Test Population Selection From Weibull-Based, Monte Carlo Simulations of Fatigue Life

Brian L. Vlcek[1]
Georgia Southern University, Statesboro, Georgia, 30460–8045

and

Erwin V. Zaretsky[2] and Robert C. Hendricks[3]
NASA Glenn Research Center, Cleveland, Ohio 44135

Fatigue life is probabilistic and not deterministic. Experimentally establishing the fatigue life of materials, components, and systems is both time consuming and costly. As a result, conclusions regarding fatigue life are often inferred from a statistically insufficient number of physical tests. A proposed methodology for comparing life results as a function of variability due to Weibull parameters, variability between successive trials, and variability due to size of the experimental population is presented. Using Monte Carlo simulation of randomly selected lives from a large Weibull distribution, the variation in the L_{10} fatigue life of aluminum alloy AL6061 rotating rod fatigue tests was determined as a function of population size. These results were compared to the L_{10} fatigue lives of small (10 each) populations from AL2024, AL7075 and AL6061. For aluminum alloy AL6061, a simple algebraic relationship was established for the upper and lower L_{10} fatigue life limits as a function of the number of specimens failed. For most engineering applications where less than 30 percent variability can be tolerated in the maximum and minimum values, at least 30 to 35 test samples are necessary. The variability of test results based on small sample sizes can be greater than actual differences, if any, that exists between materials and can result in erroneous conclusions. The fatigue life of AL2024 is statistically longer than AL6061 and AL7075. However, there is no statistical difference between the fatigue lives of AL6061 and AL7075 even though AL7075 had a fatigue life 30 percent greater than AL6061.

I. Introduction

A common method of engineering design is to use deterministic equations for strength and life. That is, it is assumed that at loads and/or times below that calculated, no failure will occur. The calculated lives and load limits are usually coupled with a safety factor that is dictated by experience and design code requirements. Complicating the issue is that failure is extremely variable and dependent on materials, processing, and operating conditions.

For a finite life where it is assumed that failure can occur, reliability is dictated by the design application. If a failure does occur a criteria is that it does not cause injury and secondary damage. A statistical distribution is assumed about the calculated value of life or strength. From this method an acceptable life and/or load is determined at a predetermined probability of survival. Usually for critical applications that affect human safety and third party liability, these assumptions and calculations are re-examined and modified based on field experience, condition monitoring and continued inspection of critical components.

For most products that have an established design and application, product improvement is evolutionary. Incremental changes are made to improve product performance, life, reliability, and maintainability as well as to

---

[1] Associate Professor, Dept. of Mechanical & Electrical Engineering Tech., U.S. Highway 301 S., P.O. Box 08045, ASME Member.
[2] Chief Engineer, Structures and Materials Division, 21000 Brookpark Road/Mail Stop 23–3, ASME Life Fellow
[3] Senior Technologist, Research and Technology Directorate, 21000 Brookpark Road/Mail Stop 5–9, AIAA and ASME member.

decrease cost of manufacture. As the criticality (usually with respect to safety) of the designed component increases, the required probability of survivability for a given application increases.[1] Usually, for critical applications with competing failure modes, fatigue becomes the failure mode that limits the components life. That is, if the component does not fail from other causes, fatigue is the limiting mode of component failure (assuming that the component has been operated at stresses above which fatigue can occur, that is, above the material's fatigue limit).

Fatigue life is probabilistic and not deterministic. For example, it is not uncommon to find the ratio of longest to shortest fatigue life of bearings to be 20 to 1 or greater.[2] A sound knowledge of material fatigue life is thus essential to a good design. Experimentally establishing the fatigue life of materials, components, and systems is both time consuming and costly. As a result, conclusions regarding fatigue life are often inferred from a dangerously insufficient number of physical tests.

In 1939, Weibull[3–5] developed a method and an equation for statistically evaluating the fracture strength of materials based upon small population sizes. The method can be and has been applied to analyze, determine and predict the cumulative statistical distribution or fatigue failure or any other phenomenon or physical characteristic that manifests a statistical distribution.

Vlcek, Hendricks and Zaretsky[6] virtually tested 31 400 rolling-element bearings that were randomly assembled by Monte Carlo (random) number generation. The method of Johnson[7] was used to plot the virtual data on Weibull plots. The Monte Carlo results were compared with endurance data from 51 bearing sets comprising 5321 bearings. A simple algebraic relation was established for the upper and lower L_{10} life limits as a function of the number of bearings failed for any bearing geometry. There was a fifty percent (50 percent) probability that the resultant bearing life will be less that that calculated. The maximum and minimum variation between the bearing resultant life and the calculated life correlated with the 90-percent confidence limits for a Weibull slope of 1.5.

From the above described Monte Carlo analysis, Vlcek, Hendricks and Zaretsky[6] established rules to compare and distinguish tests of identical bearings either from two or more sources or made from different manufacturing sources and/or materials. These include variability due to Weibull parameters, variability between successive trials, and variability due to size of the experimental population with which the initial Weibull parameters were determined.

Zaretsky, et al.,[8] extended this method to compare two computational models to determine the fatigue life and reliability of a commercial turboprop gearbox with each other and with field data from 64 gearboxes. These models were (1) Monte Carlo simulation of randomly selected lives of individual bearings and gears comprising the system and (2) two- parameter Weibull distribution function for bearings and gears comprising the system using strict-series system reliability to combine the calculated individual component lives in the gearbox. A series of equations were empirically developed from the Monte Carlo simulation to determine the statistical variation in predicted life and Weibull slope as a function of the number of gearboxes failed. These results were successfully compared to field data from the 64 gearboxes.

In view of the aforementioned, the following are objectives of the work reported herein: (a) to determine a methodology for comparing life results as a function of variability due to size of the experimental population; (b) to determine the variation in the fatigue lives of aluminum alloy AL6061 rotating rod fatigue specimens and distribution parameters as a function of sample size using Monte Carlo simulation; and (c) compare the resultant fatigue life results from AL2024 and AL7075 to that of AL6061.

I. Apparatus, Specimens, And Procedure

Fatigue lives of notched, rotating 12.7-mm- (½-in.-) diameter by 101-mm- (4-in.-) long aluminum AL6061 rods were determined using a commercial rotational fatigue tester. This test is commonly referred to as either the Krause, Budd, or Vishay fatigue test. Figure 1 is a schematic of the primary components of the rotational fatigue tester.

The notches in the test sample were cut on a lathe using a tungsten carbide tool insert. No surface finish or polish was performed on the cut surface. The notches were located midway 50.8 mm (2 in.) along the length of the test sample (see Fig. 1 insert). The reduced diameters of the shafts at the notches were 6.35 mm (0.25 in.), and the walls of the notch were at a 60° angle.
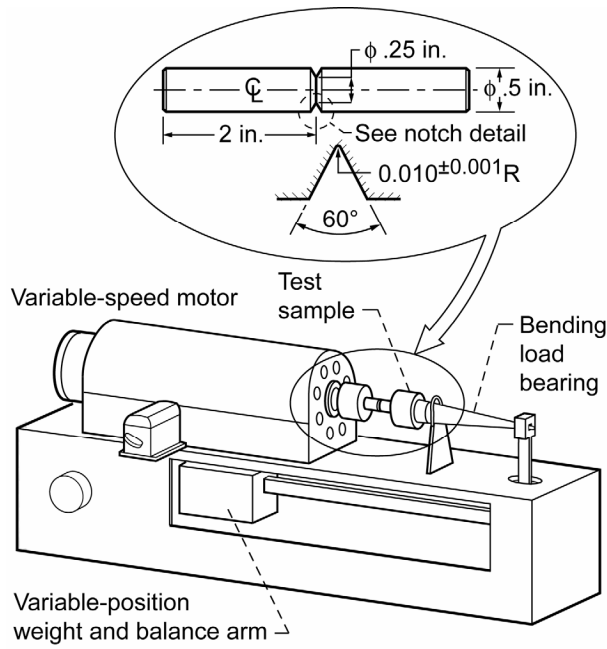
Figure 1. Schematic of Vishay rotating fatigue tester and test specimen geometry insert.

Figure 2. Generic Weibull plot where (Weibull) slope of line is *e*; probability of survival, S_β, is 36.8 percent at which $L = L_\beta$, or $L/L_\beta = 1$.

For each test, one end of the test sample was mounted in a 12.7-mm- (½-in.-) diameter rotating collet. The test sample extended from the collet as a horizontal cantilever (Fig. 1). The free end was then mounted in a collet that was part of a symmetric, tapered shaft extension. The collet was rotated by a variable-speed electric motor. A bearing from which a constant load was suspended was placed on the free end of the tapered shaft extension. As a point on the surface of the test specimen rotates about the axis, it experiences a sinusoidal load that is totally reversing each rotation of the shaft. By sliding a weight along a balance arm on the apparatus, the bending stress at the minimum sample cross section can be varied from 0 to 900 MPa (130.3 ksi). For this study, all tests were conducted at a bending stress of 157.2 MPa (22.8 ksi). All tests were performed at room temperature and 7500 rpm, and each test was conducted to failure; there were no suspensions or censoring of the data.

II. Statistical Analysis

A. Weibull Analysis

In 1939, W. Weibull[3–5] developed a method and equation for statistically evaluating the fracture strength of materials. He also applied the method and equation to fatigue data based upon small sample (population) sizes, where the two-parameter expression relating life, *L*, characteristic life, L_β and probability of survival, *S* is

$$\ln \ln \left(\frac{1}{S} \right) = e \ln \left(\frac{L}{L_\beta} \right) \quad \text{where } 0 < L < \infty; \, 0 < S < 1 \tag{1}$$

When plotting the ln ln [1/*S*] as the ordinate against the ln *L* as the abscissa, fatigue data are assumed to plot as a straight line (Fig. 2). The ordinate ln ln [1/*S*] is graduated in statistical percent of components failed or removed for cause as a function of ln *L*, the log of the time or cycles to failure (Fig. 3). The tangent of the line is designated the Weibull slope *e*, which is indicative of the shape of the cumulative distribution or the amount of scatter of the data. The method of using the Weibull distribution function for data analysis for determining component life and reliability was later developed and refined by L.G. Johnson.[7]

The Johnson-Weibull method has been in continual use by NASA for over 50 years to statistically analyze cumulative failure data such as occurs in bearings and gears for which an extensive data base now exists in the open literature.[9] The method allows for reasonable engineering estimates and comparisons of cumulative life distributions
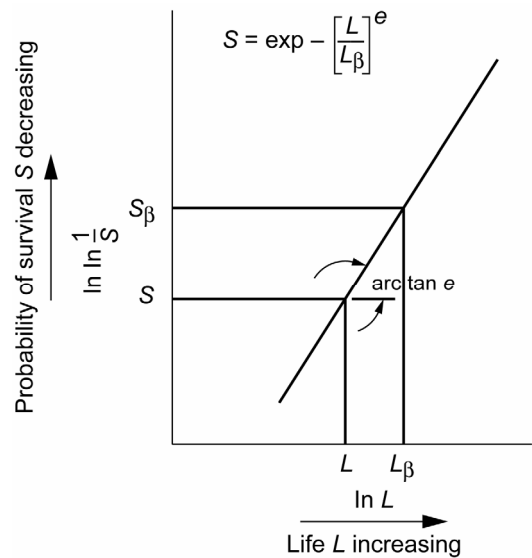
with small sample sizes or populations. The Weibull parameters were determined using a spreadsheet program that generates Weibull plots.

The Weibull slope of the resultant Weibull plot approximates the statistical distribution of the data. As an example, a Weibull slope of 1 approximates an exponential distribution. A Weibull slope of 2 approximates a Raleigh distribution. A Weibull slope of 3.57 approximates a Gaussian or normal distribution. The resulting values of life compare reasonably well with other statistical distributions such as log normal. However, the ease of use and consistency of results offers an advantage of the Johnson-Weibull method over these other distribution functions.

B. Monte Carlo Analysis

To determine the input parameters for the Monte Carlo simulation a total of 357 aluminum alloy AL6061 rotating test rods (Fig. 1) were tested to failure. A Weibull plot (Fig. 3) was computer generated for these data. The individual data points were omitted from Fig. 3 for purposes of clarity. From these results, the Weibull parameters were as follows: Weibull slope e = 2.878, L_β = 79 457 stress cycles, and L_{10} = 36 354 stress cycles. These values are summarized in Table 1 and will be referred to as the experimental "baseline" values for comparison purposes.

For purposes of Monte Carlo simulation, it was assumed that there were 1000 virtual aluminum AL6061 rod specimens contained in a virtual bin. According to the method of Johnson,[7] each rod specimen in the bin is assigned an order number (1, 2, 3,…1,000) which is also the order in which they would fail. The life of each component at the probability of survival corresponding to the randomly selected rank can be determined using Eq. (1), knowing the survivability (from the order and ranking), characteristic life, and Weibull slope from the experimental "baseline" values.

Using Monte Carlo techniques,[10] various-sized groups or populations were randomly assembled from the virtual bin. This group or population represented failure distributions of varying sizes; sizes of 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 35, 40, 45, 50, 75, 100, 125, 150, 175, and 200 were arbitrarily selected. At population sizes greater than 30, the interval between population size was increased. Each random number generated within a population corresponded to an order number of a component in the virtual bin, with which an associated failure life was determined. From each population size n, a distribution of associated lives was thus determined.

Next, this distribution of fatigue lives was sequentially ranked and then reordered according to the method of Johnson.[7] From the order number, a percent survivability was determined corresponding to each fatigue life using median ranks.

A curve fit of a Weibull plot with the independent variable of the natural logarithm of the life (ln L) and a dependent variable of the natural logarithm of the natural logarithm of the inverse of the probability of survival (ln ln(1/S)) was created. A straight line was fitted to these n data points or number of specimens failed using the method of least squares. The Weibull slope and the L_{10} life were determined from the resultant line.

For each value of n, the procedure was repeated 21 times to estimate variation between trials and to determine the maximum, minimum, and median values for each series of 21 trials. Ninety percent of the trials (19 trials) were bounded by the maximum and minimum L_{10} life values. This entire process in turn was repeated 10 times to establish statistical variation between repetitions. Thus, the lives and slopes reported herein are averaged values from these 10 repetitions. These results are shown in Fig. 4.

Best-fit curves obtained using the linear regression package of a commercial spreadsheet were fitted through the points for the minimum and maximum values shown in each of the plots of Fig. 4. The maximum and minimum variation in the distribution of the fatigue data as measured by the variation in the Weibull slope (Fig. 4 (b)) decreased as the number of test samples in a population increased.
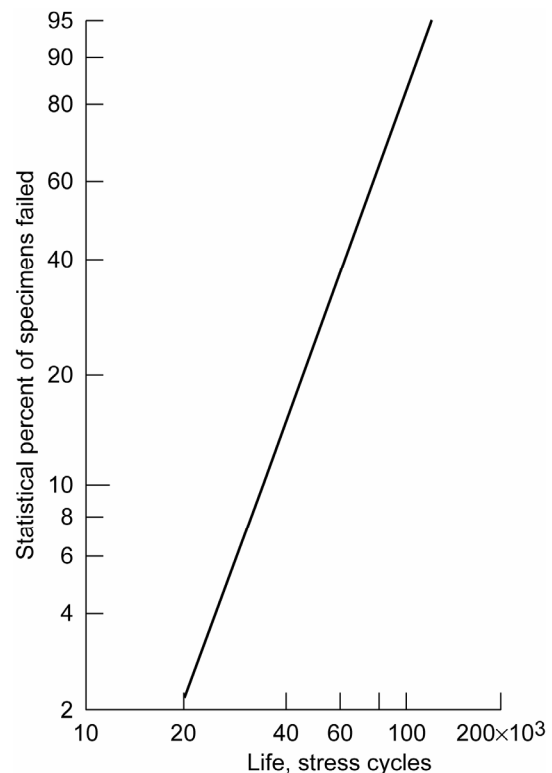

Figure 3. Weibull plot of fatigue life of aluminum alloy AL6061 rotating test rods. Number of failures, 357 (failure points omitted from Weibull plot for clarity); bending stress range, 0 to 157.2 MPa (22.8 ksi); speed, 7500 rpm; temperature, room.
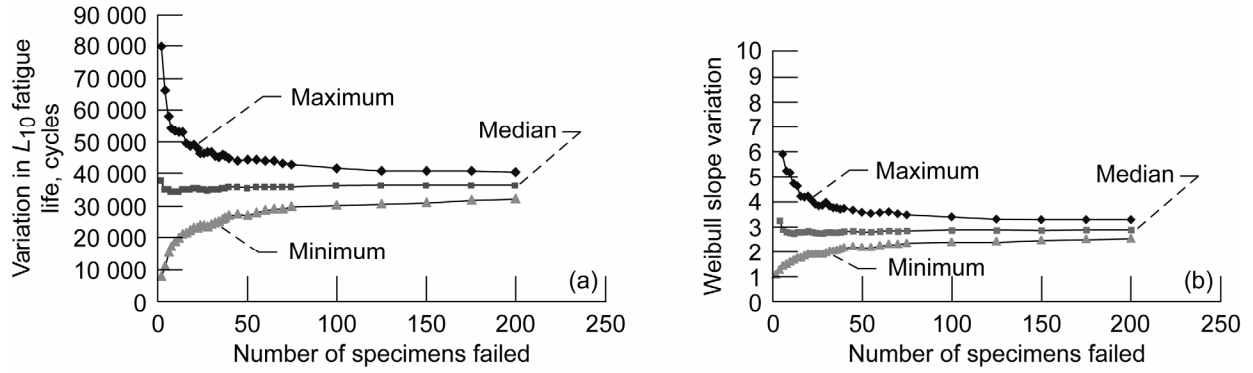
Figure 4. Maximum and minimum variation of L_{10} life and Weibull slope as a function of number of specimens failed for aluminum alloy AL6061 rotating test rods. (a) Fatigue life variation. (b) Weibull slope variation.

The bounds on the maximum and minimum L_{10} life values (Fig. 4(a)) were derived for each set of the respective curves as follows:

$$Maximum\ L_{10}\ Life = Baseline\ L_{10}\ Life\ (1 + 2n^{-0.53})\tag{2a}$$

$$Minimum\ L_{10}\ Life = Baseline\ L_{10}\ Life\ (1 - 1.25n^{-0.4})$$

$$where\ n > 2\tag{2b}$$

$$Minimum\ L_{10}\ Life \to 0\ where\ n \le 2\tag{2c}$$

where n is the number of specimens failed.

To facilitate comparison, the results of Fig. 4(a) are presented using percent variation in L_{10} life. The percent variation in L_{10} life is then expressed as Eq. (2):

$$Percent\ variation\ in\ L_{10}\ life = \frac{L_{10,\text{Monte Carlo}} - L_{10,\text{baseline}}}{L_{10,\text{baseline}}}x100\tag{3}$$

The normalizing value $L_{10,\text{baseline}}$ is 36,354 cycles. These results are plotted in Fig. 5.

C. Monte Carlo Verification

In order to verify the Monte Carlo simulation, the experimental data for the aluminum alloy AL6061 was ordered in the sequence in which they were tested, independent of the actual experimental lives. Using Weibull analysis, the L_{10} lives of 21 sequential sets of 3 each test specimens ($n = 3$) were determined for the first 63 specimens tested. Next, the L_{10} lives of 21 sequential sets of 5 each test specimens ($n = 5$) were determined for the next 105 specimens tested. Last, the L_{10} lives of 21 sequential sets of 9 each test specimens ($n = 9$) were determined for the next 189 specimens tested. The results were compared to the maximum and minimum values of Fig. 4(a) and Eqs. (2a) and (2b). The L_{10} lives of each of the subgroups fell on or within the maximum and minimum values of Fig. 4(a) as defined by Eq. (2). There was very good agreement between the maximum and minimum values of the respective L_{10} lives.
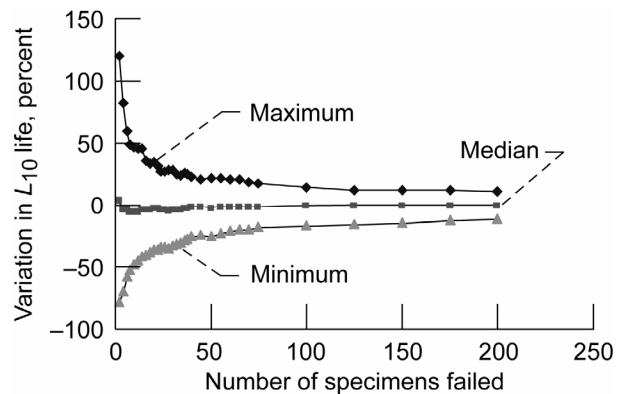

Figure 5. Maximum and minimum variation of L_{10} life as a function of number of specimens failed normalized to L_{10} life equal 36354 cycles. (Data from Fig 4(a)).

III. Results and Discussion

Incremental changes are made to improve product performance, life, reliability, and maintainability as well as to decrease cost of manufacture. As the criticality (usually with respect to safety) of the designed component increases, the required probability of survival for a given application increases. Usually, for critical applications with competing failure modes, fatigue becomes the failure mode that limits the components life and reliability. Fatigue life is probabilistic and not deterministic. Experimentally establishing the fatigue life of materials, components, and systems is both time consuming and costly. As a result, conclusions regarding fatigue life are often inferred from a statistically insufficient number of physical tests.

A. Comparing Life Results

From the results of the Monte Carlo simulation of the L_{10} life of aluminum alloy AL6061 (Figs. 4 and 5), there is an even (50 percent) probability that the resultant L_{10} test life for any randomly selected test group of the same material will be greater or less than its established baseline life. As the number of specimens in a test group increases, the magnitude of the variation in life decreases. From these observations rules can be established to compare and distinguish tests of identical or different aluminum alloys either from one another, two or more sources, or made from different processing methods. The following rules are suggested:

1. If the L_{10} life of the material is between the Maximum and Minimum L_{10} life variations, there can be no conclusion that there is a significant difference between the new material and the baseline material regardless of the ratio of the L_{10} lives. The new material is acceptable for its intended application (Fig. 6(a)).
2. If the L_{10} life of the new material is less than the minimum variation, the new material is not acceptable for its intended application (Fig. 6(b)).
3. If the L_{10} life of the new material is greater than the maximum variation, there exists a significant difference between the materials with the new material being superior for its intended application. (Fig. 6(c)).
4. If the L_{10} life of the new material is greater than the baseline life (zero variation) but less than the maximum variation, there can be no conclusion that there is a significant difference between the new material and the baseline material regardless of the ratio of the L_{10} lives. The new material is acceptable for its intended application (Fig. 6(d)).
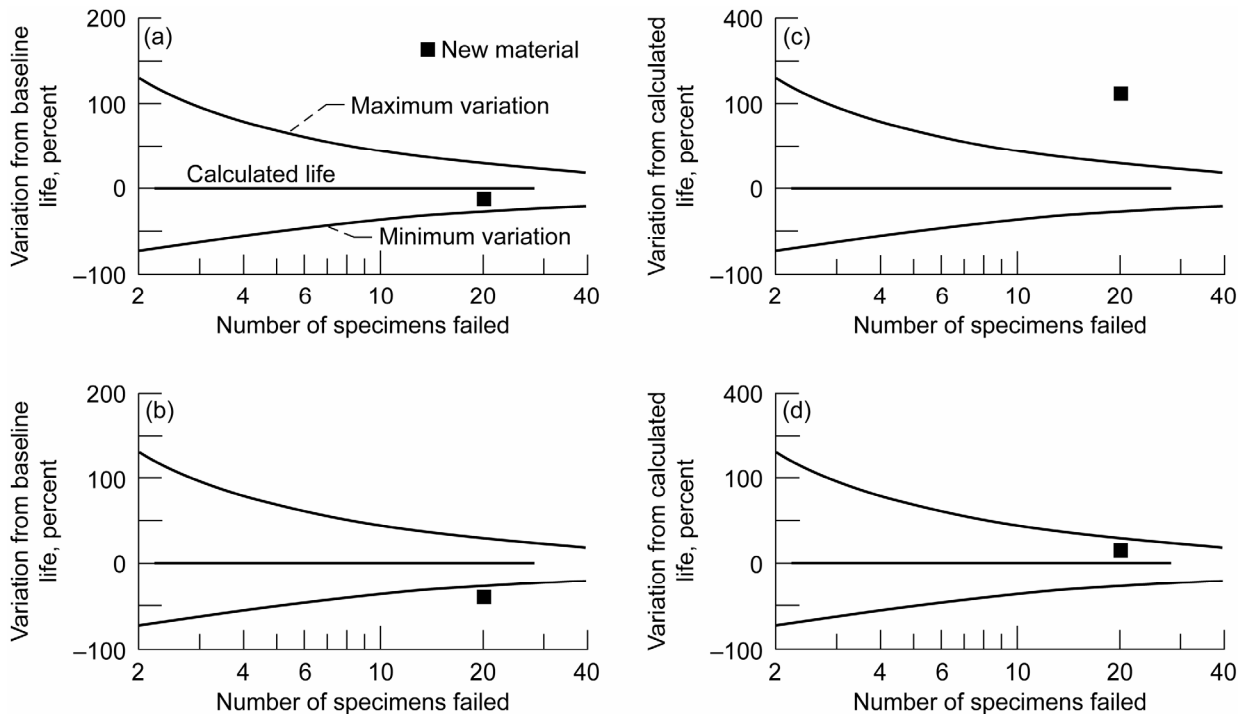

Figure 6. Rules for comparing material specimen fatigue life results to AL6061 (based on Fig. (5)). (a) New material is acceptable. (b) New material is not acceptable. (c) New material is superior to baseline material. (d) New material is acceptable.

B. Experimental Life Results

With a reasonable understanding of how the Monte Carlo simulations responded to variations in inputs, it is possible to compare fatigue data for aluminums AL7075 and AL2024 with that of aluminum AL6061. Ten (10) each rotating notched rod fatigue tests were conducted for AL7075 and AL2024, respectively. Baseline verification tests were conducted with 10 additional specimens of aluminum AL6061. These tests were conducted at a bending stress of 157.2 MPa (22.8 ksi), room temperature and a speed of 7500 rpm. Each test was conducted to failure; there were no suspensions or censoring of the data. Using Eq. (3) the resultant L_{10} lives of the 3 materials were compared to the L_{10} life from the baseline tests of the 357 test samples of AL6061 (L_{10} = 36 354 cycles). The results are summarized in Table 1.

Referring only to the resultant L_{10} lives based upon 10 test specimens for each aluminum alloy, the materials would be intuitively ranked in the order of their relative L_{10} lives. That is, AL2024 first, AL7075 second, and AL6061 third. There is a strong suggestion that the AL7075 has a 30 percent higher life than AL6061. However, referring to the baseline tests of AL6061 and to the Monte Carlo results (Fig. 4), and using the criteria of Fig. 6 to compare results, it can be reasonably concluded that AL2024 is statistically superior to the other 2 alloys. However, there is no statistical difference between AL6061 and AL7075. The variability of test results based on small sample sizes can be greater than actual differences, if any, that exists between materials and can result in erroneous conclusions. It is worth noting that the L_{10} life of even the baseline material (AL6061) differs between that determined for only 10 trials (Table 1) and the baseline of 357 trials by 35 percent.

Table 1. Comparison of Rotating Rod Fatigue Lives of Three Aluminum Alloys
[Test conditions: Bending stress, 157.2 MPa (22.8 ksi); speed, 7500 rpm; temperature, room.]

Aluminum alloy	Experimental Weibull slope	Characteristic life, L_β (cycles)	L_{10} Life (cycles)	Comparison to baseline L_{10} life (Eq. (3)), percent
[a]AL6061	2.55	65,249	26,987	−25.8
[a]AL7075	2.89	76,226	35,029	−3.6
[a]AL2024	6.22	224,304	156,207	331
[b]Baseline AL6061	2.878	79,457	36,354	-------

[a]Ten (10) test specimens failed.
[b]From Weibull plot of Fig. 3, 357 test specimens failed.

C. Minimum Population Size and Probabilistic Variability

Engineers and designers tend to consider fatigue lives as absolutes, when actually there is probabilistic variability associated with reported values. There is a significant range between the maximum and minimum 90-percent bounding fatigue lives for small population sizes of 2 to 10, where the percent variation exceeds 100-percent (Fig. 5). For population sizes ranging from 14 to 25, the range between the maximum and minimum L_{10} life continues to decrease, but nonetheless has differences of 75 to 100 percent variation. It is not until 30 to 100 samples are tested that these differences in minimum life begin to level at 40 to 60 percent variation.

The minimum life locus is most likely used in conservative designs and was thus focused on herein. Again, there is a significant variation between the simulated and calculated minimum lives for small population sizes of 2 to 6. This difference is nearly 100 percent. The difference decreases significantly with increasing sample size. Over the range of 10 to 28 specimens, the simulated minimum life differs from the known calculated life by about 50 percent. At approximately 30 specimens, the difference is less than 30-percent variation, and at 70 specimens the difference is less than a 20-percent variation. It should be noted that even for 200 test specimens, there is still a range observed between the upper and lower 90-percent bounding curves. This range is within 15 percent.

Vlcek, Hendricks, and Zaretsky[6] performed a similar Monte Carlo simulation on a bearing system comprising three components. While the life of multiple components had to be taken into account, similar trends as those evident in Figs. 4 and 5 were observed. Similar observations were also made by Zaretsky, Savage, Lewicki, and Vlcek,[8] where a Monte Carlo simulation was used to determine the 90-percent bounds on the L_{10} life of a 21 component gear box.

The number of fatigue tests performed is dependent upon the probabilistic variability that can be tolerated. Figure 5 helps establish the range in variability associated with fatigue life and test population size. For most engineering applications where less than 30 percent variability can be tolerated in the maximum and minimum values, at least 30 to 35 test samples are necessary. Design decisions can be made based upon smaller population sizes, as long as the inherent increase in variability in the life value and the associated risk are recognized and accounted for.

IV.   Summary of Results

Fatigue life is probabilistic and not deterministic. Experimentally establishing the fatigue life of materials, components, and systems is both time consuming and costly. As a result, conclusions regarding fatigue life are often inferred from a statistically insufficient number of physical tests. A proposed methodology for comparing life results as a function of variability due to Weibull parameters, variability between successive trials, and variability due to size of the experimental population is presented. Using Monte Carlo simulation of randomly selected lives from a large Weibull distribution, the variation in the L_{10} fatigue life of aluminum alloy AL6061 rotating rod fatigue tests was determined as a function of population size. These results were compared to the L_{10} fatigue lives of small (10 each) populations from AL2024, AL7075 and AL6061. The following results were obtained:

1. For aluminum alloy AL6061, a simple algebraic relationship was established for the upper and lower L_{10} fatigue life limits as a function of the number of specimens failed.
2. For most engineering applications where less than 30 percent variability can be tolerated between the maximum and minimum L_{10} fatigue life values, at least 30 to 35 test samples are necessary. Design decisions can be made based upon smaller population sizes, as long as the inherent increase in variability in the life value and the associated risk are recognized and accounted for.
3. The variability of test results based on small sample sizes can be greater than actual differences, if any, that exists between materials and can result in erroneous conclusions. Based on a test population of 10 specimens each, the fatigue life of AL2024 is statistically longer than AL6061 and AL7075. However, there is no statistical difference between the fatigue lives of AL6061 and AL7075 even though AL7075 had a fatigue life 30 percent greater than AL6061.
4. The maximum and minimum L_{10} life limits determined experimentally for Aluminum alloy AL6061 rotating rod fatigue specimens for test population sizes of 3, 5, and 9 correlated with those obtained from the Monte Carlo simulation.

V.   References

[1]Zaretsky, E. V., "Design for Life, Plan for Death," Machine Design, 66(15), Aug. 8, 1994, pp. 55–59.

[2]Zaretsky, E. V., STLE Life Factors for Rolling Bearings, STLE SP–34, STLE, Park Ridge, IL, 1992.

[3]Weibull, W. (1939), "A Statistical Theory of the Strength of Materials," Ingenioersvestenskapsakademiens. Handl., 151.

[4]Weibull, W. (1939), "The Phenomenon of Rupture of Solids," Ingeniorsvetenkapsakademiens. Handl., 153.

[5]Weibull, W. (1951), "A Statistical Distribution Function of Wide Applicability," J. Appl. Mech. Trans. ASME, 18(3), pp. 293–297.

[6]Vlcek, B. L., Hendricks, R. C., and Zaretsky, E. V., (2003), "Determination of Rolling-Element Fatigue Life from Computer Generated Bearing Tests," STLE Tribology Trans., 46 (4), pp. 479–493.

[7]Johnson, L.G., The Statistical Treatment of Fatigue Experiments, Elsevier Publishing Co., Amsterdam, The Netherlands, 1964.

[8]Zaretsky, E. V., Savage, M., Lewicki, D., and Vlcek, B., "Determination of Turboprop Gearbox System Fatigue Life and Reliability," STLE Tribology Trans., 50 (4) pp. 507–516, Oct. 2007.

[9]Zaretsky, E. V., Tribology for Aerospace Applications, STLE SP–37, STLE, Park Ridge, IL, 1997.

[10]Mooney, C. Z., Monte Carlo Simulation, Sage Publications, Thousand Oaks, CA, 1997.