U.S. Department
of Transportation
Federal Aviation
Administration

Advisory Circular

Subject: Sample Size Considerations for Comparative Test and Analysis for Turbine Aircraft Engine and Auxiliary Power Unit (APU), Parts Manufacturer Approval (PMA), and Third-Party Repair Parts

Date: MM/DD/YY
Initiated By: AIR-625

AC No: 33-13
Change No: N/A

1. **PURPOSE.**

1.1 This advisory circular (AC) describes two approaches to define minimum sample sizes of products and articles for turbine aircraft engines and APUs that applicants for a parts manufacturer approval (PMA) may use to show compliance with the test reports and computations requirements of Title 14, Code of Federal Regulations (14 CFR) § 21.303(a)(4), using the comparative test and analysis approach to support showing the engine or APU comply with the applicable airworthiness requirements of 14 CFR Part 33 and Technical Standard Order (TSO) C77.

1.2 This guidance also supports the development of technical data needed for major repairs (as defined in Appendix A to Part 43). Parts fabrication and their implementation for repairs must be accomplished in accordance with 14 CFR Part 43.13, which requires that "[e]ach person maintaining or altering, or performing preventive maintenance, shall do that work in such a manner and use materials of such a quality, that the condition of the aircraft, airframe, aircraft engine, propeller or appliance worked on will be at least equal to its original or properly altered condition (with regard to aerodynamic function, structural strength, resistance to vibration and deterioration, and other qualities affecting airworthiness)." The method in this AC supports showing the engine or APU continues, after such maintenance, to comply with the applicable airworthiness requirements of 14 CFR Part 33, and TSO C77.

2. **APPLICABILITY.**

2.1 The guidance in this AC is directed to applicants requesting FAA approval for PMA, TC, STC, or repair or alteration of turbine engine and APU parts Federal Aviation

Administration (FAA) engine type certification engineers, and the Administrator's designees.

2.2 This is a guidance document. Its content is not legally binding in its own right and will not be relied upon by the Department as a separate basis for affirmative enforcement action or other administrative penalty. Conformity with the guidance document is voluntary only. Nonconformity will not affect rights and obligations under existing statutes and regulations.

2.3 This AC is intended only to provide information to the public regarding existing requirements under the law or agency policies. The FAA will consider other methods of demonstrating compliance that an applicant may elect to present. Terms such as "should," "may," and "must" are used only in the sense of ensuring applicability of this particular method of compliance when the acceptable method of compliance in this document is used. If the FAA becomes aware of circumstances in which following this AC would not result in compliance with the applicable regulations, the FAA may require additional substantiation as the basis for finding compliance.

2.4 The material in this AC does not change or create any additional regulatory requirements, nor does it authorize changes in, or permit deviations from, existing regulatory requirements.

3. RELATED READING MATERIAL.

3.1 Title 14, Code of Federal Regulations.

The following 14 CFR regulations are related to this AC. You can download the full text of these regulations from the Federal Register website at www.ecfr.gov.

- Part 33, *Airworthiness Standards: Aircraft Engines.*
- Section 21.303, *Application.*
- Section 43.13, *Performance rules (general).*

3.2 FAA Publications.

The following ACs are related to the guidance in this AC. The latest version of each AC referenced in this document is available on the FAA website at FAA Advisory Circulars and on the Dynamic Regulatory System.

- AC 20-107B, *Composite Aircraft Structure.*
- AC 21.303-4, *Application for Parts Manufacturer Approval Via Tests And Computations Or Identicality.*
- AC 33-8, *Guidance for Parts Manufacturer Approval of Turbine Engine and Auxiliary Power Unit Parts under Test and Computation.*
- AC 33-9, *Developing Data for Major Repairs of Turbine Engine Parts.*

- AC 33.83-1, *Comparative Method to Show Equivalent Vibratory Stresses and High Cycle Fatigue Capability for Parts Manufacturer Approval of Turbine Engine and Auxiliary Power Unit Parts.*
- AC 33.87-2, C*omparative Endurance Test Method to Show Durability for Parts Manufacturer Approval of Turbine Engine and Auxiliary Power Unit Parts.*
- AC 43-18, *Fabrication of Aircraft Parts by Maintenance Personnel.*

3.3 **FAA Orders.**

The following orders are related to the guidance in this AC. The latest version of each order referenced in this document is available on the Dynamic Regulatory System.

- Order 8110.42, *Parts Manufacturer Approval Procedures.*
- Order 8300.16, *Major Repair and Alteration Data Approval.*
- TSO-C77, *Gas Turbine Auxiliary Power Units.*

3.4 **Industry Publications.**

International Organization for Standardization. (1976). *ISO 2854:1976, Statistical interpretation of data — Techniques of estimation and tests relating to means and variances*. ISO Central Secretariat.

4. BACKGROUND.

The FAA has observed that PMA applicants use a variety of statistical principles and methods in their comparative test and analyses to show the proposed PMA part meets airworthiness standards by substantiating that it is at least equal to the original article approved under a type certificate in a particular area or attribute. However, the data often provided is based on statistical methods that are designed to detect statistical differences, not to show equivalency. The sample sizes submitted to the FAA as part of the application often have not provided an adequate demonstration of comparability between the proposed PMA or repair part population and the approved part population. Note that the term “approved part” is used throughout this AC to reference parts approved by type certification, supplemental type certification, or major repair or alteration. Often, proper use of statistical analysis requires larger sample sizes of the comparative populations than what was provided to the FAA to demonstrate equivalency.

5. STATISTICAL PRINCIPLES DISCUSSION.

5.1 Statistical equivalency tests begin with the null hypothesis, or assumption, that two parts are different by more than an allowed amount (refer to the discussion about parameter *d* in Appendix B.1.1.1 for more details) and seek evidence to reject that hypothesis. The equivalency tests are in contrast to the more common question “are these two populations significantly different?” that inherently assumes a baseline of equivalence.

5.2 In the past, some applicants have proposed analyses that used a simple t-test to compare sample means (i.e., averages) from two populations. Tests, such as the t-test, were used to find a difference, beginning with the null hypothesis that two populations are the same, and attempted to show a difference within a given level of confidence. This type of test puts the burden on finding a difference, and therefore, this approach is not statistically valid to prove that the two different populations are equally compliant to airworthiness standards. Many applicants are only familiar with this common test for differences and assume that if a t-test fails to detect a difference, the two populations must, therefore, be equivalent. In reality, a test result of "no statistically significant difference" means the current evidence is not strong enough to demonstrate that the two parts are different; this is not the same as showing that the two populations are equivalent.

5.3 Using a small number of samples (small sample sizes) in this regard may compromise the statistical method's ability to identify differences. This situation arises because a small sample size can make it difficult to identify statistical indications that a difference exists. Therefore, if applicants use a small sample size, the applicant may miss differences in the population because the applicant evaluated too few samples. This test would result in an incorrect conclusion that the populations are equivalent.

5.4 A potential consequence of using a small sample size for a statistical hypothesis test is a failure to find a difference when one exists, called beta error *(β)* or "Type II error." ($1 - \beta$) is the "Power" of finding a given difference. Saying a difference exists when a difference does not is called alpha error, *(α)* or "Type I error." ($1 - \alpha$) is the "Confidence" level in saying a difference exists. The FAA is concerned with beta errors when demonstrating equivalence for parts being approved under a major repair or under PMA by comparative test and analysis. The common practice of using the simple t-test, high-confidence levels, and small sample sizes leads to a high beta error; the inability to find real differences between the samples. Appendix B of this AC provides a detailed summary of how to address beta errors.

5.5 The finding of equivalency requires a different approach to that described in paragraphs 5.1 and 5.2. It requires a test to determine, with a given level of confidence, that the two populations are not different by more than an allowed amount established by parameter *d*. This AC describes two acceptable approaches to define a minimum sample size to be used as part of these methods of compliance, either through a practical method of evaluating approved parts, or through a more analytically rigorous and statistically accurate method of evaluating the approved parts (refer to Appendices A and B of this AC, respectively). Typically, there is a considerable difference in the necessary sample size between these two methods.

5.6 The approach in Appendix B for attempting to show equivalency may be appropriate when the PMA and approved part populations are normally distributed, believed to have the same variance, and when the PMA and approved part sample sizes are identical. Similar approaches with appropriate modifications may be used when these conditions are not met.

5.7 This AC does not provide a comprehensive technical background for each statistical concept involved in showing equivalency. Applicants may consult published statistical reference books or industry standards such as *ISO 2854:1976, Statistical Interpretation of Data — Techniques of Estimation and Tests Relating to Means and Variances* for related information. Similarly, unique materials such as composites may have additional or unique requirements; applicants may consider available FAA publications (such as AC 20-107B, *Composite Aircraft Structure*) associated with the part being proposed.

6. SUGGESTIONS FOR IMPROVING THIS AC.

If you have suggestions for improving this AC, you may use the Advisory Circular Feedback Form at the end of this AC.

Appendix A. Practical Approach to Define Sample Size

A.1 PRACTICAL APPROACH TO DEFINE SAMPLE SIZE.

A.1.1 This appendix describes a practical approach to define acceptable sample sizes. This approach typically allows applicants to use smaller sample sizes than the statistically derived sample size calculations described in Appendix B.

A.1.2 Using the sample size in this appendix starts with the fundamental assumption that both samples come from populations having the same mean. While results from testing this sample size can support that assumption, it does not rigorously demonstrate equivalence. Passing the test using this appendix's sample size only shows there is no reason to overturn the assumption that both samples have the same population mean. Failing the test using this sample size would indicate there is strong evidence that the assumption is wrong, and it should prompt additional investigation.

A.1.3 The applicant must provide qualitative and/or quantitative rationale for choosing the sample size in this appendix, which is typically smaller than the sample size calculated using Appendix B.

A.2 IDENTIFYING SAMPLES WITHIN POPULATIONS.

A.2.1 Minimum sample sizes depend on the particular parameter being evaluated. For measurements such as geometric dimensions and material composition and static properties such as elastic modulus and tensile strength, the FAA considers a minimum sample of 10 approved parts from each of three or more different lots to be acceptable. For example, 30 parts equally distributed over three or more lots are an acceptable minimum sample using this method. The intent is for the applicant to take measurements of multiple lots to observe a true range of population variation. Unique factors may bias any given lot within a single manufacturing run. Including several lots in sampling provides a better evaluation of population variability, thereby providing a more representative description of the approved part.

A.2.2 Larger sample sizes may provide a more accurate representation of the approved design, as well as the additional benefit of potentially expanding the acceptable tolerances for the proposed part.

A.2.3 Note that the guidance in this AC does not override guidance in other ACs that list a specific sample size to use for a particular subject. This AC provides a method of selecting a sample size if no other guidance material exists.

A.2.4 Inability to Meet Measured Tolerances.

A.2.4.1 It may be desirable to increase the number of approved part samples in an effort to increase the acceptable range for a given parameter. These additional samples should be random.

A.2.4.2 In the case of A.2.4.1, the additional approved parts would be added to the overall sample. Parts already included in the analysis may not be eliminated, unless either they clearly fail an outlier test (refer to paragraph A.2.6 for more details on outliers) or an entirely new proposed part sample population is provided. For example, if a new set of manufacturing controls are put in place when measuring proposed parts, replace the previously manufactured parts with these new parts.

A.2.5 **Rationale to Stay within Observed Measurements.**

A.2.5.1 Occasionally, part measurements seem to imply a population distribution that can be identified statistically. Even in those cases, it may be important to stay within observed measurements because the true approved part manufacturing tolerance may be "inspected in" to the approved part population. In this situation, the approved part manufacturing process produces a normal distribution, but the tails of the distribution lie outside the acceptable limits for production acceptance. As a result, the applicant must either scrap or rework the part to be within the acceptable limits.

A.2.5.2 As shown in figure A-1, data analysis may imply a normal distribution closely about the mean, but the tails of the distribution are actually truncated. Unless an adequate number of parts are sampled, it is unlikely that the tolerance truncation will be recognized. In addition, any attempt to identify statistical tolerance limits, as opposed to using measured tolerance limits, will likely result in a proposed part that is not within demonstrated experience, because it is not within the approved part tolerance.

A.2.5.3 For data that does not meet the assumption of being normally distributed, transformation into a normal distribution is often possible (e.g., square-root normal, log-normal, etc.). If the data represents other than a normal distribution, and cannot be transformed into a normal distribution, then the analysis should use the distributional parameters appropriate to the distribution type.

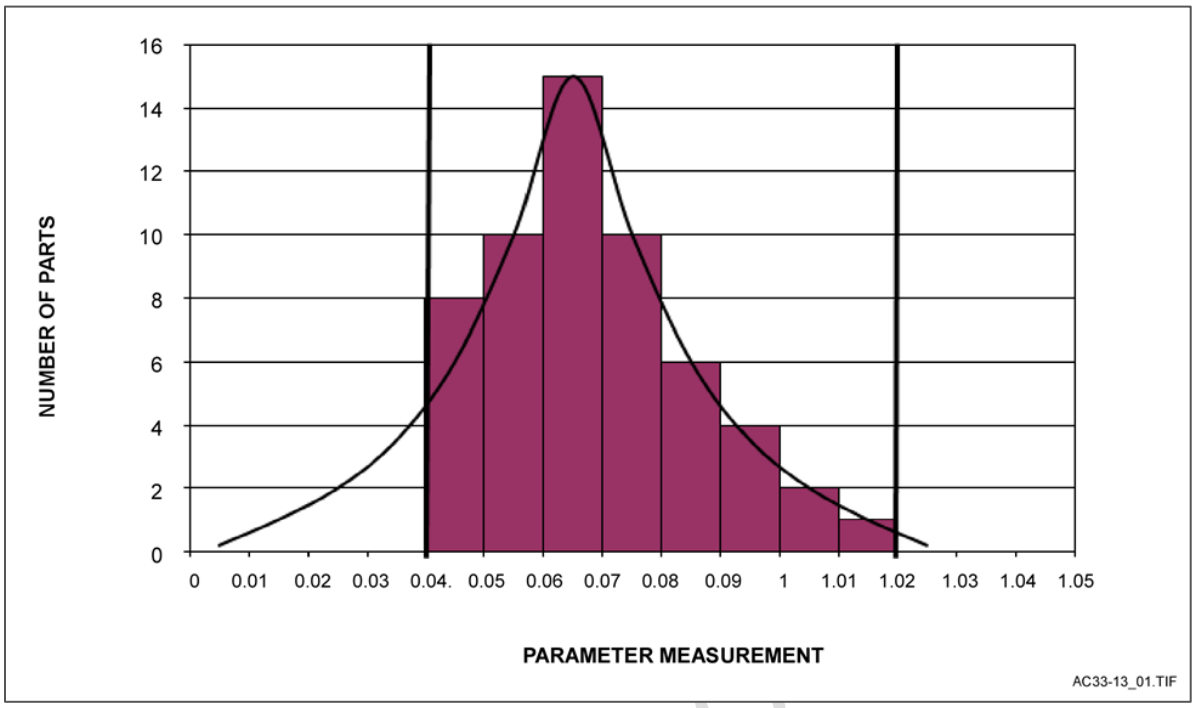

Figure A-1. Histogram of Parameter Measurements by Number of Parts

A.2.6 **Outliers**.

A.2.6.1 Applicants should examine the inspection results to ensure the data is representative of the sample populations. If either an approved or a proposed part could be an outlier, it may be identified using statistical tests. A typical outlier test involves determining the probability of an extreme measurement existing within the measured sample size. The statistical test may only be used once on any given sample, or data set, meaning that you may only eliminate the data point(s) identified by the initial outlier test. Do not eliminate the initial outlier(s) and then perform the statistical outlier test again for the remaining data points.

A.2.6.2 The reason(s) for any potential outlier must be investigated and explained before eliminating the data point(s). Applicants should retest outliers attributed to a faulty test procedure, preferably by use of an improved test procedure. Applicants must not eliminate extreme parts from the statistical population without proof that the part did not conform to design intent.

A.2.6.3 An extreme part that is not attributable to an identifiable cause, and was produced by the proposed production process, may be an indication of an unacceptable production process or part drawing. Such an outlier may require either improvement in the production process or changes to the proposed part design.

A.2.6.4 Often, the resulting data, including what may seem to be unusual data, reflects the true variation in the proposed part design or production process. Therefore, applicants may not remove potential outliers from the analysis without appropriate justification. This

justification must be in the FAA-approved certification test plan, and include, at a minimum, the following:

- A statistical test;
- A histogram of all the parts to document the outlier's status visually; and
- Explanations of factors unique to the outlier that provide a basis for not including the part in the rest of the population.

Appendix B. Statistical Method for Determining Sample Size

B.1 STATISTICALLY ACCURATE METHOD FOR DETERMINING SAMPLE SIZE.

This appendix provides a statistically accurate method for determining the sample size that is necessary to show two sample populations are statistically equivalent within an allowed amount. When applicants start with an agreed assumption, that the two sample populations are different by more than the allowed amount, they can use a statistically meaningful sample size towards showing a proposed part is sufficiently equivalent to an approved part. Before performing the statistical test, a sample size must be determined.

B.1.1 **Calculating Sample Size**.

B.1.1.1 Overview of Sample Size and Allowed Difference.

The following formula assumes that the standard deviation is known and equivalent between the two populations. The sample size for both proposed and approved parts, *N*, required to find an allowed difference *d* given alpha error, α, beta error, β, and standard deviation, *s*, is:

$$N = [(Z_{1-\alpha/2} + Z_{1-\beta})^2 * 2 * s^2] / d^2$$

The allowed difference, *d*, is an important parameter in calculating the sample size and is based on experience for what is an acceptable range for a given property of interest. For example, a proposed definition may be "within 3 KSI," or "within 0.0005 inch tolerance," or other relevant parameter. Discussions with the FAA are strongly recommended to reach consensus on the proposed allowed difference.

In the formula above, $Z_{1-\alpha/2}$ is the two-sided normal deviate for the given error level, and $Z_{1-\beta}$ is the one-sided normal deviate for the power level. These values are obtained from a standard normal table or may be calculated in Excel™ via the NORM.S.INV function. NORM.S.INV returns the inverse of the standard normal cumulative distribution for a probability corresponding to the normal distribution. The input to NORM.S.INV is $(1 - \beta)$ or $(1 - \alpha/2)$.

B.1.1.2 Example Calculation.

For a typical α of 0.20 (80% confidence) and β of 0.05 (95% power), the *z* values are 1.28 and 1.64, respectively. *N* should be rounded up to the next integer.

For α=0.2, β=0.05, *s*=8, and *d*=5, we get:

$$N = [(1.28 + 1.64)^2 * 2 * (8)^2] / (5)^2 = 43.8 \text{ rounded up to } 44$$

Note: *N* is the required number of both proposed and approved part samples, so the total number of parts for both combined is 88, 44 of each the approved and the proposed parts.

B.1.2 **Identifying samples within populations.**

As discussed in Appendix A, paragraph A.2.1, testing parts from more than one lot, melt, or other relevant production factors helps establish the variation introduced by these factors. The intent is for the applicant to take measurements of multiple lots to observe a true range of population variation. Unique factors may bias any given lot within a single manufacturing run. Including several lots in sampling provides a better evaluation of population variability, thereby providing a more representative description of the approved part. Therefore, the sample size, *N*, determined using the approach in Section B.1.1 should be divided among at least three lots.

B.2 **ASSUMPTION OF NORMALLY DISTRIBUTED DATA.**

The possibility of a non-normal distribution and truncation of approved part data should be considered and evaluated per the guidance in paragraph A.2.5 of this AC.

B.3 **OUTLIERS.**

Applicants must apply the outlier principles as defined in paragraph A.2.6 of this AC when using Appendix B to determine sample population size.