Designation: F3263 – 17

Standard Guide for Packaging Test Method Validation[1]

This standard is issued under the fixed designation F3263; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

INTRODUCTION

The tests often used by engineers in regulated industries such as medical device or pharmaceuticals are well known and referenced in both ASTM and ISO literature. However, questions around the validation of these tests are not nearly as well understood. Questions that often arise are; how should one validate these test methods? Should they be validated at all? To what degree should they be validated?

One answer to this is the guidance provided by ISO 11607-1 and ISO 11607-2 where it is stated that "all test methods used to show compliance with this part of ISO 11607 shall be validated and documented."

Unfortunately, this does not answer all questions as little is provided in how to demonstrate conformance to these requirements. This is due to the fact that there needs to be a great deal of flexibility in how these test methods are used. Not all circumstances and test methods require the same degree of scrutiny. Therefore, when assessing when, why, and how a test method should be validated, it is critical to keep this flexibility in mind and use the best tools available to answer the above questions appropriately for a given situation. A robust risk assessment process is arguably the best tool for determining the risk associated with a particular design element being tested. For example, there are clear differences in the risk associated with testing the adhesion of a label versus testing the integrity of a sterile barrier when viewed from the perspective of patient safety. If a label is missing, the product would be discarded, and a new one that is properly labeled chosen. However, if the sterile barrier has been compromised due to a seal breach or pinhole in the web of the material, this may go undetected, a contaminated device may be used, and the patient may become infected.

The typical process for determining the level of risk associated with medical device packaging components is the failure mode effects analysis tool, commonly referred to as an FMEA. The FMEA process is intended to identify potential failure modes for a product or process, to assess the risk associated with those failure modes, to rank the issues in terms of importance, and to identify and document mitigation strategies that address the most serious concerns. There are many guides and standards available that describe this process, such as SAE J1739, AIAG FMEA-3 and MIL-STD-1629A. The present guide will be helpful in proposing ways to go about defining what approaches to test method validation that will work best in a given application based on the associated risk, and will also provide guidance on the execution of the validation.

1. Scope

1.1 This guide provides information to clarify the process of validating packaging test methods specific for an organization utilizing them as well as through inter-laboratory studies (ILS), addressing consensus standards with inter-laboratory studies (ILS) and methods specific to an organization.

1.1.1 ILS discussion will focus on writing and interpretation of test method precision statements and on alternative approaches to analyzing and stating the results.

1.2 This document provides guidance for defining and developing validations for both variable and attribute data applications.

[1] This test method is under the jurisdiction of ASTM Committee F02 on Primary Barrier Packaging and is the direct responsibility of Subcommittee F02.50 on Package Design and Development.

Current edition approved Dec. 15, 2017. Published March 2018. DOI: 10.1520/F3263–17.

1.3 This guide provides limited statistical guidance; however, this document does not purport to give concrete sample sizes for all packaging types and test methods. Emphasis is on statistical techniques effectively contained in reference documents already developed by ASTM and other organizations.

1.4 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety, health, and environmental practices and determine the applicability of regulatory limitations prior to use.*

1.5 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods

E456 Terminology Relating to Quality and Statistics

E691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method

E2282 Guide for Defining the Test Result of a Test Method

E2782 Guide for Measurement Systems Analysis (MSA)

F17 Terminology Relating to Primary Barrier Packaging

F2097 Guide for Design and Evaluation of Primary Flexible Packaging for Medical Products

2.2 *ISO Standards:*[3]

ISO 11607-1: 2006/A1: 2014 Packaging for terminally sterilized medical devices—Part 1: Requirements for materials, sterile barrier systems, and packaging, Amendment 1

ISO/TS 16775 Packaging for terminally sterilized medical devices—Guidance on the application of ISO 11607-1 and ISO 11607-2

3. Terminology

3.1 *Definitions of Terms Specific to This Standard:*

3.1.1 *accuracy, n*—see E177.

3.1.2 *alpha risk error (α), n*—the probability that an inspector will reject a conforming unit. Also referred to as producers risk or type I error. For the purposes of this document this error type will be referred to as Alpha risk error.

3.1.3 *appraiser, n*—term used to identify individual(s) that will execute test method validation activities. May commonly also be referred to as appraisers or technicians.

3.1.4 *as defined by team with rationale, n*—the validation team determines a performance level or sample size with acceptance criteria. When a test method falls under this category another option may be no testing required.

3.1.5 *attribute test method, n*—tests that return a pass/fail output measurement on a characteristic that is either conforming or nonconforming. Variable measurement data treated as attribute also qualifies.

3.1.6 *acceptable quality level (AQL), n*—represents a level of quality that a sampling plan routinely accepts. Lots at or below the AQL are accepted at least 95% of the time. The AQL may be determined from the sampling plan's Operating Characteristic (OC) Curve.

3.1.7 *beta risk error (β), n*—the probability that an inspector will accept a nonconforming unit. Also referred to as beta error (escape rate) or type II error. For the purposes of this document this error type will be referred to as Beta risk error or (β).

3.1.8 *borderline samples, n*—marginally passing or failing samples.

3.1.9 *comparative test method, n*—a test method that is used for comparing the means of two or more populations using a statistical test (e.g. 2-sample t test, ANOVA test). A comparative test method is NOT used for accepting or rejecting individual units, and the output usually does NOT have specification limits.

3.1.10 *failure modes effects analysis, n*—Failure modes and effects analysis (FMEA) is a step-by-step approach for identifying all possible failures in a design, a manufacturing or assembly process, or a product or service.[4]

3.1.11 *highly instrumental method, n*—a test method where the result is not dependent on the operator.

3.1.12 *lot tolerance percent defective (LTPD), n—in a sampling plan*, represents a level of quality that a sampling plan routinely rejects. Lots at or above the LTPD are rejected at a probability level determined by the confidence level. The LTPD may be determined from the sampling plan's Operating Characteristic (OC) Curve. Also known as the Rejectable Quality Level (RQL), Limiting Quality Level (LQ), and Unacceptable Quality Level (UQL).

3.1.13 *measurement resolution, n*—the smallest detectable increment that can be measured by the test method.

3.1.14 *precision, n*—see E177.

3.1.15 *%P/T (precision to tolerance ratio), n*—%P/T is a test method performance metric of a Gage R&R study. It measures the percentage of the tolerance attributable to test method variation. Depending on the component of test method variation being assessed, %P/T has three forms: $\%P/T_{repeatability}$, $\%P/T_{reproducibility}$, and $\%P/T_{total}$.

3.1.16 *operating characteristic (OC) curve, n*—plot of process or lot quality versus the probability of acceptance; the protection offered by a sampling plan shown graphically.

3.1.17 *repeatability, n*—see E177.

3.1.18 *reproducibility, n*—see E177.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

[3] Available from American National Standards Institute (ANSI), 25 W. 43rd St., 4th Floor, New York, NY 10036, http://www.ansi.org.

[4] http://asq.org/learn-about-quality/process-analysis-tools/overview/fmea.html

3.1.19 *%R&R (reproducibility and repeatability), n*—%R&R is a test method performance metric of a Gage R&R study. It measures the percentage of the historical process variation attributed to test method variation. Calculating %R&R requires a known historical standard deviation.

3.1.20 *self-evident, n*—an inspection that meets both of the following criteria: *(1)* The nonconformance is discrete in nature, meaning there cannot be a transition region between conforming and non-conforming product which has at least the potential for misclassification. *(2)* Little or no training is required to discriminate between conforming and non-conforming product.

3.1.21 *% study variation (%SV), n*—a test method performance metric of a Gage R&R study. It measures the percentage of the total variation of a Gage R&R study attributed to the test method variation.

3.1.22 *subject matter expert (SME), n*—subject matter expert on the product and/or process. Engineers, inspector, technicians, trainers and production supervisors who have a strong understanding of the failure modes may be considered SMEs.

3.1.23 *test method, n*—see ASTM E2282.

3.1.24 *test systems, n*—instrument and associated materials required to perform the test.

3.1.25 *trial, n*—a trial is defined as one inspector or a piece of equipment in the case of a highly instrumental method making one measurement or pass/fail decision. If three inspectors each evaluate the same device once, it counts as three trials. Similarly, if one inspector evaluates the same device twice during the test, it counts as two trials.

3.1.26 *user defined with minimum sample size restrictions, n*—the validation team selects the performance level, but the test shall satisfy minimum requirements for the proportion of conforming and nonconforming trials or for measurement collected to calculate variability.

3.1.27 *validation team, n*—the responsible party for the test method validation that seeks cross-functional input, validates the effectiveness of the test method, and completes corrective actions associated with any test failures.

3.1.28 *variable test method, n*—a test method that produces numerical results with reference to a continuous scale.

3.1.29 *visual aid, n*—visual media used for training purposes or to illustrate manufacturing process steps.

3.2 *Acronyms:*

3.2.1 *AQL*—Acceptable Quality Level

3.2.2 *ATMV*—Attribute Test Method Validation

3.2.3 *DV*—Design Verification

3.2.4 *FMEA*—Failure Modes and Effects Analysis

3.2.5 *LTPD*—Lot Tolerance Percent Defective

3.2.6 *MVR*—Master Validation Record

4. Significance and Use

4.1 Addressing consensus standards with inter-laboratory studies (ILS) and methods specific to an organization. Test methods need to be validated in many cases, in order to be able to rely on the results. This has to be done at the organization performing the tests but is also performed in the development of standards in inter-laboratory studies (ILS), which are not substitutes for the validation work to be performed at the organization performing the test.

4.1.1 *Validations at the Testing Organization*—Validations at the test performing organization include planning, executing, and analyzing the studies. Planning should include description of the scope of the test method which includes the description of the test equipment as well as the measurement range of samples it will be used for, rationales for the choice of samples, the amount of samples as well as rationales for the choice of methodology.

4.1.2 *Objective of ILS Studies*—ILS studies (per E691-14) are not focused on the development of test methods but rather with gathering the information needed for a test method precision statement after the development stage has been successfully completed. The data obtained in the interlaboratory study may indicate however, that further effort is needed to improve the test method. Precision in this case is defined as the repeatability and reproducibility of a test method, commonly known as gage R&R. For interlaboratory studies, repeatability deals with the variation associated within one appraiser operating a single test system at one facility whereas reproducibility is concerned with variation between labs each with their own unique test system. It is important to understand that if an ILS is conducted in this manner, reproducibility between appraisers and test systems in the same lab are not assessed.

4.1.3 *Overview of the ILS Process*—Essentially the ILS process consists of planning, executing, and analyzing studies that are meant to assess the precision of a test method. The steps required to do this from an ASTM perspective are; create a task group, identify an ILS coordinator, create the experimental design, execute the testing, analyze the results, and document the resulting precision statement in the test method. For more detail on how to conduct an ILS refer to E691-14.

4.1.4 *Writing Precision and Bias Statements*—When writing Precision and Bias Statements for an ASTM standard, the minimum expectation is that the Standard Practice outlined in E177-14 will be followed. However, in some cases it may also be useful to present the information in a form that is more easily understood by the user of the standard. Examples can be found in 4.1.5 below.

4.1.5 *Alternative Approaches to Analyzing and Stating Results—Variable Data:*

4.1.5.1 *Capability Study:*

(1) A process capability greater than 2.00 indicates the total variability (part-to-part plus test method) of the test output should be very small relative to the tolerance. Mathematically,

$$Pp = \frac{Specifiction\ Tolerance}{6\sigma_{Total}} \geq 2.00$$
$$\geq \sigma_{Total} \leq \frac{1}{12} Specification\ Tolerance \quad (1)$$

(2) Notice, σ_{Total} in the above equation includes σ_{Part} and σ_{TM}. Therefore, two conclusions can be made:

(a) The test method can discriminate at least 1/12 of the tolerance and hence the test method resolution is adequate Therefore, no additional analysis such as a Gage R&R Study is necessary.

(b) The measurement is precise relative to the specification tolerance.

(3) In addition, since the TMV capability study requires involvement of two or more operators utilizing one or more test systems, a high capability number will prove consistent test method performance across operators and test systems.

4.1.5.2 *Gage R&R Study:*

(1) The proposed acceptance criteria below for %SV, %R&R, and %P/T came from the industry-wide adopted requirements for measurement systems. According to Automotive Industry Action Group (AIAG) Measurement System Analysis Manual (4th edition, p. 78), a test method can be accepted if the test method variation (σ_{TM}) counts for less than 30 percent of the total variation of the study (σ_{Total}).

(2) This is equivalent to:A process capability greater than 2.00 indicates the total variability (part-to-part plus test method) of the test output should be very small relative to the tolerance. Mathematically,

$$\%SV = \frac{\sigma_{TM}}{\sigma_{Total}} \leq 30\% \quad (2)$$

(3) When historical data is available to evaluate the variability of the process, we should also have:

$$\%R\&R = \frac{\sigma_{TM}}{\sigma_{Process}} \leq 30\% \quad (3)$$

(4) For %P/T, another industry-wide accepted practice is to represent the population using the middle 99% of the normal distribution.[5] And ideally, the tolerance range of the output should be wider than this proportion. For a normally distributed population, this indicates:

$$Specification\ Tolerance \geq 5.15\sigma_{Total} \quad (4)$$

(5) The factor 5.15 in the above equation is the two-sided 99% Z-score of a normal distribution. Therefore:

$$\%P/T = \frac{\sigma_{TM}}{Specification\ Tolerance} \leq \frac{\sigma_{TM}}{5.15 \times \sigma_{Total}} \leq \frac{30\%}{5.15} = 5.8\% \quad (5)$$

(6) In practice this means that a test method with up to 6% P/T reproducibility would be effective at assessing the P/T for a given design.

4.1.5.3 *Power and Sample Size Study:*

(1) When comparing the means of two or more populations using statistical tests, excessive test method variability may obscure the real difference ("Signal") and decrease the power of the statistical test. As a result, a large sample size may be needed to maintain an adequate power (≥ 80%) for the statistical test. When the sample size becomes too large to accept from a business perspective, one should improve the test method before running the comparative test. Therefore, an accept /reject decision on a comparative test method could be made based on its impact on the power and sample size of the comparative test (ex. 2 Sample T-test).

[5] Design and Analysis of Gage R&R Studies by Burdick, Borror, and Montgomery, page 3.

4.2 *Attribute Test Method Validation:*

4.2.1 *Objective of Attribute Test Method Validation*—Attribute test method validation (ATMV) demonstrates that the training and tools provided to inspectors enable them to distinguish between good and bad product with a high degree of success. There are two criteria that are used to measure whether an ATMV has met this objective. The primary criterion is to demonstrate that the maximum escape rate, β, is less than or equal to its prescribed threshold of βmax. The parameter β is also known as Type II error, which is the probability of wrongly accepting a non-conforming device. The secondary criterion is to demonstrate that the maximum false alarm rate, α, is less than or equal to its prescribed threshold of αmax. The parameter α is also known as Type I error, which is the probability of wrongly rejecting a conforming device.

4.2.2 *Overview of the ATMV Process*—This section describes how an ATMV typically works. In an attribute test method validation, a single, blind study is conducted that is comprised of both conforming and non-conforming units. The ATMV passes when the requirements of the both sampling plans are met. The first sampling plan demonstrates that the test method meets the requirements for the maximum allowable beta error (escape rate), and the second sampling plan demonstrates that the test method meets the requirements for the maximum allowable alpha error (false alarm rate). In other words, the test method is able to demonstrate that it accepts conforming units and rejects non-conforming units with high levels of effectiveness. The beta error sampling plan will consist entirely of nonconforming units. The total number of beta trials conducted by each inspector[6] are pooled together, and their total number of misclassifications (nonconforming units that were accepted) need to be less than or equal to the number of failures prescribed by the beta error sampling plan. The alpha error sampling plan will consist entirely of conforming units. The total number of alpha trials conducted by each inspector are pooled together, and their total number of misclassifications (conforming units that were rejected) need to be less than or equal to the number of failures prescribed by the alpha error sampling plan.

4.2.3 *ATMV Examples*—Attribute test methods cover a broad range of testing. Examples of these test method categories are listed in Table 1. The right half of the table consists of test methods that return qualitative responses, and the left half of the table contains test methods that provide variable measurement data.

[6] Inspector may be a machine.

TABLE 1 ATMV Examples

Inspector	Quantitative		Qualitative	
	Tactile	Visual	Tactile	Visual
Human	Pin gages	Dimensional templates	Bumps or burrs on a finished surface	Bubbles, voids or discoloration of product
Machine	Bed of nails used in printed circuit board testing	Automated imaging systems	Contact profilometer	Automated inspection systems

4.2.4 *ATMV for Variable Measurement Data*—It is a good practice to analyze variable test methods as variable measurement data whenever possible. However, there are instances where measurement data is more effectively treated as qualitative data. Example: A Sterile Barrier System (SBS) for medical devices with a required seal strength specification of 1.0-1.5 lb./in. is to be validated. A tensile tester is to be used to measure the seal strength, but it only has a resolution of 0.01 lbs. As a result, the Ppk calculations typically fail, even though there is very rarely a seal that is out of specification in production. The validation team determines that the data will need to be treated as attribute, and therefore, an ATMV will be required rather than a variable test method validation.

4.2.5 *Self-evident Inspections*—This section illustrates the requirements of a self-evident inspection called out in the definitions above. To be considered a self-evident inspection, a defect is both discrete in nature and requires little or no training to detect. The defect cannot satisfy just one or the other requirement.

4.2.5.1 The following may be considered self-evident inspections:

(1) Sensor light illuminates when lubricity level on a wire is correct and otherwise does not light up when lubrication is insufficient – Since the test equipment is creating a binary output for the inspector and the instructions are simple, this qualifies as self-evident. However, note that a test method validation involving the equipment needs to be validated.

(2) Component is present in the assembly – If the presence of the component is reasonably easy to detect, this qualifies as self-evident since the outcome is binary.

(3) The correct component is used in the assembly – As long as the components are distinct from one another, this qualifies as self-evident since the outcome is binary.

4.2.5.2 The following would generally not be considered self-evident inspections:

(1) Burn or heat discoloration – Unless the component completely changes color when overheated, this inspection is going to require the inspector to detect traces of discoloration, which fails to satisfy the discrete conditions requirement.

(2) Improper forming of S-bend or Z-bend – The component is placed on top of a template, and the inspector verifies that the component is entirely within the boundaries of the template. The bend can vary from perfectly shaped to completely out of the boundaries in multiple locations with every level of bend in-between. Therefore, this is not a discrete outcome.

(3) No nicks on the surface of the component – A nick can vary in size from "not visible under magnification" to "not visible to the unaided eye" to "plainly visible to the unaided eye". Therefore, this is not a discrete outcome.

(4) No burrs on the surface of a component – Inspectors vary in the sensitivity of their touch due to callouses on their fingers, and burrs vary in their degree of sharpness and exposure. Therefore, this is neither a discrete condition nor an easy to train instruction.

(5) Component is cracked – Cracks vary in length and severity, and inspectors vary in their ability to see visual defects. Therefore, this is neither a discrete outcome nor an easy to train instruction.

4.2.6 *ATMV Steps:*

4.2.6.1 *Step 1 – Prepare the test method documentation:*

(1) Make sure equipment qualifications have been completed or are at least in the validation plan to be completed prior to executing the ATMV.

(2) Examples of equipment settings to be captured in the test method documentation include environmental or ambient conditions, magnification level on microscopes, lighting and feed rate on automatic inspection systems, pressure on a vacuum decay test and lighting standards in a cleanroom, which might involve taking lux readings in the room to characterize the light level.

(3) Work with training personnel to create pictures of the defects. It may be beneficial to also include pictures of good product and less extreme examples of the defect, since the spectrum of examples will provide better resolution for decision making.

(4) Where possible, the visual design standards should be shown at the same magnification level as will be used during inspection.

(5) Make sure that the ATMV is run using the most recent visual design standards and that they are good representations of the potential defects.

4.2.6.2 *Step 2 – Establish acceptance criteria:*

(1) Identify which defects need to be included in the test.

(2) Use scrap history to identify the frequency of each defect code or type. This could also be information that is simply provided by the SME.

(3) Do not try to squeeze too many defects into a single inspection step. As more defects are added to an inspection process, inspectors will eventually reach a point where they are unable to check for everything, and this threshold may also show itself in the ATMV testing. Limits will vary by the type of product and test method, but for visual inspection, 15-20 defects may be the maximum number that is attainable.

4.2.6.3 *Step 3 – Determine the required performance level of each defect:*

(1) If the ATMV testing precedes completion of a risk analysis, the suggested approach is to use a worse-case outcome or high risk designation. This needs to be weighed against the increase in sample size associated with the more conservative rating.

(2) Failure modes that do not have an associated risk index may be tested to whatever requirements are agreed upon by the validation team. If a component or assembly can be scrapped for a particular failure mode, good business sense is to make sure that the inspection is effective by conducting an ATMV.

(3) Pin gages are an example of a variable output that is sometimes treated as attribute data due to poor resolution combined with tight specification limits. In this application, inspectors are trained prior to the testing to understand the level of friction that is acceptable versus unacceptable.

(4) Incoming inspection is another example of where variable data is often treated as attribute. Treating variable measurements as pass/fail outcomes can allow for less complex measurement tools such as templates and require less training for inspectors. However, these benefits should be weighed against the additional samples that may be required and the degree of information lost. For instance, attribute data would say that samples centered between the specification limits are no different than samples just inside of the specification limits. This could result in greater downstream costs and more difficult troubleshooting for yield improvements.

4.2.6.4 *Step 4 – Determine acceptance criteria:*

(1) Refer to your company's predefined confidence and reliability requirements; or

(2) Refer to the chart example in Appendix X1.

4.2.6.5 *Step 5 – Create the validation plan:*

(1) Determine the proportion of each defect in the sample.

(a) While some sort of rationale should be provided for how the defect proportions are distributed in the ATMV, there is some flexibility in choosing the proportions. Therefore, different strategies may be employed for different products and processes, for example 10 defective parts in 30 or 20 defects in 30. The cost of the samples along with the risk associated with incorrect outcomes affects decision making.

(b) Scrap production data will often not be available for new products. In these instances, use historical scrap from a similar product or estimate the expected scrap proportions based on process challenges that were observed during development. Another option is to represent all of the defects evenly.

4.2.6.6 *Step 6 – Determine the number of inspectors and devices needed:*

(1) When the number of trials is large, consider employing more than three inspectors to reduce the number of unique parts required for the test. More inspectors can inspect the same parts without adding more parts to achieve additional trials and greater statistical power.

(2) Inspectors are not required to all look at the same samples, although this is probably the simplest approach.

(3) For semi-automated inspection systems that are sensitive to fixture placement or setup by the inspector, multiple inspectors should still be employed for the test.

(4) For automated inspection systems that are completely inspector independent, only one inspector is needed. However, in order to reduce the number of unique parts needed, consider controlling other sources of variation such as various lighting conditions, temperature, humidity, inspection time, day/night shift, and part orientations.

4.2.6.7 *Step 7 – Prepare the Inspectors:*

(1) Train the inspectors prior to testing:

(a) Explain the purpose and importance of ATMV to the inspectors.

(b) Inspector training should be a two-way process. The validation team should seek feedback from the inspectors on the quality and clarity of visual standards, pictures and written descriptions in the inspection documentation.

(1) Are there any gray areas that need clarification?

(2) Would a diagram be more effective than an actual picture of the defect?

(c) Review borderline samples. Consider adding pictures/diagrams of borderline samples to the visual standards. In some cases there may be a difference between functional and cosmetic defects. This may vary by method/package type.

(d) Some validation teams have performed dry run testing to characterize the current effectiveness of the inspection. Note that the same samples should not be used for dry run testing and final testing if the same inspectors are involved in both tests.

4.2.6.8 *Step 8 – Select a representative group of inspectors as the test group:*

(1) There will be situations, such as site transfer, where all of the inspectors have about the same level of familiarity with the product. If this is the case, select the test group of inspectors based on other sources of variability within the inspectors, such as their production shift, skill level or years of experience with similar product inspection.

(2) The inspectors selected for testing should at least have familiarity with the product, or this becomes an overly conservative test. For example, a lack of experience with the product may result in an increase in false positives.

(3) Document that a varied group of inspectors were selected for testing.

4.2.6.9 *Step 9 – Prepare the Test Samples:*

(1) Collect representative units.

(a) Be prepared for ATMV testing by collecting representative defect devices early and often in the development process. Borderline samples are particularly valuable to collect at this time. However, be aware that a sample that cannot even be agreed upon as good or bad by the subject matter experts is only going to cause problems in the testing. Instead, choose samples that are representative of "just passing" and "just failing" relative to the acceptance criteria.

(2) Use the best judgment as to whether the man-made defect samples adequately represent defects that naturally occur during the sealing process, distribution simulation, or other manufacturing processes, for example. If a defect cannot be adequately replicated and/or the occurrence rate is too low to provide a sample for the testing, this may be a situation where the defect type can be omitted with rationale from the testing.

(3) Estimate from a master plan how many defects will be necessary for testing, and try to obtain 1.5 times the estimated number of samples required for testing. This will allow for weeding out broken samples and less desirable samples.

(4) Traceability of samples may not be necessary. The only requirement on samples is that they accurately depict conformance or the intended nonconformance. However, capturing traceability information may be helpful for investigational purposes if there is difficulty validating the method or if it is desirable to track outputs to specific non-conformities.

(5) There should preferably be more than one SME to confirm the status of each sample in the test. Keep in mind that a trainer or production supervisor might also be SMEs on the process defect types.