Generic Basis Values and Acceptance Criteria for Composite Materials

E. Clarkson[1]

[1] National Institute for Aviation Research, Wichita State Univerity, Wichita, Kansas, USA

Correspondence: E. Clarkson, National Institute for Aviation Research, Box 93, Wichita State Univerity, Wichita 67260-0093, Kansas, USA. Tel: 316-978-3952. E-mail: bclarkson@NIAR.wichita.edu

Abstract

The approaches used to compute engineering design values (A-basis and B-basis) and acceptance sampling criteria were developed independently in the twentieth century. This was a practical approach for that time period but it isn't working well for new materials with process dependent strength and modulus characteristics, such as carbon fiber composite materials. This paper lays out an approach designed to meet industry needs for identification of engineering design values applicable to the majority of manufacturing facilities using approved processing procedures for carbon-fiber composite materials along with corresponding acceptance criteria set to specific values for both the consumer's and the producer's risks.

Keywords: B-basis, A-basis, design values, composite materials, acceptance criteria, producer's risk, consumer's risk

1. Introduction

The acceptance sampling plans developed in the twentieth century were created and indexed based on the producer's risk (the probability of rejecting acceptable material) rather than the consumer's risk (the probability of accepting poor quality material). This was a practical approach for that time but it isn't working well for many new materials, such as carbon fiber composites, with characteristics that are highly dependent on processing variables. This leads to high variability between manufacturing sites. A very different approach is needed for certification of new materials being developed in the 21st century, one that allows many different manufacturers to use these materials without requiring a complete qualification testing program. Generic design values and acceptance criteria are a solution to this problem.

This economically feasible approach allows users to specify both the consumer's risk and the producer's risk. Given a new manufacturing facility or a change to a process procedure for a previously qualified material, it will allow engineering basis values to be set for the new procedure with a reduced dataset by making a comparison with the original qualification data. If the new product is sufficiently similar to the original qualification sample, then the two can be considered equivalent in terms of the engineering basis values.

The economic state of the aircraft industry is reaching a critical point as large transport aircraft manufacturers and airlines are investigating all methods of reducing manufacturing costs and increasing operational efficiency. Currently, establishing and protecting engineering basis values for a commercially available material requires considerable resources from every manufacturing facility that wants to use the material. NCAMP – the National Center for Advanced Materials Performance - has been working to develop a publicly accessible database of properties of composite materials.

1.1 Consumer's Risk and Producer's Risk

In the terminology of acceptance sampling, the consumer's risk is the risk of accepting material that should have been rejected, and the producer's risk is the risk of rejecting material that should have been accepted[1].

As illustrated in Figure 1, there are four possible outcomes to any hypothesis test: two correct conclusions and two erroneous conclusions. The two wrong conclusions are termed Type I and Type II errors to distinguish them. The probability of making a Type I error (wrongly concluding that the two composite part manufacturers do NOT make equivalent material when, in fact, they do) is specified by α, while the probability of a Type II error (wrongly concluding that the two composite part manufacturers make equivalent material when, in fact, they do not) is specified by β. The probability of correctly concluding the two samples come from equivalent populations is termed power of the test $(1-\beta)$.

	Materials are equal	*Materials are not equal*
Conclusion that materials are equal	*Correct Decision*	*Type II Error Consumer's Risk*
Conclusion that materials are not equal	*Type I Error Producer's Risk*	*Correct Decision*

Figure 1. Relationship of Type I and Type II Errors

One puzzling aspect to the current standard practices of acceptance sampling for materials is that most products will have more than one key characteristic that must be monitored, yet sampling plans are set up for a single characteristic. Although sampling plans that allow for multiple types of defects exist, generally, each key material characteristic is evaluated separately, with an unstated (and frequently inaccurate) underlying assumption that the key characteristics are independent.

Another puzzling aspect of the standard practice is that acceptance plans only give probabilities for the producer's risk. This equates to the default hypothesis that the material is acceptable – material is considered to be good until proven bad. For example, the sampling plans detailed in Mil-Std-105, a very widely used set of acceptance sampling plans, are indexed by the producer's risk.

Why weren't sampling plans based on the consumer's risk, since acceptance sampling plans are typically constructed by consumers for their own benefit? This was due to the difficulty of the computations, given they were done prior to the computing era, not to mention that they required a value for the acceptable difference which was typically undefined. In addition, for small acceptable differences, they will result in extremely large sample size requirements.

In order to create sampling plans that were computationally and economically viable at that time, the producer's risk was used. This approach worked well enough for most applications during the 20th century. Unfortunately, using the producer's risk had the side effect of conflating failure to reject the null (which assumes the material good) with acceptance of the null (concluding the material is good).

1.2 Current Approach for Computing Property Design Values

Before a new material, such as a carbon fiber composite, can be used in the design of an aircraft part, the design engineers must understand its property values. They must ensure that a part exposed to stresses, such as a strut in an airplane wing, is composed of materials that will not fail under the greatest anticipated stress. The material's design values represent the lowest property values that the material might have and still be considered acceptable for that use. Acceptance criteria are set to ensure that design values are upheld.

Specifically, A-basis and B-basis design values are computed to comply with the following statistical definitions[2]:

- A-basis value: An engineering value at the lower end of a 95% confidence interval for the 1st percentile.
- B-basis value: An engineering value at the lower end of a 95% confidence interval for the 10th percentile.

These design values are computed for all of the key **strength** properties of a composite material, such as compressive strength in the warp direction. Because composite materials are considered more sensitive to environmental conditions than metals, there are requirements for tests run under different conditions such as in a cold, dry environment or in hot, wet condition. Data is collected on modulus properties as well, but the mean values, rather than A- or B-basis values, are used for design.

1.4 Current Approach for Testing Equivalence

Two mutually exclusive hypotheses, termed the null (H_0) and the alternative (H_1), are defined along with an α value specifying the maximum probability of Type I errors as defined above. The null hypothesis is assumed true and must contain the equality. In current methodology, the null hypothesis is that materials are equivalent.

M represents the true – unknown – mean of the original population. M_1 represents the unknown mean of the same material produced by another composite part manufacturer (CPM):

$$\text{Two-sided tests:} \quad H_0: M = M_1 \quad H_1: M \neq M_1$$

$$\text{One-sided tests:} \quad H_0: M \leq M_1 \quad H_1: M > M_1 \tag{1}$$

Strength properties use one-sided tests, while modulus properties use a two-sided test. A test statistic is computed using the test results. If the probability of the computed test statistic under the assumption of the null hypothesis is less than α, then the null hypothesis is rejected and the alternative hypothesis is accepted as true. If not, then the null hypothesis is retained as plausible.

While the exact number of different properties and conditions that are tested varies, for a material to meet the standards of the Composite Materials Handbook (CMH17 Rev G) the minimum sample size required to determine B-basis values for any given property is a total of 18 specimens, six from each of three independent batches with specimens from each batch divided into two separate cure cycles. For A-basis values, it's a minimum of 55 specimens from five different independent batches. For establishing equivalency due to a change in the processing recipe or manufacturing facility, one batch is considered sufficient, with a minimum of eight specimens created using at least two separate cure cycles.

A separate equivalence test is performed for each material property and condition being evaluated. As we examine the results of two dozen or more different property tests from different facilities, the probability that one or more statistical failures occur due to random chance alone is quite high – see Figure 2 below.
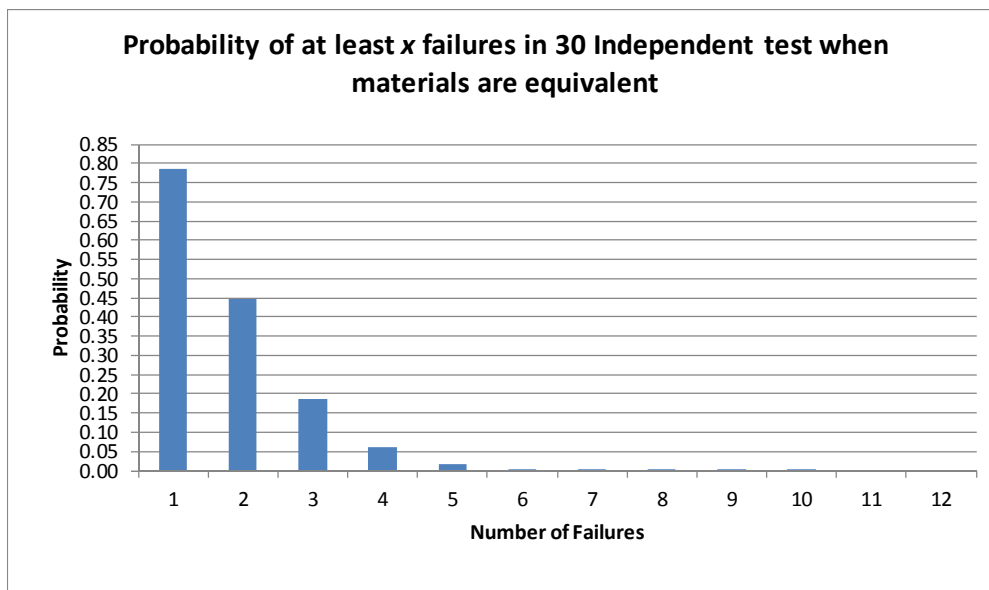

Figure 2. Binomial Distribution for 5% Failure rate

Any test that results in a detectible difference is considered to have failed equivalency. A detectible difference is defined as a test result giving less than 5% probability that the new sample is from the same population as the original. A single failure does not require concluding that the new sample is NOT equivalent, but it does require a subjective judgment on the part of any certification authorities, which can lead to delays and difficulties for a user of the composite material.

In practical terms, this results in nearly every equivalence dataset requiring subjective evaluation by experienced engineers to determine if the test failures of the new dataset are minor enough to justify calling it 'equivalent' to the original qualification dataset. This is not an objective decision, but a judgment call based on an evaluation of the overall performance of the material and the uses to which it will be put. This situation is deliberate because, while processing and examining all failures is expensive and time-consuming, it is preferable to mistakenly accepting substandard material. Is it any wonder that there are then substantial delays in working out whether the actual parts produced can be certified for use in aircraft structures?

1.3 Problem Statement

Because acceptance sampling criteria must, first and foremost, uphold the design values, the methodology and assumptions used to compute acceptance criteria should not be changed without making a corresponding change in the methodology and assumptions used to compute design values.

There are several underlying assumptions of the current computations that contribute to our current methodology being less than ideal for composite materials:

- Each property is assumed to be independent of all other tested properties.
- All properties conform to an underlying normal distribution (acceptance criteria only)
- A new sample is from a population that has the same mean (or higher for strength properties) as the original qualification dataset for all properties.
- Different manufacturing facilities using the same batch of material and the same processing procedures will produce specimens with the same property values.

These issues combine to result in a paradoxical situation; if we use larger samples that allow more accurate assessment of the true means and variances of the different properties, we decrease the probability of a sample passing the acceptance criteria.

Small samples have a large uncertainty about the true property values for both qualification samples and equivalence samples, making detection of small differences unlikely. When more data is available for acceptance testing, smaller differences become detectable. This has led to a situation where competent users of a material are essentially 'punished' for increasing the number of tests performed and 'rewarded' for using the minimum allowable sample sizes. Using this approach, a multivariate analysis only increases the sensitivity of the test to small differences, exacerbating this problem.

1.3 Example

NCAMP data for Hexcel 8552 IM7 longitudinal tension strength in the cold dry condition[3]. Figure 3 shows the results of the qualification sample tests, the calculated normal distribution curve, the minimum acceptable mean value, and the minimum acceptable specimen value.
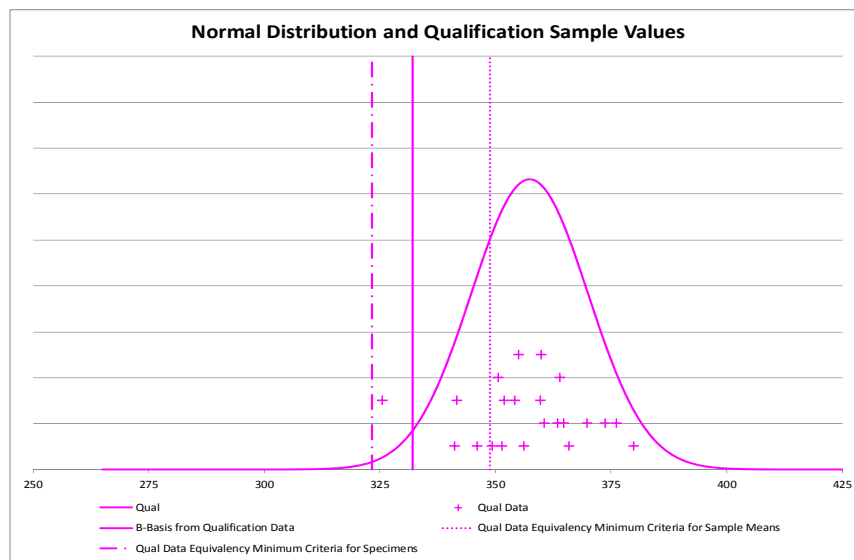

Figure 3. Qualification Sample Results

Figure 4 adds the results of additional equivalency samples produced by different manufacturers. The added blue line represents the normal distribution of the combined datasets. It's clear from the figure that there is a great deal of variability between manufacturers that was not included in the computations based on the qualification sample alone. Out of nine equivalency samples, five failed to meet the equivalency criteria for this property.
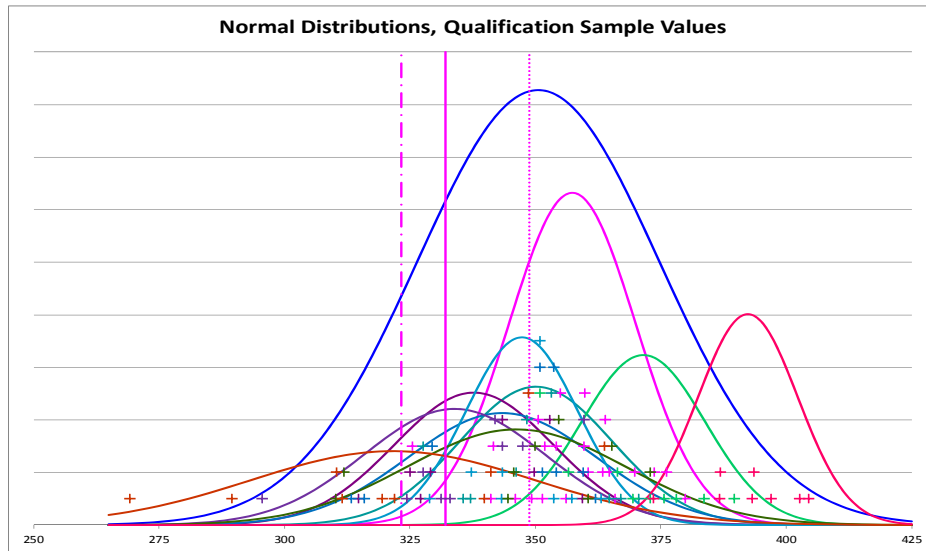
Figure 4. Qualification and Equivalency Sample Results

Figure 5 shows the results of using the combined sample to compute the B-basis and equivalence criteria. While the B-basis is reduced and a relatively good fit for the combined sample, the acceptance criteria have actually tightened relative to the B-basis. Out of the nine equivalency samples, two of them are still considered 'failures' with respect to the equivalency criteria: a failure rate of ~ 20% for the data that was used to generate the acceptance criteria.

Multiply this result by 30 different properties and condition combinations to determine if a company should be certified as meeting the equivalency requirements and allowed to use the original design values. It leads to nearly all such samples getting flagged for one or more failures and requiring expert subjective engineering judgment to deal with the inherent uncertainty of determining whether a new sample is equal 'enough'.
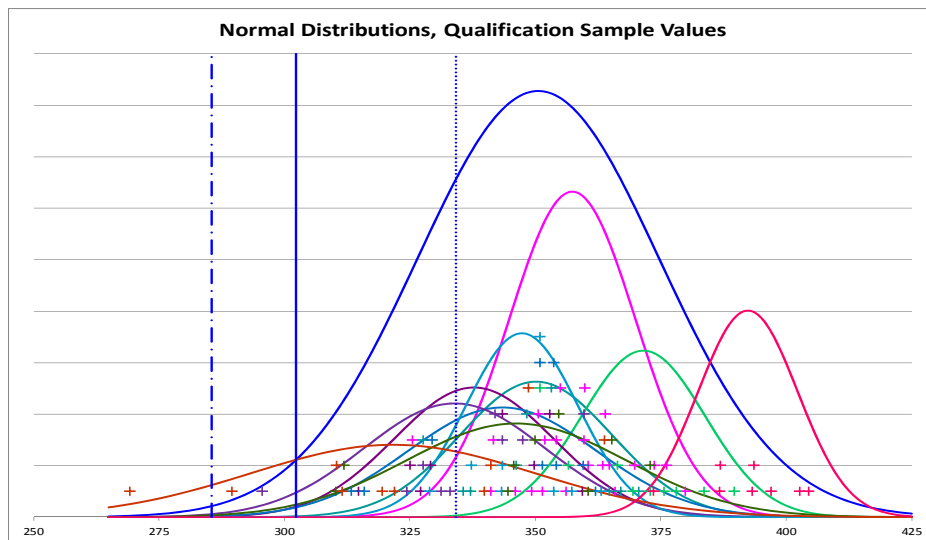

Figure 5. B-basis and Equivalency Criteria from Combined Samples

What is needed is a way to set both design values and acceptance criteria such that an equivalency sample can be evaluated according to objective criteria that have a high probability of identifying any discrepancies (consumer's risk) combined with realistic acceptance values such that the majority of users can expect to pass (producer's risk) without requiring subjective judgment with respect to the specific application of the material. To achieve this, we need an approach that sets both producer's and consumer's risk and base the decision on a single test that combines the test results of many different properties and conditions.

2. Proposed Approach: Flip the Null Hypothesis and Use Multivariate Analysis

The development of a system for setting acceptance criteria for both consumer's and producer's risk requires a fundamental reshaping of the null hypothesis that underlies acceptance testing. Specifically, it requires that we flip the null hypothesis. Instead of assuming materials equivalent and testing that assumption, as is the case currently, assume that the samples come from materials that lie within a specified area, termed the **equivalence** region (R_e). This is equivalent to assuming the samples are from similar populations rather than assuming they are from identical populations and test that assumption instead. This seemingly small change is significant in that we now have two user specified parameters, α_c and α_p. The producer's risk (α_p) is used to determine the acceptance region while α_c for the reversed null hypothesis specifies the consumer's risk.

The use of the multivariate approach allows the incorporation of knowledge about the correlation between properties rather than assuming them independent and identifies whether the combined test results for a new sample are sufficiently similar to judge equivalence with a single test statistic rather the testing each property individually. To do this, an estimate of the covariance matrix is required. This estimate can be supplied by the NCAMP database.

It reduces the subjectivity of the overall choice by replacing a decision based on the subjective weighting of many different test results with an objective decision based on the combined results of test for multiple properties and conditions.

There will remain an area between the acceptance region (definitely good) and the rejection region (definitely bad) that will require in-depth examination and subjective engineering judgment to determine the best course of action, but these subjective judgments will only be required for a fraction of the datasets tested rather than, as is currently the case, virtually all of them.

2.1 Setting the Producer's risk

Let α_p be the producer's risk. Define the acceptance region (R_a) for a material as the $1-\alpha_p$ confidence region around the mean using the covariance matrix computed from all available test results for that material. Thus, we can expect α_p% of the output to be rejected. We can easily alter the producer's risk by altering the acceptance region.

2.2 Setting the Consumer's risk

Let α_c be the consumer's risk. Once the **acceptance** region (R_a) has been defined, define the **equivalency** region (R_e) such that a sample produced from a CPM with mean property values that lie *within* R_a has a probability of $1-\alpha_c$ of producing a sample with property values lying *outside* R_e. This slightly larger region encompasses the acceptance region (R_a). A CPM with following all proper procedures has a probability of less than α_p of being erroneously rejected. At the same time, customers or inspectors can be confident that a CPM producing parts that lie outside of the **equivalency** region (R_e) has a probability of less than α_c of being accepted.

2.2 The Hypothesis Test for Generic Basis Values

To determine whether or not a new sample comes from an 'equivalent' population, set the null hypothesis as follows:

$$H_0 : M_1 \notin R_a$$
$$H_1 : M_1 \in R_a$$

(2)

2.2.1 Test Statistic for New Hypothesis Test

The following test statistic is used with to implement this hypothesis test[4]:

$$T = \frac{n_1 n_2}{n_1 + n_2} \left(\bar{X}_1 - \bar{X}_2 \right)' \Sigma^{-1} \left(\bar{X}_1 - \bar{X}_2 \right)$$

(3)

where

-   p is the number of material properties included in the analysis
-   n_1 is the number of units in the database.
-   n_2 is the number of units in the sample being evaluated for equivalency.
-   $\bar{X}_1$ is the p-dimensional mean vector computed from the database.
-   $\bar{X}_2$ is the p-dimensional mean vector of the sample being tested for equivalency .
-   Σ is the pxp covariance matrix computed from the database.

Under the revised null hypothesis, T has a non-central chi-squared distribution:

$$T \sim \chi_p^2 \left(\frac{n_1 n_2}{n_1 - n} d' \Sigma^{-1} d \right)$$

(4)

with d being any vector in the boundary of R_e which equates to the largest possible value of the non-centrality parameter (NCP) for the hypothesis test. The null hypothesis can be rejected and the new production facility concluded equivalent when $T < \alpha^{th}$ percentile of the non-central chi-squared distribution.

To determine R_e it is necessary to find a value for the non-centrality parameter NCP = $[n_1 n_2/(n_1+n_2)]d'\Sigma^{-1}d$ such that the critical value of the test statistic T is greater than the value of the corresponding $\chi_{p,0.90}^2$ distribution used to compute the acceptance ellipsoid. A table of these values is provided in the appendix A.

For example, if there are six tests in the grouping being evaluated and the NCP of a non-central chi-square distribution with 6 degrees of freedom is 17.871, then the critical value to reject the null is 10.645 > 10.6446 = $\chi_{p,0.90}^2$. This value of 17.871 is then used to define R_e as follows: $R_e=\{v \in \Re^P: (v-M)'\Sigma^{-1}(v-M)<17.871\}$ with M referring to the vector of the population means of the property values being tested.

With R_e so defined, the rejection region for H_0 will contain the acceptance region (R_a) previously established. Thus any sample vector that falls within the acceptance region has a probability of less than α of falling outside the equivalence region and a probability of at least $(1-\alpha)$ of lying within the equivalence region.

Figure 6 shows an example of the results of this approach[3]. The dotted green ellipse defines the limits of the generic acceptance area for Short Beam Strength (SBS) tests in the RTD and ETW conditions. The solid green ellipse defines the limits of the generic equivalence area. The generic acceptance and equivalency areas follow the general pattern of the data itself, with the correlation between the different properties being imbedded into the computation.

Notice that all but one of the NCAMP units are contained within the generic acceptance region. Further investigation revealed that that unit also had excessive variability, so it was deemed unacceptable. This graph also shows that the normal production process used by fabricators is not achieving the strength values reported in the Hexcel product data information for this property, although the different ETW condition results may be due to the temperature difference, 180°F for the Hexcel Product data versus 250°F for the NCAMP results.
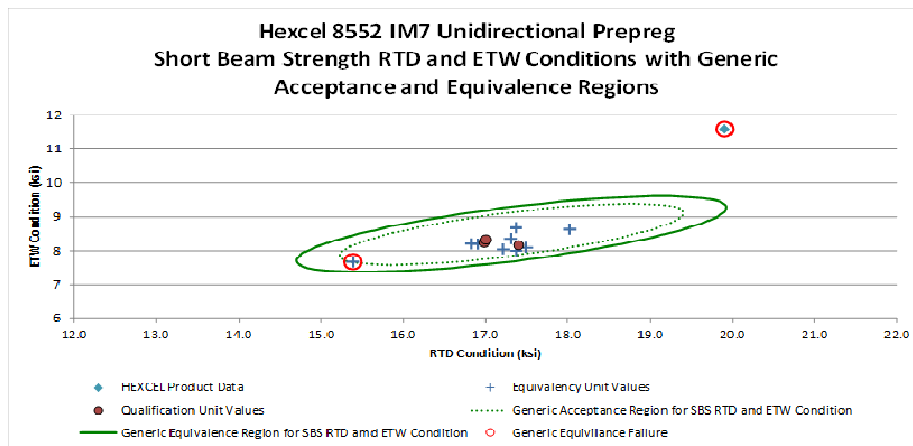

Figure 6. 2-D representation of Data, Acceptance Region and Equivalence Region

2.3 Defining the Design Values

B-basis values are defined as 95% lower confidence bound on the 10th percentile of a specified population of measurements. By this definition, if α_p is set to 10% of the population α_c set at 5%, then the lowest value of the equivalence region for each property will meet the definition of a B-basis value. A-basis values can be computed by setting α_c to 1%. However, there it should be noted that if the within site variability of a property is greater than the between site variability, to maintain conservative basis values, the design value computation should be done separately for that property using the within site variability.

3. Discussion

The main issue with equivalence testing under this paradigm is that it requires both R_a and R_e to be defined prior to setting engineering design values. Thus, the concept of equivalence testing must be in place from the beginning. Basis values are computed assuming a batch is produced with each property mean at the lowest possible boundary point of R_a for that property.

To set R_a and R_e, an estimate of the co-variance matrices for the properties being tested is required. The importance of the NCAMP database in developing this approach for prepreg composite materials cannot be overestimated. Prior to the NCAMP program, there was no publicly available information on the variability between manufacturing sites. Being able to include this variability in computations is key to developing the generic approach. If the co-variance matrices are consistent for different materials with a common form (such as Unitape or plain weave), then the co-variance matrix could be combined with predicted property mean values to develop generic values and acceptance criteria prior to any testing being performed.

Using this methodology, when the acceptance criteria are met, the probability that the manufacturer has produced acceptable material is the stated confidence level $(1 - \alpha)$. Using the traditional approach, a manufacturer that has failed the equivalency test has established with 95% confidence that they are NOT equivalent, but a manufacturer that has passed the test has no basis to compute the probability of actually being equivalent.

3.1 Advantages of this Approach

- A single objective measure of how similar a new material process is to the original material process.
- Categories developed with A- and B-basis values that apply to all materials included in the category.
- Ability to create a system allowing designers to input their material needs and run a search of the shared database to find all materials that would qualify for their application. (Note: This will require some additional funding.)
- Manufacturers will be motivated to include their material test results in our database because:
  o Any revisions to category basis values will retroactively apply to all materials previously included in the database.
  o A search engine to help designers find the material that best suits their needs will be limited to materials in our database.
  o Once a process has been shown to produce material that falls into the equivalence range, it needs to be monitored to make sure that the quality is maintained.
  o A single multivariate T2 control chart can be used to monitor production and make sure that the process continues to produce acceptable product. Control limits and spec limits can be set based on the characteristics routinely tested by the manufacturer.

3.1 Disadvantages of this Approach

Expense: It requires data from multiple composite part manufacturers. This is expensive in terms of the resources and requires cooperation between different users of composite materials. Thus, it is only achievable through projects like NCAMP that bring industry, government and academia together and provide a public database of results.

Lower Design values: Because the generic design values incorporate the variability between users of composite materials, they are lower than the computations resulting from a single site. For parts where weight is a significant constraint factor, design engineers may prefer to use basis values developed from the site that will be manufacturing their parts.

Acknowledgments

References

Clarkson, E. (2010). *Equivalence Testing for Mean Vectors of Multivariate Normal Populations*. Wichita, KS: Wichita State University, Ph.D. dissertation. Retrieved from http://soar.wichita.edu/handle/10057/3280

Clarkson, E. (2013). *A Comparison of Equivalence Criteria and Basis Values for Hexcel 8552 IM7 Unidirectional Tape computed from the NCAMP shared database* (pp. 19, 58). Retrieved from https://www.cmh17.org/Portals/0/Generic%20Basis%20Values%20for%20Hexcel%208552%20IM7%20Unidirectional%20Material%20May%2024%202013_FINAL.pdf

Composite Materials Handbook (CMH-17). (2012). Rev G, Volume 1. In *Warrendale, PA: SAE International*. Retrieved from http://store.sae.org/cmh-17/

Juran, J. M., & Gryna, F. M. (1988). *Juran's quality control handbook* (4th ed., pp. 25-29). New York etc.: McGraw-Hill book co.

Notes

Note 1. Details of the mathematics of this approach can be found in the above references by the same author. It has been submitted by the author to the "*Composite Materials Handbook (CMH-17)*" for inclusion in Rev H.

Appendix A

Non-Centrality Parameter Values for Generic Acceptance Criteria

DF	90% of Central Chi Squared	DF	If NCP is this:	Then the Critical Value for 95% confidence ellipsoid is
2	4.605170186	2	13.0240	4.60519
3	6.251388631	3	14.5727	6.251389
4	7.77944034	4	15.8280	7.779478
5	9.2363569	5	16.9086	9.236378
6	10.64464068	6	17.8708	10.64466
7	12.01703662	7	18.7461	12.01706
8	13.36156614	8	19.5542	13.36159
9	14.68365657	9	20.3083	14.68375
10	15.98717917	10	21.0180	15.98722
11	17.27500852	11	21.6901	17.27505
12	18.54934779	12	22.3300	18.54939

Copyrights