

How To Establish Sample Sizes For Process Validation Using Statistical Tolerance Intervals

By Mark Durivage, ASQ Fellow

The first article in this series, Risk-Based Approaches To Establishing Sample Sizes For Process Validation (June 2016), established the relationship between risk and sample size. Subsequent articles have introduced different methods for determining sample sizes to ensure validation activities will yield valid results. This installment will demonstrate how to use statistical tolerance limits, which use the confidence level (how sure we are) and reliability value (population value) to determine appropriate statistically valid sample sizes for process validation.



Start With FMEA

Before we begin, we must establish our definitions of risk and their associated confidence level and reliability value. These definitions can and should vary based upon the organizational needs. A good place to determine the risk level is failure mode and effects analysis (FMEA), a systematic group of activities designed to recognize, document, and evaluate the potential failure of a product or process and its effects. FMEA uses a risk priority number (RPN), which is based on the frequency, detection, and severity of a

potential failure mode. The higher the RPN, the higher the risk. However, a low probability of occurrence in conjunction with high severity and high probability of detection may still necessitate the appropriate controls for high risk.

Table 1 depicts an example FMEA with the associated risk levels. Once the risk level has been determined (low, medium, high), the appropriate confidence level and reliability can be selected using Table 3. Figure 1 depicts the linkage between FMEA, risk, and confidence level and reliability.

Table 1: Example FMEA

Process	Failure Mode	Numeric Ranking			RPN	Risk
		Frequency	Detection	Severity		
Pouch sealing	Seal not intact	1	5	5	25	High
Container fill	Incorrect amount	2	3	3	18	Medium

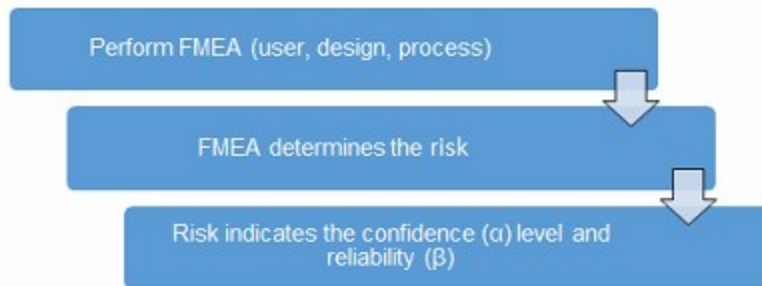


Figure 1: Risk process for determining the appropriate confidence level and reliability

Table 2 shows an example of risk level definitions with accompanying defect classifications. These definitions can and will vary based upon the product(s) produced and its intended and unintended uses.

Table 2: Example of Risk Level Definitions

Risk	Defect	Definition
High	Critical	Life threatening, or may result in death
Medium	Major	May result in temporary or permanent injury requiring medical intervention
Low	Minor	May result in minor injury, discomfort, or inconvenience not requiring medical intervention

Table 3 depicts example confidence levels and reliability values based upon risk. Of course, different confidence and reliability levels can and should be utilized based upon an organization's risk acceptance determination threshold, industry practice, guidance documents, and regulatory requirements.

Table 3: Example Confidence and Reliability Levels Based upon Risk Acceptance

Risk	Defect	Confidence (α)	Reliability (β)
High	Critical	95%	99%
Medium	Major	95%	95%
Low	Minor	95%	90%

Statistical Tolerance Intervals

When using statistical tolerance intervals, there is an assumption and expectation that the data is normally distributed. There are many ways to determine if data is normally distributed, including computer programs and spreadsheets. However, for small samples (15 or fewer) normal probability plots can be used to assess normality.

Normal probability plots can be constructed to look for linearity when using one variable. Normal probability plots provide a visual way to determine if a distribution is approximately normal. If the distribution is close to normal, the plotted points will lie close to a line. Normal probability plots are constructed by doing the following:

1. Arrange the data from smallest to largest.

1. Determine the percentile of each data value.
1. From these percentiles, do normal calculations to determine their corresponding z-scores.
1. Plot each z-score against its corresponding data value.
1. Perform a graphical review of the normal probability plot for the data points to verify if they generally fall on the line of best fit — sometimes referred to as the “pencil” test. If the data is not normally distributed, it is best to use an alternative method.

(For an example of how to create and analyze normal probability plots, refer to Chapter 4 of *Practical Engineering, Process, and Reliability Statistics*, ASQ Quality Press, 2014.)

Statistical tolerance intervals can be one-sided or two-sided, depending if the specification requires a minimum or maximum value or if the specification has an allowable range. However, it is generally best practice to use one-sided tolerance intervals — this method will provide a more conservative approach, because the risk is placed on one side rather than being split. If the specification is bilateral, use the specification that is closest to the sample mean, which is calculated from the initial sample.

The first step in using statistical tolerance intervals to determine sample sizes for process validation is to calculate the mean and standard deviation from a small sample, which should capture the expected range of variation that can reasonably be expected from the process (different batches of materials, operators, etc.). To determine this minimum initial sample for a one-sided test:

$$n = (z_{1-\alpha} + z_{1-\beta})^2 \left(\frac{\sigma}{\delta}\right)^2$$

And for a two-sided test:

$$n = \left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2 \left(\frac{\sigma}{\delta}\right)^2$$

Where:

z = normal distribution value

α = confidence

β = reliability

δ = the difference or shift we want to detect

Table 4 provides the number of samples required to detect either a 1.0σ shift or a 1.5σ shift. If the testing to be performed is destructive, expensive, or on high-value parts, using a 1.5σ shift is suggested; otherwise use a 1.0σ shift. Please note that these are the minimum number of samples to establish the sample mean and standard deviation.

Table 4: Initial Sample Table

α	β	δ=1.0σ	δ=1.5σ
One-sided Test			
0.05	0.01	16	8
0.05	0.05	11	5
0.05	0.10	9	4
Two-sided Test			
0.05	0.01	19	9
0.05	0.05	13	6
0.05	0.10	11	5

Single-Sided Specification Example

A pouch-sealing operation is considered high risk according the FMEA in Table 1. High risk requires 95% confidence (0.05α) and 99% reliability (0.01β), as shown in Table 3. The validation team decided to use a δ of 1.0σ. Table 4 requires an initial sample of 16. The 16 samples were found to be normally distributed, with a mean of 3.2 lbs and sigma of 0.23. The pouch-sealing process specification requires a minimum pull strength of 2.5 lbs.

For a single-sided specification:

$$k = \frac{|\bar{X} - T|}{s}$$

Where:

\bar{X} = the sample mean

s = the sample standard deviation

T = the tolerance (specification)

$$k = \frac{|\bar{X} - T|}{s} = \frac{|3.2 - 2.5|}{0.23} = 3.043$$

Table 5: 95% One-Sided Tolerance k Factors (Partial)

n	95% One-Sided Tolerance k Factors		
	90%	95%	99%
30	1.777	2.220	3.064
35	1.732	2.167	2.995
40	1.697	2.125	2.941

The calculated k value is 3.043. (The k value selected should be equal to or less than the calculated value.) High risk requires 95% confidence (0.05 α) and 99% reliability (0.01 β), for which Table 5 indicates a minimum of 35 samples. A minimum of 35 samples must be used for the validation. Because the company typically uses three batches for process validation, 12 samples will be randomly drawn from each of three batches for a total of 36 samples.

Three validation lots were manufactured. The 36 samples were found to be normally distributed with a mean of 3.1 lbs and sigma of 0.21.

For a one-sided tolerance:

$$TI = \bar{X} - ks \text{ or } TI = \bar{X} - ks$$

Where:

\bar{X} = sample mean

k = table values based on confidence (α) and reliability (β)

s = sample standard deviation

$$TI = \bar{X} - ks = 3.1 - (2.995 * 0.21) = 3.037$$

Since the calculated tolerance interval (3.037 lbs) is greater than the minimum pull strength of 2.5 lbs, the validation passed. We can state that we are 95% confident that the process is 99% reliable. In other words, we can claim that 95% of the time, 99% of the products produced will exceed the minimum pull strength of 2.5 lbs.

Sample size does matter! Let Mark Durivage show you which method for determining sample size is best for your process validation efforts:

Risk-Based Approaches To Establishing Sample Sizes For Process Validation

February 16, 2017 | 1:00-2:30PM EST



Double-Sided Specification Example

A filling operation is considered to be medium risk according the FMEA in Table 1. Medium risk requires 95% confidence (0.05α) and 95% reliability (0.05β), as shown in Table 3. The validation team decided to use a δ of 1.0σ . Table 4 requires an initial sample of 13. The 13 samples were found to be normally distributed, with a mean 995 g and sigma of 4.3. Filling process specification is 1000 g \pm 5 g.

For a double-sided specification:

$$k = \frac{T}{s}$$

Where:

\bar{X} = the sample mean

s = the sample standard deviation

T = the total tolerance (specification)

$$k = \frac{T}{s} = \frac{10}{4.3} = 2.326$$

The calculated k value is 2.326. (The k value selected should be equal to or less than the calculated value.) Medium risk requires 95% confidence (0.05α) and 95% reliability (0.05β), which Table 6 indicates a minimum of 80 samples. A minimum of 80 samples must be used for the validation. Because the company typically uses three batches for process validation, 27 samples will be randomly drawn from each of three batches for a total of 81 samples.

Three validation lots were manufactured. The 81 samples were found to be normally with a sample mean of 994 g and sample standard deviation of 4.2.

Table 6: 95% Two-Sided Tolerance k Factors (Partial)

n	95% Two-Sided Tolerance k Factors		
	90%	95%	99%
60	1.958	2.333	3.066
80	1.907	2.272	2.986
100	1.874	2.233	2.934

For a two-sided tolerance:

$$TI = \bar{X} \pm ks$$

Where:

\bar{X} = sample mean

k = table values based on confidence (α) and reliability (β)

s = sample standard deviation

$$TI = \bar{X} \pm ks = 994 - (2.272 * 4.2) = 994 \pm 9.542 \text{ or } 984.5 \text{ to } 1003.5$$

Recall the tolerance is 1000 g \pm 5 g (995 g to 1005 g). Since the lower tolerance interval (984.5 g) is below the lower specification of 995 g, the validation has failed. It should be apparent that there are two issues with the process. The first is that the process is not centered, and second is that there is excessive variation. The validation team must first center the process and then determine the best method(s) to reduce the variation within the process.

Choosing The Right Method For Your Process

Although the examples provided use single-sided and double-sided specifications, in practice I prefer to use the single-sided method, to account for an off-center process (same principle as C_{pk}). Remember, the criteria used should be based upon risk and should be documented and proceduralized. When a validation fails, there are three distinct options available:

- Center the process
- Reduce the variation
- Change the specification (tolerance)

I want to reinforce that different confidence levels and reliability values can and should be utilized based upon an organization's risk acceptance determination threshold, industry practice, guidance documents, and regulatory requirements. Using statistical tolerance intervals to establish sample sizes for process validation activities is a widely used and accepted practice in FDA-regulated industries.

The methods presented here have been used and successfully defended during audits and inspections. I cannot emphasize enough the importance of proceduralizing (documenting) the statistical methods and rationale your organization may use for process validation activities. Table 2 provides an example of documenting and standardizing risk levels, defect classifications, and defect definitions. Table 3 provides an example for documenting the confidence level and reliability requirements for process validation activities. I also recommend that validation and statistical technique procedures include the formulas as well as fully worked examples, like those demonstrated above, to provide clarity and guidance for those individuals writing, performing, executing, and approving process validation activities.

Subsequent articles in this series will provide additional how-to examples for applying risk-based sample size techniques to process validations in your organization. Previous articles include:

- Risk-Based Approaches To Establishing Sample Sizes For Process Validation
- How To Establish Sample Sizes For Process Validation Using The Success-Run Theorem
- How To Use Reliability-Based Life Testing Sampling For Process Validation
- How To Establish Sample Sizes For Process Validation Using C=0 Sampling Plans

About the Author

Mark Allen Durivage is the managing principal consultant at Quality Systems Compliance LLC and an author of several quality-related books. He earned a B.A.S in computer aided machining from Siena Heights University and an MS in quality management from Eastern Michigan University. Durivage is an ASQ Fellow and holds several ASQ certifications including CQM/OE, CRE, CQE, CQA, CHA, CBA, CPGP, and CSSBB. He also is a Certified Tissue Bank Specialist (CTBS) and holds a Global Regulatory Affairs Certification (RAC). Durivage resides in Lambertville, Michigan. Please feel free to email him at mark.durivage@qscompliance.com with any questions or comments, or connect with him on LinkedIn.



References:

1. Durivage, M.A., *Practical Engineering, Process, and Reliability Statistics*, Milwaukee, ASQ Quality Press, 2014.
1. Durivage, M.A. and Mehta B., *Practical Process Validation*, Milwaukee, ASQ Quality Press, 2016.
1. Durivage, M.A., *Risk-Based Approaches To Establishing Sample Sizes For Process Validation*, Life Science Connect, 2016.
1. International Organization for Standardization, *ISO 16269-6:2013 Statistical interpretation of data - Part 6: Determination of statistical tolerance intervals*, 2014.