standard deviation function box. The standard deviation of the data points appears in cell F19.

All the appropriate calculations are then performed by the computer.

The above template is for data with upper and lower specification limits. Suitable modifications produce a template for data having an upper limit only (e.g., tablet disintegration time) or a lower limit only (e.g., drug dissolution rate from tablets).

CONCLUSION

A tool has been developed that allows the risk of batch failure during commercial production to be estimated with encouraging accuracy from a limited number of scale-up batches during development. The tool also has value in setting or resetting specification limits or in estimating the impact of changes in specifications proposed by the registration authorities and may be used to assist defense of specifications on the basis of submitted data.