Designation: F2118 – 14

Standard Test Method for Constant Amplitude of Force Controlled Fatigue Testing of Acrylic Bone Cement Materials[1]

This standard is issued under the fixed designation F2118; 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.

1. Scope

1.1 This test method describes test procedures for evaluating the constant amplitude, uniaxial, tension-compression uniform fatigue performance of acrylic bone cement materials.

1.2 This test method is relevant to orthopedic bone cements based on acrylic resins, as specified in Specification F451 and ISO 16402. The procedures in this test method may or may not apply to other surgical cement materials.

1.3 It is not the intention of this test method to define levels of performance of these materials. It is not the intention of this test method to directly simulate the clinical use of these materials, but rather to allow for comparison between acrylic bone cements to evaluate fatigue behavior under specified conditions.

1.4 A rationale is given in Appendix X2.

1.5 The values stated in SI units are to be regarded as standard. No other units of measurement are included in this standard.

1.6 *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 and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E466 Practice for Conducting Force Controlled Constant Amplitude Axial Fatigue Tests of Metallic Materials

E467 Practice for Verification of Constant Amplitude Dynamic Forces in an Axial Fatigue Testing System

E1823 Terminology Relating to Fatigue and Fracture Testing

F451 Specification for Acrylic Bone Cement

2.2 *ISO Standard:*

ISO 16402 Flexural Fatigue Testing of Acrylic Resin Cements Used in Orthopedics[3]

3. Terminology

3.1 Unless otherwise given, the definitions for fatigue terminology given in Terminology E1823 will be used.

3.2 *Definitions:*

3.2.1 *mean fatigue life at N cycles*—the average number of cycles to failure at the specified load level. For the purposes of this test method, the fatigue life will be determined at 5 million load cycles. A rationale for this is provided in X2.4.

3.2.2 *median fatigue life at a given stress level*—the number of cycles to failure at which 50 % of the tested samples failed at the specified stress level.

3.2.3 *runout*—a predetermined number of cycles at which the testing on a particular specimen will be stopped, and no further testing on that specimen will be performed. For the purposes of this test method, the runout will be 5 million load cycles.

3.2.4 *specimen failure*—the condition at which the specimen completely breaks or is damaged to such an extent that the load frame is no longer able to apply the intended stress within the required limits.

3.2.5 *stress level*—the value of stress at which a series of duplicate tests are performed. For the purposes of this test method, the stress level is reported as the maximum stress applied to the specimen.

4. Summary of Test Method

4.1 Uniform cylindrical reduced gage section test specimens are manufactured from acrylic bone cement and mounted in a uniaxial fatigue frame. The specimen is subjected to fully reversed tensile and compressive loading in a sinusoidal cyclic manner at a specified frequency in phosphate buffered saline (PBS). The fatigue loading is continued until the specimen fails or a predetermined number of cycles (run-out limit) is reached.

[1] This test method is under the jurisdiction of ASTM Committee F04 on Medical and Surgical Materials and Devices and is the direct responsibility of Subcommittee F04.15 on Material Test Methods.

Current edition approved Dec. 1, 2014. Published February 2015. Originally approved in 2001. Last previous edition approved in 2010 as F2118-10. DOI:10.1520/F2118-14.

[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.

5. Significance and Use

5.1 This test method describes a uniaxial, constant amplitude, fully reversed fatigue test to characterize the fatigue performance of a uniform cylindrical waisted specimen manufactured from acrylic bone cement.

5.2 This test method considers two approaches to evaluating the fatigue performance of bone cement:

5.2.1 Testing is conducted at three stress levels to characterize the general fatigue behavior of a cement over a range of stresses. The stress level and resultant cycles to failure of the specimens can plotted on an *S-N* diagram.

5.2.2 Another approach is to determine the fatigue life of a particular cement. The fatigue life for orthopaedic bone cement is to be determined up to 5 million (5×10^6) cycles.

5.3 This test method does not define or suggest required levels of performance of bone cement. This fatigue test method is not intended to represent the clinical use of orthopaedic bone cement, but rather to characterize the material using standard and well-established methods. The user is cautioned to consider the appropriateness of this test method in view of the material being tested and its potential application.

5.4 It is widely reported that multiple clinical factors affect the fatigue performance of orthopaedic bone cement; however, the actual mechanisms involves multiple factors. Clinical factors which may affect the performance of bone cement include: temperature and humidity, mixing method, time of application, surgical technique, bone preparation, implant design, anatomical site, and patient factors, among others. This test method does not specifically address all of these clinical factors. The test method can be used to compare different acrylic bone cement formulations and products and different mixing methods and environments (that is, mixing temperature, vacuum, centrifugation, and so forth).

6. Apparatus

6.1 *Uniaxial Load Frame*—A testing machine capable of applying cyclic sinusoidal tensile and compressive loads.

6.1.1 The crossheads of the load frame shall be aligned such that the alignment meets the requirements of section 8.2 of Practice E466. The alignment should be checked at both the maximum tensile and minimum compressive load to be applied during the course of a test program.

6.2 *Cycle Counter*—A device capable of counting the number of loading cycles applied to a specimen during the course of a fatigue test.

6.3 *Load Cell*—A load cell capable of measuring dynamic tensile and compressive loads in accordance with Practice E467.

6.4 *Limit*—A device capable of detecting when a test parameter (for example, load magnitude, actuator displacement, DC error, and so forth) reaches a limiting value, at which time the test is stopped and the current cycle count recorded.

6.5 *Environmental Chamber*—A chamber designed to immerse the fatigue specimen completely in a solution. The chamber should have provisions for maintaining a constant temperature to an accuracy of ±2°C.

7. Test Specimen

7.1 Test specimens shall be fabricated from cement that is representative of the final product with regard to materials, manufacturing processes, sterilization, and packaging. Certain sterilization methods have been shown to have an effect on fatigue performance (for example, gamma sterilization of the powder). Any deviations of the test cement from the clinically used product must be reported.

7.2 Cylindrical reduced gage section test specimens with a straight 5-mm diameter by 10-mm-long gage section shall be used. The diameter of the specimen ends shall be substantially greater than the gage diameter to ensure that fracture occurs in the gage section. A smooth surface of the test specimen in the radius or taper between the specimen ends and gage section is essential to reduce variation in reported fatigue life. Suggested specimen dimensions are provided in Fig. 1.

8. Specimen Preparation

8.1 *Cement Mixing:*

8.1.1 Store the liquid and powder portions of the cement according to the manufacturer's instructions before mixing.

8.1.2 Allow the mixing equipment to equilibrate to room temperature before mixing. Record the room temperature at the onset of mixing.

8.1.3 Mix the powder and liquid components according to the manufacturer's instructions and begin recording the time using a stopwatch when the liquid and powder are initially mixed. Report any deviations from the manufacturer's storage and mixing recommendations.

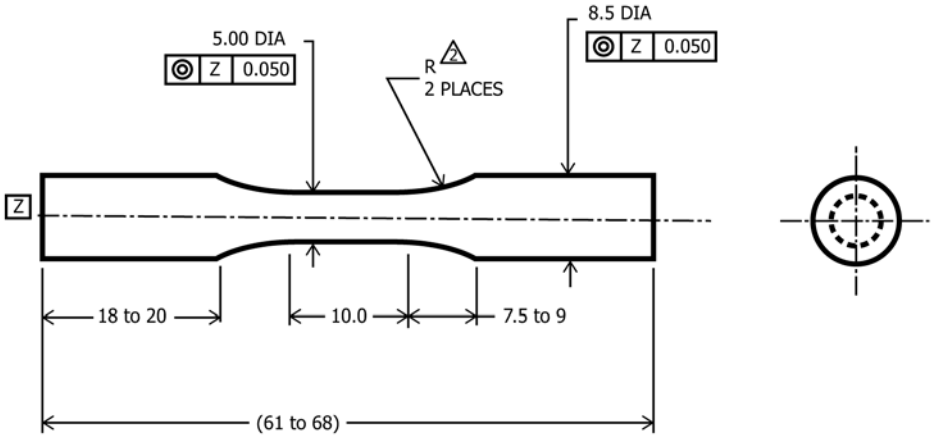
8.1.4 Report the mixing method and any equipment used. The method used for mixing the cement may affect its fatigue behavior. See X2.13 for further information.

8.2 *Specimen Fabrication*—The cylindrical reduced gage section test specimens are fabricated using the following method:

8.2.1 *Direct Molding:*

8.2.1.1 Inject the mixed cement into a specimen mold during the dough phase as determined by Specification F451 (manufactured from silicone material, see Appendix X3 (suggested specimen molding method)) with an internal cavity which has the same dimensions as the final cement test specimen. Record the method of cement insertion into the mold (that is, syringe injected). A 150 mL syringe with an inner diameter of 38 mm and a nozzle tip diameter of 10 mm should be considered for use. The mold should be placed on a flat surface. The cement injection should be performed from top to bottom in direction allowing the cement to flow down axially to the bottom . The bottom of the mold is placed on a flat surface as the bone cement is being injected into the mold unixaially from the top down. If air is entrapped and leads to resistance to injection, the mold should be rocked back and forth to release trapped air from the bottom of the mold. This will allow for air to escape from the bottom of the mold. (See X3.6 for standard operating procedure for making bone cement specimens.)

8.2.1.2 Place the mold in a container of phosphate buffered saline (PBS). The PBS solution should be maintained at 37 ±



1. All dimensions in mm
2. Radius to blend smoothly with gage section
3. Tolerances: X. = ± 1.0, X.X = ± 0.5, X.XX = ± 0.1

FIG. 1 Specimen Dimensions

2°C. After at least 1 h in the PBS bath, the specimens may be removed from the mold. Appendix X3 describes a suggested procedure for molding cement specimens.

8.3 *Specimen Examination:*

8.3.1 Visually examine specimens for surface defects. Surface defects in the gage or transition sections (radii) shall be rejected from testing and discarded. A surface defect is defined as a surface discontinuity greater than 250 mm in major diameter. All specimens should be photographed to document surface finish prior to testing. In addition, the specimens' straightness should be compared to the metal positive blank to ensure that the specimen is will not product bending moments during the uniaxial fatigue testing. Straightness can be assessed by rolling the specimens and determining if there is a visible wobble as compared to the straight metallic blank used to make the mold. Specimens with surface defects or deemed not to be straight shall be rejected from testing and discarded. The total number of specimens rejected divided by the total number of specimens manufactured (rejection rate) shall be reported. A rationale for these rejection criteria is provided in X2.11.

8.4 *Specimen Finishing*—If necessary, lightly polish the gage length of the specimens with 600-grit abrasive paper in the longitudinal direction until the surface is free of machining and/or mold marks. It should be noted that molds can wear over time as they are used, and a visual inspection of the surface roughness of each specimen should be done to ensure smoothness. New molds should be made when the smoothness can no longer be achieved with light polish.

8.5 *Specimen Measurement*—Measure the diameter of the specimens at a minimum of three places along the gage length of each specimen. The average of these measurements shall be used as the specimen's gage diameter for calculation of the required load.

8.6 *Specimen Conditioning:*

8.6.1 Place the test specimens in PBS which is maintained at a temperature of 37 ± 2°C.

8.6.2 Maintain the specimens in the PBS solution for a minimum of 7 days. The cement specimens shall be maintained in the PBS solution for 7 to 60 days. The specimens shall be continually immersed in the test solution so that they do not dry out. Distilled water shall be added to the soaking chamber during the soaking period to make up for evaporation loss. Each specimen should be soaked up to the time immediately before its being mounted on the load frame. See X2.5 for further information.

9. Fatigue Test Procedures

9.1 Mount one specimen at a time in a test frame test such that a uniaxial load is applied. Collets, Jacob's chucks, or pressurized grips should be used to firmly grip the specimen at each end. Ensure the longitudinal centerline of the test specimen is aligned with test machine loading axis such that bending moments are minimized. Testing of multiple specimens on the same fixture in parallel or series shall not be performed as this complicated and changes the stress state in the individual specimens when cracks initiate and propagate through the specimen occurs, effectively changing the modulus of each individual specimen being tested.

9.2 Mount an environmental chamber on the load frame and fill with fresh PBS solution immediately after the specimen is mounted to keep the specimen from drying out. The chamber should be filled to a level such that the entire specimen is immersed. Distilled water shall be added to the test chamber during the course of a test to make up for any evaporation loss. The temperature controller should be programmed and activated to heat the test solution to 37°C, and then maintain that temperature within ±2°C. Fatigue testing should not begin until at least ½ h after the solution temperature has reached 37°C to ensure equilibration.

9.3 Program the test frame controller to apply a fully reversed sinusoidal cyclic waveform at a constant frequency. When testing at frequencies above 5 Hz, the user should verify that, for the formulation being tested, the chosen frequency has a negligible effect on the test results. See X2.6 for further information.

9.4 Program the test frame controller to apply the desired maximum stress level and a stress ratio of $R = -1$, indicating fully reversed loading. A rationale for using fully reversed loading is provided in X2.10. The load shall be calculated by multiplying the desired stress by the specimen's cross-section area, based on each specimen's gage diameter as determined in 8.5.

9.4.1 Report the stress level to the nearest 0.5 MPa.

9.4.2 Determine the appropriate data acquisition frequency to adequately document the loads and displacements achieved during the testing.

9.4.3 When developing an *S-N* curve (see 10.1), it is recommended that testing be conducted at the following maximum stress levels: 15, 12.5, and 10 MPa. Other stress levels may also be appropriate for orthopedic applications such as the hip and knee. However, stress levels of 5, 7, and 9 MPa should be considered for spinal applications in vertebroplasty and kyphoplasty. See X2.7 for a rationale regarding the selection of the recommended stress levels.

9.5 *Number of Specimens*—When developing an *S-N* curve, a minimum of 15 specimens shall be tested at each stress level. The desired statistical power of the comparison and the variability to be expected from the cement formulation(s) being investigated should be considered when determining the appropriate sample size; while this may require more than 15 specimens per bone cement formulation at each stress level, 15 is the recommended minimum number to test. See X2.12 for further information.

9.6 Set the cycle counter and limit settings of the test frame controller to record the cumulative number of cycles applied to the test specimen and the appropriate test limits values to indicate specimen failure or deviations from the intended load system performance.

9.7 After the solution has reached the temperature requirements in 9.2, activate the test frame controller to begin the test.

9.8 Testing shall continue until specimen failure or the run-out limit is reached.

10. Calculation and Interpretation of Results

10.1 The maximum stress and the cycles to failure for each specimen should be recorded and plotted on an Stress Level versus number of cycles diagram, which is a plot of the number of cycles to failure on the *x*-axis at each of the stress levels examined on the *y*-axis. The techniques used to measure mean fatigue lives, as well as to compare statistical differences between sample groups, and calculate fatigue life are described in 10.2 – 10.6.

10.2 *Mean Fatigue Life*—For each stress level, the mean fatigue life and standard deviation about the mean shall be determined assuming a log-normal distribution; that is, assuming that the log-transformed number of cycles to failure is approximately normally distributed (**1**).[4] The mean log fatigue life is determined as follows. A sample size of N specimens is tested, and the total number of cycles to failure for each (denoted N_i) is recorded. Next take the natural log of the number of cycles: $X_i = \ln(N_i)$. The mean log number of cycles to failure is obtained via the sample mean:

$$\bar{X}_{\log} = \sum_{i=1}^{N} \frac{X_i}{N} \tag{1}$$

where:

N = total number of specimens in the sample group,
N_i = number of cycles to failure of ith specimen,
X_i = log-transformed number of cycles to failure of ith specimen: $X_i = \ln(N_i)$, and
$\bar{X}_{\log}$ = mean log fatigue life.

10.2.1 Using a similar approach, the sample standard deviation of the log fatigue life ($S_{X_{\log}}$) is determined.

$$S_{\bar{X}_{\log}} = \sqrt{\sum_{i=1}^{N} \frac{\left(X_i - \bar{X}_{\log}\right)^2}{N-1}} \tag{2}$$

10.2.2 These are expressed in more familiar terms, as cycles to failure, by calculating the following:

$$\text{Mean fatigue life} = e^{\bar{X}_{\log}} \tag{3}$$

10.2.3 A 95 % lower and upper bound for the mean number of cycles to failure can be obtained using the following formulas (using the delta method, see X4.1):

$$e^{\bar{X}_{\log}} \pm -1.96 * \left(e^{\bar{X}_{\log}} S_{\bar{X}_{\log}}\right) \tag{4}$$

10.3 *Parametric Statistical Comparisons*—Statistical differences between specimen groups may be determined by commonly used methods such as a two-sample independent *t*-test to compare two groups, or analysis of variance (ANOVA) to compare more than two groups. This comparison is performed at each stress level using published methods (**2**) which are available through many commercial statistical software packages. The use of these tests requires several assumptions; the two most relevant are normality and equal variances. That is, these tests assume that the number of cycles to failure in each bone cement at each stress level is approximately normally distributed, and that the variance of these normal distributions is the same for all of the bone cements. These are relatively strong assumptions, which may not be upheld. It is therefore recommended that these assumptions be assessed. Tests to assess normality include the Lillie for test and the Shapiro-Wilk test (**3**). However, these tests are based on large samples approximations, and having a sample size on the order of 15–30 observations per group may not be sufficient to guarantee reliable performance.

10.3.1 Often, the decision as to whether to analyze data on an untransformed or log-scale is based on a test for normality; the most common of these is the Shapiro-Wilk test (**4**). Based on a small simulation study using the results of the "round-robin" experiment, we found that the test rejects samples from

[4] The boldface numbers in parentheses refer to the list of references at the end of this standard.

a true normal distribution approximately 7.5 % of the time (out of an expected 5 %). If the data is assumed to arise from a gamma distribution (a highly skewed distribution which appears to be a reasonable fit to this data), the untransformed data is not rejected approximately 27 % of the time. This implies that reliance on the Shapiro-Wilk test may lead to incorrect application of statistical tests assuming normality; this is likely if a relatively small number of specimens are tested (N=15).

10.3.2 It is often recommended that a parametric analysis be performed using the log-transformed data—this assumes that the number of cycles to failure follows a log-normal distribution. If this is the case, then analyzing on the log scale would be expected to improve the normality of the data; the number of cycles is highly skewed with all values being non-negative, and some having extremely high values. Taking the log of the number of cycles is believed to make the resulting data more approximately normally distributed. In addition, calculating the mean based on the log scale reduces the effect of extremely large or small values (for example, outliers) on the sample mean. The disadvantage of analyzing on the log scale is that the units are in terms of log cycles rather than cycles. However, the transformed value can be back-transformed to the original scale (and an approximate 95 % confidence interval can be estimated via the delta method as shown in 10.2).

10.4 *Non-parametric Statistical Comparisons*—In situations in which the parametric statistical tests are not appropriate (for example, the number of cycles is not approximately normally distributed, or the variances of the different bone cements are unequal), non-parametric statistical methods are suggested for use in determining statistical differences between sample groups. Non-parametric tests are based upon the median, rather than the mean, and are therefore more robust because they are less influenced by the highly skewed nature of the data. In addition, as these tests are based on ranks, rather than upon the actual observation values, the results are the same regardless of whether or not the data are log-transformed. The Mann-Whitney U test (equivalent to the Wilcoxon rank sum test) is recommended for comparing two groups, and the Kruskal-Wallis test is recommended for comparing three or more groups. This comparison is performed at each stress level using published methods **(2)** which are available through many commercial statistical software packages.

10.5 *Recommendations for Analysis*—In light of these discussions, as well as an examination of the "round-robin" data and the observation that the number of cycles to failure must be non-negative and may be highly skewed (Appendix X5), an assumption of normality is somewhat tenuous. For the number of samples suggested here (n=15 per bone cement) it is recommended that non-parametric tests, which are more robust to non-normal data, be used for statistical inference and to compare different types of bone cement.

10.6 A brief description of the fracture characteristics; results of post-test photography or scanning electron microscopy or both; identification of fatigue mechanism; and the relative degree of transgranular and intergranular cracking would be highly beneficial. In addition, all fractured specimens will be examined visually for pores and failure occurring outside the gauge area.

11. Report

11.1 The test report shall include the following:

11.1.1 Manufacturer and brand of bone cement.

11.1.2 Product catalog number, lot number, and expiration date. If the cement is not in its final packing or sterilized, then the manufacturing date should be provided and noted that the bone cement components were not sterilized.

11.1.3 Composition of bone cement polymer powder and liquid.

11.1.4 Deviations from clinically used product (if applicable).

11.1.5 Description of cement storage, temperature of room during bone cement mixing and relative humidity, mixing method (that is, report duration of mixing, wait time (if applicable), determination of dough time, application time, and hardening time), and any deviations from the manufacturer's recommendations.

11.1.5.1 If vacuum mixing is used, the information and parameters described in 8.1.3 shall be reported.

11.1.6 Description of specimen fabrication method.

11.1.7 Description of specimen examination procedures, rejection rate, rejection criteria and rationale for the rejection criteria.

11.1.8 Duration of preconditioning, provided either for each specimen, or expressed as an average and range of duration.

11.1.9 Cyclic frequency.

11.1.10 A summary of the maximum cyclic stress and cycles to failure for each specimen tested.

11.1.10.1 Peak/valley load and displacement data in order to document the loads and displacements each sample experienced during testing.

11.1.11 A summary for each sample group describing at each stress level the following parameters:

11.1.11.1 Mean fatigue life, along with the standard deviation and 95 % confidence interval as presented in 10.2.

11.1.12 A description of the failure mode and failure location for each specimen that failed. Scanning electron microscopy (SEM) is suggested to identify the failure mode.

11.1.13 The mean fatigue life at each load level. A description of the analytical or statistical techniques used for determining the fatigue life should be included.

11.1.14 Any deviations from the specified test method.

12. Keywords

12.1 acrylic bone cement; fatigue; fatigue life

APPENDIXES

(Nonmandatory Information)

X1. FORMULAS

X1.1 Formulas are presented following the notation of Hollander and Wolfe (**5**).

X1.2 *Formula for Wilcoxon Rank Sum Test*:

X1.2.1 The Wilcoxon rank sum test (which is equivalent to the Mann-Whitney test) is a non-parametric analog of the two-sample *t*-test.

X1.2.2 This test assumes that there are two independent groups, and the question of interest is whether the medians of the two groups are equal. To implement the test, refer to the m observations from the first group as X and the n observations from the second group as Y. Order all of the observations from smallest to largest, and assign ranks to each observation. Denote the rank of all of the values from the second group as $S_1, ..., S_n$.

X1.2.3 Calculate the sum of the ranks of the observations in the second group:

$$W = \sum_{j=1}^{n} S_j$$

X1.2.4 To test for equivalence of medians in a 2-sided test, calculate the test statistic:

$$W^* = \frac{W - [n(m+n+1)/2]}{\sqrt{mn(m+n+1)/12}}$$

and refer W^* to a standard normal distribution; that is, reject the null hypothesis of equal medians if $|W^*| \geq z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the $1 - \alpha/2$th percentile from the standard normal distribution.

X1.2.5 If there are ties in the ranks, assign the average rank to each of the tied values, and adjust the test statistic W^* as follows:

$$W^* = \frac{W - [n(m+n+1)/2]}{\sqrt{\frac{mn(N+1)}{12} - \left\{ \frac{mn}{12N(N-1)} \sum_{j=1}^{g} (t_j - 1)t_j(t_j+1) \right\}}}$$

where g represents the number of tied groups (thus, if there are no ties, $g=N$ and the formula simplifies to the first form).

X1.3 *Formula for Kruskal-Wallis Test*:

X1.3.1 The Kruskal-Wallis test is an extension of the Wilcoxon rank sum test to more than two independent groups; it is a non-parametric analog of the 1-way ANOVA.

X1.3.2 To implement this test, first order all N of the observations from all of the k groups from smallest to largest. Denote as r_{ij} the rank of observation X_{ij}, and the number of samples in the jth group as n_j. Calculate:

$$R_j = \sum_{i=1}^{n_j} r_{ij} \quad \text{and} \quad R_j = \frac{R_j}{n_j}$$

X1.3.3 The test statistic for the Kruskal-Wallis test is then calculated as:

$$H = \left(\frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} \right) - 3(N+1)$$

X1.3.4 Compare H to a ${\chi_{k-1}}^2$ (chi-square with k-1 degrees of freedom) distribution, and reject the null hypothesis of equality of medians across groups if $H \geq {\chi_{k-1,\alpha}}^2$, where ${\chi_{k-1,1-\alpha}}^2$ is the 1 – αth percentile from a chi-square distribution with k-1 degrees of freedom.

X1.3.5 If there are ties present in the data, calculate the modified test statistic:

$$H' = \frac{H}{1 - \left(\sum_{j=1}^{g} (t_j - 1)/(N^3 - N) \right)}$$

where g represents the number of tied groups (thus, if there are no ties, $g=N$ and the formula simplifies to the first form).

X2. RATIONALE

X2.1 This test method is intended to provide the user with standard and well-established procedures for evaluating the fatigue properties of bone cement materials. Specimen parameters, test procedures, data analysis techniques, and reporting requirements are provided.

X2.2 The test method does not specify the mixing conditions to use for the preparation of the test specimens. Considerable research is currently being performed on bone cement and the committee did not want to unnecessarily limit the conditions or parameters that are being investigated by excluding them from the standard.

X2.3 It is important to realize that this test method is intended to characterize the bone cement material—not the bone cement which is used *in vivo*. Some consideration has been given to the parameters which the cement encounters during *in vivo* use (37°C temperature and PBS solution); however, it is not practical to try and completely simulate the clinical use of bone cement. The results obtained from this test method characterize the bone cement material for a specified set of conditions, but they may not necessarily reflect the

cement's clinical performance.

X2.4 The orthopedic literature generally reports that joint replacement patients may be expected to take 2 million to 3 million steps per year (1 million to 1.5 million gait cycles). Therefore bone cement, when used for securing artificial hip and knee joints, is exposed to millions of loading cycles during its use. It is appropriate to expect that the fatigue testing of bone cement would likewise subject the test specimens to millions of cycles. However, it should be kept in mind that the fatigue testing cycles described herein may not be directly correlated with the duration of clinical implantation because of the limitations described in X2.3. The committee has chosen a runout limit of 5 million load cycles to provide a reasonable representation of the high cycle fatigue loading to which bone cement is exposed while also addressing the economic and practical considerations of testing at realistic load rates (see X2.6) in a reasonable period of time.

X2.5 It is recognized that the total time for which the specimens are presoaked may have an important effect on their fatigue performance since fluid uptake and polymer degradation are functions of time. Most articles in the literature have reported presoaking cement specimens for a minimum of 7 days. This test method provides a maximum presoaking time of 60 days to reasonably minimize the effect of different presoaking times on the results. It has been shown that most formulations of acrylic cement will experience a weight gain of 2.0 to 2.5 % during an extended soak period of 100 days **(6)**. It is recommended that the user identify a uniform presoak time that brings the specimens to a weight-gain plateau at which they are gaining less than 0.2 % of their weight per week. As reasonably possible, all of the test specimens should have the same soaking time before testing.

X2.6 Because acrylic bone cement is a viscoelastic material, its cyclic stress-strain behavior is rate dependent. However, frequency up to 5 Hz has been shown not to affect the cycles to failure for PMMA based bone cement that were tested **(7)**. It has been shown in tension-tension tests that an elevation in the testing frequency tends to increase the fatigue life of bone cement **(8)**. The user is cautioned to verify from the literature or from new tests that for the formulation being tested the use of any elevated frequency should not have an effect on the reported results.

X2.7 When establishing load levels to test bone cements at, it is important that the specimens are subjected to stress levels which the cement would likely experience *in vivo*. For normal joint loading, the nominal tensile stress levels in the cement mantle surrounding a stable hip stem are reported to be between 3 and 11 MPa **(9-11)**. The specified maximum stress levels are chosen to provide sufficient finite life fatigue data to develop an *S-N* curve, while providing some data in the range of expected *in vivo* stresses. The load levels may depend on the location of use (for example, hip versus knee versus spine). In the past, some investigators have recommended fitting a survival curve to the failure data and comparing bone cements from different experiments by calculating the expected number of cycles to failure for a common load level for both cements; this is not recommended. Fitting such curves makes a number of assumptions about the data, which may not be valid, or testable. Further, attempting to fit a model with several parameters based on only three load levels could lead to over-fitting the dataset, resulting in the model performing poorly for other data. Finally, using such a model to extrapolate (for example, predicting the expected number of cycles to failure at a load level not examined in the experiment) is statistically questionable and could lead to unstable and inaccurate estimates. It is instead recommended that any comparisons be performed by matching the stress level directly (for example, comparing the number of cycles to failure at 10, 12.5, and 15 MPa across different labs).

X2.8 Differences in specimen fabrication method (user experience, cement application to mold technique mold materials,) may lead to different test results for the same cement, tested under identical conditions **(12)**. The scientific literature does not provide a clear indication as to the preferred method of specimen fabrication. For the current time, the standard provides a recommended procedure, while allowing alternative methods, provided they are fully described. The user is cautioned against comparing different sets of data generated using this even though the same procedures are used for specimen preparation because of variability in specimen preparation from one investigator to another. Whenever possible, investigator(s) should plan to test their own concurrent controls for comparisons and not rely on previously published values.

X2.9 Fatigue of the cement mantle has been implicated as one of the mechanisms leading to orthopaedic prosthesis loosening and eventual arthroplasty failure **(13)**. Fractographic analysis of cement explanted from failed prostheses demonstrate characteristics consistent with PMMA fatigue crack initiation and propagation to failure **(14)**. The polymer chemistry, molecular weight, radiopacifier, voids, mixing method, and sterilization method have all been presented by various authors as influencing the fatigue properties of bone cement. The test method described herein provides a means for evaluating the effect of these various parameters in a controlled manner.

X2.10 The cement mantle surrounding hip and knee implants is subjected to complex tensile and compressive stresses. Generally, fatigue cracks will only initiate and extend under localized tensile stresses. Fully reversed loading has been selected for this test method for two reasons: *(1)* for a given maximum stress, fully reversed loading provides the most conservative estimate of fatigue performance, and *(2)* the vast majority of the bone cement fatigue data in the U.S. literature uses fully reversed loading.

X2.11 The rejection criteria should be used to identify specimens based on the specimens having surface defects. The surface finish should be free of any surface defect that may influence the fatigue performance of the specimen. While internal defects may result lower number of cycles to failure, the current method suggest that these specimens be randomized and tested at the different stress levels. Therefore, specimens

with surface defects should be not be tested because surface defects can have a great influence on the cycles to failure reported for this specific material in using this specific test set up. However, internal defect detected on X-ray or micro CT should be used to help determine if there is a correlation between cycles to failure and porosity in the gauge length.

X2.12 A minimum of 15 specimens was considered to be an appropriate balance of *(1)* the requirement for having sufficient data to allow statistical comparisons and generation of the *S-N* curve with *(2)* the resources required to perform high-cycle fatigue testing. The user is encouraged to calculate the power of the test comparisons, using well-published methods (**2**), for the particular cement formulation(s) being investigated to determine the appropriateness of the sample size used. Based on the variability seen in the data from the "round-robin" experiment, a sample size of 15 specimens per bone cement at 12.5 MPa stress level would have approximately 80 % power to detect a difference in the number of cycles of approximately 140 000 cycles[5]. As a result, if only 15 specimens per bone cement are used and the statistical analysis is not significant, the correct conclusion is not that the two bone cements are equivalent, but that the difference in fatigue life is less than 140,000 cycles. If more precision on the equivalence of the bone cements is desired, a larger number of specimens from each should be tested to ensure sufficient power to rule out a difference of a given magnitude.

X2.12.1 *Elaboration on Sample Size for Non-parametric Approaches Recommended in the ASTM Bone Cement Analysis Standard*—Parametric analyses (for example, (**8**) *t*-test and ANOVA) assume that the data being analyzed are approximately normally distributed. Non-parametric analogs (for example, Mann-Whitney / Wilcoxon rank sum test and Kruskal-Wallis test) do not make this assumption, and are therefore more appropriate if there is concern regarding the normality of the data being analyzed. A limitation to the use of these methods is the perceived absence of sample size formulas for non-parametric methods. Here, we provide three options for estimating the sample size required to detect a significant difference between two (or more) different bone cements: use of parametric sample size formulas as a ballpark estimate, a direct sample size formula for the Wilcoxon rank sum test, or simulation. These are discussed in turn in the following sections.

X2.12.1.1 *Use Parametric Approaches as a Rough Estimate*—One approach to sample size estimation in non-parametric analyses is to calculate the sample size for the corresponding parametric test, then use that sample size as a ballpark estimate for the sample size required for the non-parametric approach. This method is generally reasonably accurate, due to the statistical concept of relative efficiency. Relative efficiency is a means of comparing various statistical tests based on the standard error of the test statistic. For example, Hollander and Wolfe (**5**) report that the relative efficiency of the Wilcoxon rank sum test relative to the parametric analog (two sample *t* -test) is at least 86.4 %. This means that if the data were analyzed with the Wilcoxon rank sum test rather than the two-sample *t*-test, the efficiency loss would be approximately 14 % or less. This loss in efficiency can be dealt with by slightly increasing the sample size if the non-parametric analysis is used. Many commercially available programs will conduct sample size estimates for the two-sample *t*-test.

X2.12.1.2 *Sample Size Formulas Presented by Noether*—Noether (**15**) provides sample size formulas for several commonly encountered non-parametric tests, including the Wilcoxon rank sum test. Assuming that an equal number of observations will be selected from both groups, the total sample size (N) associated with a type I error rate (α) and power $1 - \beta$ is obtained via:

$$N = 2n = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{3(p'' - 0.5)^2}$$

where $z_{1-\alpha}$ and $z_{1-\beta}$ are the $1 - \alpha$ and $1 - \beta$ percentiles of the standard normal distribution, respectively. p'' represents the expected probability that samples from group 1 are larger than samples from group 2; that is, $1/n^2 U$, where n is the number of specimens in each group ($N/2$ if the two groups have the same number of specimens), and U is the expected Mann-Whitney test statistic.

X2.12.1.3 *Simulation*—A third approach is to use computer simulation (sometimes referred to as Monte Carlo studies) to estimate the required sample size. Simulations routinely used in statistical research to obtain estimates which are impossible or intractable to solve for directly (**16**). An example of how simulation could be used in the context of sample size estimation for non-parametric approaches is the following. First, a candidate distribution for the number of cycles to breaking for the two different bone cements is selected in order to have characteristics similar to those expected in the experiment (for example, mean, median, or standard deviation expected). Next, some number of specimens (N) are drawn from each of these two distributions and the Wilcoxon rank sum test applied to the resulting samples. This process is repeated 1,000 times, and the number of times the test is statistically significant divided by 1,000 is the estimated statistical power of the test at that sample size. The process can be repeated using a larger value of N until one associated with the desired power is obtained. Typically, simulation studies require programming in some computer language (for example, R, SAS, Fortran, C++) and may benefit from input from a statistician.

X2.13 In general, hand mixing under ambient pressure will produce specimens with the shortest fatigue life. Other methods of mixing (for example, vacuum mixing and centrifugation) generally produce specimens with similar or greater fatigue life than hand-mixed specimens; however, exceptions to this have been reported (**17**, **18-21**).

[5] This number was estimated as follows. The data from the round-robin analysis was used to fit a gamma distribution for the number of cycles to failure. 15 random draws from this gamma distribution were sampled. 15 different random draws were selected from a gamma distribution with the same parameters, but shifted to the right by parameter delta. By varying delta, and comparing the number of times the simulated data found a statistically significant difference between the two groups, the threshold value of 140,000 was determined, as it has approximately 85 % power when the data are analyzed with the Wilcoxon rank sum test.

X2.14 In 2002, a round-robin experiment was conducted in order to establish the precision and accuracy to be expected from this test method have been established based on a multiple-laboratory experiment, described in Appendix X5. Six different laboratories followed a standardized procedure based on the previous version of this standard. The main findings of this experiment are summarized below; further details of the procedure, data, and analysis can be found in the appendix.

X2.14.1 The data were log-transformed, but the data from most of the labs was still not approximately normal based on significant *p*-values for the Shapiro-Wilk test. Samples which were rejected due to radiographic defects showed a significant decrease in the number of cycles to failure relative to radiographically acceptable samples; this was consistent regardless of whether the data were analyzed as number of cycles to failure or log-transformed number of cycles to failure, or analysis method (parametric or non-parametric).

X2.14.2 There was also significant variation between the results obtained at each laboratory. This highlights the importance of following the testing procedure presented in this standard as closely as possible, as well as documenting all relevant testing parameters. Clarifications and modifications have been made to the molding material and preparation of the specimens to improve the quality of the specimens made using in this test method.

X3. SUGGESTED SPECIMEN MOLDING METHOD

X3.1 Scope

X3.1.1 This appendix provides a suggested fixture and method for molding cement specimens.

X3.2 Summary of Procedure

X3.2.1 A silicone mold is produced by curing liquid silicone around metallic positive blanks. After curing, the metallic positive blanks are removed to leave internal cavities in the mold with the intended cement specimen dimensions. The metallic positive blanks should be used for comparison with the molded PMMA specimens. The silicone mold should be free of voids or surface defects adjacent to the metallic blank. If the mold has voids the remake the mold to ensure that good quality bone cement specimens can be produced through the use of silicone mold.

X3.2.2 Liquid cement is poured or injected into the cavities in the silicone mold, which is then placed on a water bath. The cement is allowed to polymerize and are then ejected from the mold to produce the cement specimens.

X3.3 Apparatus

X3.3.1 *Positive Blank*—A metallic blank which is machined to the specimen dimensions provided in Fig. 1. The master is used to produce the internal cavity of the final silicone mold. The surface finish of the positive blank should be at least *Ra* = 0.05 µm or better.

X3.3.2 *Specimen Master Holder*—An assembly of two metal plates and four bolts (or equivalent) which is used to hold the specimen masters during the pouring and curing of the silicone mold. Suggested holder dimensions are provided in Fig. X3.1.

X3.3.3 *U-Channel*—An assembly consisting of three metal plates (or equivalent) which form a trough with a U-shaped profile. Together with the specimen master holder, this forms the molding chamber. Suggested U-channel dimensions are provided in Fig. X3.2.

X3.3.4 *Molding Chamber*—The square internal cavity formed by the insertion of the specimen master holder into the U-channel. The walls of the U-channel along with the two end plates of the specimen master holder produce the chamber into which the silicone is poured during the fabrication of the mold.

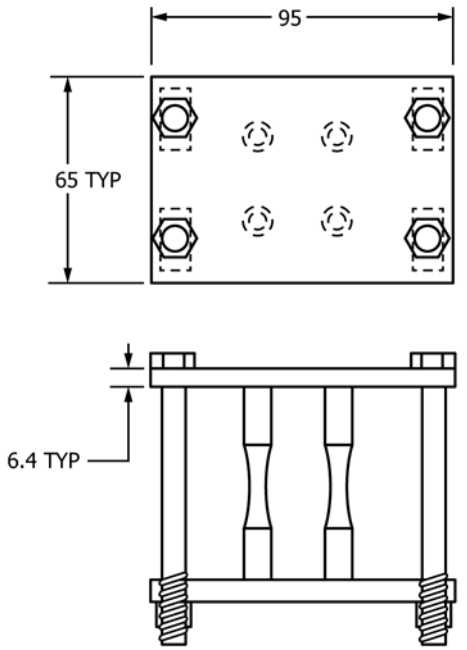


FIG. X3.1 Dimensions of Specimen Master Holder

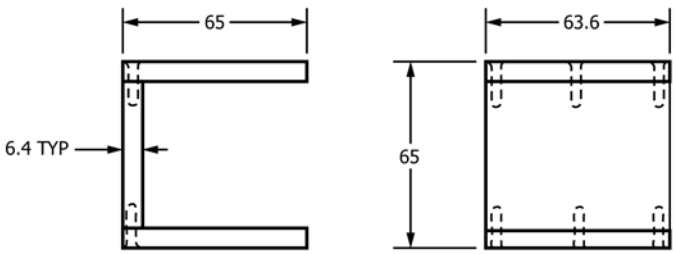


FIG. X3.2 Dimensions of U-Channel

X3.3.5 *Water Bath*—A chamber which is filled with water into which the mold is placed while the cement is polymerizing. The bath should have provisions for maintaining a constant temperature to an accuracy of ±2°C.

X3.4 Preparation of Silicone Mold

X3.4.1 Several metallic master specimens are machined to the final dimensions as provided in Fig. 1. The finish should be *Ra* = 0.05 µm or better.

X3.4.2 A number of specimen masters are placed into the specimen master holder such that they are sufficiently spaced apart from one another and then secured in the holder. Position the masters a minimum of 15 mm from sides of master holder.

X3.4.3 The holder with masters is turned on its side and inserted into the U-channel such that the two end plates of the master holder form the remaining two walls of the molding chamber. Any gaps between the master holder and U-channel should be filled with putty or appropriately sealed (for example, epoxy adhesive) to seal any joints and prevent seepage of silicone during molding. Under exhaust hood, spray inside of the mold and masters with mold release.

X3.4.4 The molding chamber should be filled with a two-part silicone system (see silicone material specifications below) to cover the specimen masters to a depth of at least 10 mm. The silicone should then be allowed to cure according to the manufacturer's instructions. Molding success with smooth surface finish and a mold that can be reused many times has been reported with silicone material of shore A durometer 25.

X3.4.5 After the silicone is cured, the silicone mold is removed from the molding chamber, and the specimen masters ejected by carefully pushing them from the mold with a blunt rod. The cavities formed by the specimen masters which remain in the mold are used for molding the cement specimens.

X3.5 Making Silicone Molds

X3.5.1 *Equipment/Supplies:*
X3.5.1.1 Fume Hood.
X3.5.1.2 Steel Rule.
X3.5.1.3 Tri-Pour 800 mL Beaker.
X3.5.1.4 Spatulas/Tongue Depresser.
X3.5.1.5 Vacuum/Oven.
X3.5.1.6 Scales/Balance.
X3.5.1.7 Aluminum Foil.
X3.5.1.8 Punch (flat nose).
X3.5.1.9 Scotch Tape.
X3.5.1.10 Instant Adhesive.
X3.5.1.11 Scotch–weld (epoxy adhesive).
X3.5.1.12 U-Shaped Mold as described in Fig. X3.1 and Fig. X3.2.
X3.5.1.13 Exacto Knife.
X3.5.1.14 Stoner, Zero Stick Mold Release.

X3.5.2 *Silicone Material:*

Description	Application
A two-part, translucent gray, pourable silicone system	mold-making material
10:1 Mix Ration (Part A:B)	

Typical Properties	Result	Metric Conversion	ASTM	NT-TM
	Uncured			
Appearance	Translucent gray	. . .	D2090	002
Viscosity, Part A	110,000 cP	110,000 mPas	D1084, D2196	001
Viscosity, Part B	1,600 cP	1,600 mPas	D1084, D2196	001
	Cured: 3 mm @ 150°C; Post-Cured: 1 h @ 150°C; Stabilize for 3 h @ ambient temp and humidity			
Specific Gravity	1.09	. . .	D792	003
Durometer, Type A	25	. . .	D2240	006
Elongation	530%	. . .	D412	007

X3.5.3 *Instructions for Use:*

X3.5.3.1 *Mixing*—Thoroughly mix Part A and Part B in a 10:1 ratio by weight. Take care to minimize air entrapment during mixing.

X3.5.4 *Procedure:*

NOTE X3.1—Use proper eye and hand protection when handling chemicals.

X3.5.4.1 *Making Silicone Mold:*

(1) Place 800 mL beaker on scale and tare. Use spatula to place desired amount of silicone part A into beaker, tare. Then, pour in required amount of part B. (Note: Ratio is 10:1.)

(2) Use spatula to mix parts A and B thoroughly.

(3) It is recommended if bubbles are noticed during the mixing of the silicone material then applying vacuum under pressure will remove the bubble from silicone material. In order to remove any bubbles formed during mixing of the 2 parts of silicone, place the beaker of the mixed silicone under vacuum and with no heat). Note that applied vacuum allows for the release of bubbles forming at the surface of the beaker.

The user should cycle the vacuum on and off for a minimum of four cycles or until all visible bubbles have been removed from the silicone material.

1st cycle:	5 ± 0.5 minutes of vacuum at 28 ± 3 mm Hg (converted to kPa = 3.73 ± 0.4 kPa), then release the vacuum to collapse the bubbles at the surface.
2nd cycle:	3 ± 0.5 minutes of vacuum at 28 ± 3 mm Hg then release the vacuum to collapse the bubbles at the surface.
3rd cycle:	3 ± 0.5 minutes of vacuum at 28 ± 3 mm Hg then release the vacuum to collapse the bubbles at the surface.
4th cycle:	3 ± 0.5 minutes of vacuum at 28 ± 3 mm Hg then release the vacuum to collapse the bubbles at the surface, if necessary.

(4) Remove from vacuum and pour mixture into U-channel mold, leaving 2 to 3 mm of masters exposed.

(5) Allow mold to sit at room temperature approximately 1 h. This will allow trapped air to rise to the surface.

(6) Using Exacto knife, pop surface bubbles. Repeat about every 30 min for 2 h.

(7) Mold can then be placed in 37°C oven overnight to accelerate curing of the silicone elastomer, or allow mold to stand at room temperature 24 h.

(8) When cure is complete, disassemble the U shape mold.

(9) Twist mold from side to side to release specimen masters and push masters free from the mold using a flat nosed punch.

X3.6 Injection Molding of Bone Cement Specimens in the Dough Phase, and Finishing of Fatigue Test Specimens

X3.6.1 The silicone mold is placed on flat surface clean surface, with the axis of the specimen cavities oriented vertically.

X3.6.2 Bone cement is stored and mixed as described in 8.1. The cement is mixed in accordance to the manufacturer's instructions for use and then injected in the dough phase into the specimen cavities so that each cavity is slightly overfilled.

X3.6.3 The following describes how two people are to make bone cement fatigue test specimens using the silicone mold in accordance with ASTM F2118.

X3.6.3.1 Two trained staff (one gunner and one time manager/mold finisher) work together as a team to mix and mold of bone cement specimens.

X3.6.3.2 Use translucent silicone mold that is dry and clean. The mold design and mold material should conform to ASTM F2118. The silicone should be transparent and free of any porosity to help allow for defect free specimens to be formed from the mold. In addition, the transparent mold will allow for visualization during bone cement filling in mold. This is helpful in allowing the users to determine the proper filling of the molds.

X3.6.3.3 Accurate measurement of temperature and relative humidity. Need to cross reference temperature and relative humidity.

X3.6.3.4 Bone cement needs to be equilibrated to room temperature before mixing.

X3.6.3.5 Stop watch (measurement of the monomer pour time, wooden spatula mixing, resting time, dough time, hardening time, determination of start of exotherm, maximum exotherm time, and of the exotherm).

X3.6.3.6 Nitrole gloves used to test for dough time and stickiness.

X3.6.3.7 Paint mask used so as minimize the exposure to MMA monomer smell.

X3.6.3.8 Surgical scalpel used to cut excess bone cement from around each hole.

X3.6.3.9 Pliers (used to uniaxially extract the cured bone cement specimen from silicone mold).

X3.6.3.10 Polymer blunt probe (assistant pushes with blunt probe until the Gunner grapes the end of the cured specimen with plies. The assistant holds the mold so that the specimen is extracted from the mold only in a uniaxial direction.

NOTE X3.2—The water acts like a natural lubricant when extracting specimens from the mold.

X3.6.3.11 Meat temperature probe to determine the length of the exotherm and maximum temperature.

X3.6.3.12 Mixing of (for example, Brand A) Bone Cement in Plastic Beaker with Wooden Spatula:

(1) Pour the bone cement powder component into the plastic beaker.

(2) Open monomer under hood and Start stop watch when you pour monomer over the powder polymer (takes approximately 8 to 10 s).

(3) Mix bone cement for 30 s with wooden spatula as to wet all the powder.

(4) Transfer to open cement gun cartridge then attach nozzle. Pull gun trigger until the bone cement extrudes 1 to 2 cc from the nozzle. (The internal diameter of the nozzle tip (10 mm) should fit into mold, an adapter may be used.)

(5) Wait until *x* min (dough phase, as reported from manufactures package insert for handling characteristics at different temperature chart) from the start, then with gloves touch the cement that was extruded from the cement gun. If sticky, determine how sticky and wait. Wait 30 s then extrude a little amount of curing bone cement and determine if you have reached the dough phase when the bone cement does not stick to the glove, then begin filling of mold.

(6) The gunner places the nozzle tip into the silicon mold and inject the bone cement when the assistant tells the other he or she is all clear. The nozzle should be correctly aligned in the end of the mold. The assistant helps to ensure the nozzle tip is properly aligned. Because the nozzle tip fits into the mold, good pressurization can be achieved. The assistant continues to hold the mold and visually watches the injection of the bone cement and tells the gunner when the bone cement has filled down to the bottom of the mold. The assistant, quickly then guides the gunner's nozzle into the next empty mold then repeats the filling step.

(7) Once all four molds have been filled, the assistant then starts the finishing process both ends of the bone cement fatigue test specimens. This occurs near then end of the working time of the bone cement. Any excess bone cement which was extruded when gun nozzle tip was transferred from specimen to specimen can be used to form a uniform 2 mm mantle on the top of mold over all four top holes. The mold is completely lifted up and turned sideways as to expose the bottom opening of the specimens. Additional bone cement is extruded from the gun and placed on the bottom surface to form a uniform bone cement mantle as previously formed on the top surface. When the mold positioned on its side, a razor should be used to cut away the bone cement mantle away from each opening of each specimen (that is, cutting the mantle away from each opening by cutting perpendicular to the orientation of the cement specimen). Then, use your index finger smooth the ends of specimen before it hardens so as to allow for safe extraction of the bone cement specimen after it has hardened.

(8) At approximately *Y* min the exotherm begins. The exotherm last for 1 to 2 min with the peak occurring around *Z* min.

(9) Place molded specimens into a 37°C water bath for at least 10 min after the exotherm as ended.

(10) Extraction of the bone cement specimens from silicone mold. A polymer blunt probe is pushed against one end of specimen by the assistant and the other person pulls the specimen from the mold with a pair of pliers. Like a dentist

using the pliers, one person grabs the exposed end of the specimen. The assistant holds the mold so that the specimen is extracted from the mold by the dentist in a uniaxial direction. The water acts like a natural lubricant when extracting specimens from the mold.

X4. SUGGESTED STATISTICAL METHODS

X4.1 Delta Method

X4.1.1 Suppose X (the number of cycles to failure) has a distribution with finite variance σ_x^2, and to improve the approximate normality of the data, the analysis is performed on the log transformed variable. If an approximate confidence interval for the number of cycles is desired, the data must be back-transformed to the scale of cycles rather than log cycles ($Y = e^X$), and the variance of Y (σ_Y^2) must be estimated.

X4.1.2 Using the delta method **(22)**:

$$Y = f(X) = e^X$$
$$\sigma_Y^2 \approx [f'(X)]^2\, \sigma_X^2$$
$$f'(X) = f'[e^X] = e^X$$

So,

$$\sigma_Y^2 \approx [f'(X)]^2\, \sigma_X^2 = (e^X \sigma_X)^2$$

and an approximate 95 % confidence interval for Y (the mean number of cycles on the log scale) is given by:

$$Y \pm 1.96 * e^X \sigma_X$$

X4.1.3 These values (see 10.2) provide an approximate 95 % confidence interval on the untransformed scale (that is, in terms of cycles, not log cycles).

X5. ANALYSIS OF "ROUND-ROBIN" EXPERIMENT (Statistical analysis of bone cement "round-robin" data)

X5.1 Executive Summary

X5.1.1 The current ASTM guidelines recommend a strict testing regimen for evaluating the number of cycles of bone cement, as well as appropriate statistical methods for evaluating and comparing different types of bone cement.

X5.1.2 The data analyzed in this report is from a "round-robin" experiment at six different labs, each of which examined at least 15 radiographically acceptable samples as well as several radiographically unacceptable samples. The same bone cement was used at each lab. The purpose of this experiment was to determine the expected level of variation present in this type of evaluation.

X5.1.3 The results of this experiment suggest that there appears to be significant variation in the number of stress cycles before breaking based on the acceptance/rejection of samples by radiographic means. As a result, it is likely that discarding samples based on radiography results leads to a systematic bias of the life of the material, because the rejected samples are more likely to break at a lower stress cycle number. It is understood that fatigue testing a bone cement as described in this appendix, with pores greater than 1 mm, may not evaluate the intrinsic properties of a given bone cement formulation. Testing all finished specimens is more representative of clinical use.

X5.1.4 In addition, there is also evidence of significant variation between labs. This suggests that either: there are differences in how each lab conducts the testing, so that they are either not following the ASTM guidelines, or there are additional variables which affect the life of the sample besides those laid out in the ASTM guidelines.

X5.1.5 Regardless of the reason, it is important that experiments which are conducted at more than one lab account for this potential lab-to-lab variability.

X5.2 Research Questions

X5.2.1 Is there a significant difference between the different labs regarding the proportion of samples rejected?

X5.2.2 Are the results significantly different between the labs (assessed using both parametric and non-parametric methods)?

X5.2.3 Are the results significantly different between accepted and rejected samples (assessed using both parametric and non-parametric methods)?

X5.2.4 Is it better to analyze on non-transformed or log-transformed scale?

X5.3 Results

X5.3.1 *Is there significant variation between labs in the proportion of samples rejected?*

X5.3.1.1 The overall acceptance probability is 57.5 %; acceptance probabilities by lab are presented in Table X5.1. There is no significant difference in the proportion of sample acceptance by lab ($\chi_5^2 = 8.58$, $p = 0.1269$).

X5.3.2 *Are the results significantly different between the labs?*

X5.3.2.1 As the samples at each lab are all made from the same batch and following the same protocol, it would be expected that the results would be the same at each lab. If this is not the case, then the standards as currently written may need to specify that lab be included in any analysis of this kind of

TABLE X5.1 Radiographic Status by Lab

Radiographic Status	Lab						
Frequency Col Pct	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Total
Reject	20	15	8	20	4	13	80
	38.5	50.0	34.8	57.1	21.1	44.8	
Accept	32	15	15	15	15	16	108
	61.5	50.0	65.2	42.9	78.9	55.2	
Total	52	30	23	35	19	29	188

stress data. The effect of site-to-site variability should be minimized to the extent possible, as the round-robin experiment suggests that there are potentially significant differences from laboratory to laboratory.

X5.3.2.2 Based on the summary statistics presented in Table X5.2 (shown graphically in Fig. X5.1), there is evidence of considerable heterogeneity in the number of cycles at each lab.

X5.3.2.3 *Parametric Analysis*—The parametric analysis of the effect of lab was implemented via a one-way ANOVA using the number of cycles as the dependent variable and lab as the independent variable. The results ($F_{5,182} = 9.37$, $p < 0.0001$) clearly indicate that there is a significant difference in the number of cycles by lab.

X5.3.2.4 *Non-parametric Analysis*—The non-parametric analysis of the effect of lab was implemented via the Kruskal-Wallis test using the number of cycles as the dependent variable and lab as the independent variable. The results, shown in the table below, clearly indicate that there is a significant difference in the number of cycles by lab (Kruskal-Wallis $\chi_5^2 = 88.47$, $p \leq 0.0001$).

X5.3.3 *Are the results significantly different between accepted and rejected samples?*

X5.3.3.1 Currently, the ASTM guidelines recommend discarding samples which are rejected by radiographic analysis and not including them in the testing. It is possible that samples which are rejected would have significantly different breaking performance relative to those samples which are not rejected. The data were analyzed to examine this possibility. See Table X5.3 and Fig. X5.2.

X5.3.3.2 *Parametric Analysis*—The parametric test to detect a significant difference between the mean number of cycles between radiographically acceptable and rejected samples is the 2-sample *t*-test (Satterthwaite approximation: $t_{146} = -4.08$, $p < 0.0001$). See Table X5.4 and Fig. X5.3 for parametric analysis using log-transformed cycles (2 sample *t*-test, Satterthwaite approximation: $t_{144} = -5.04$, $p < 0.0001$).

X5.3.3.3 *Non-parametric Analysis*—The non-parametric analysis of the effect of status was implemented via the Wilcoxon Rank Sum test using the number of cycles as the dependent variable and lab as the independent variable. The results clearly indicate that there is a significant difference in the number of cycles by lab (Wilcoxon Rank Sum test $z = -4.63$, $p < 0.0001$). No difference in results via the non-parametric approach using the log-transformed values relative to untransformed.

X5.3.4 *Two-Way Analysis Approach:*

X5.3.4.1 As the above analyses suggest that both radiographic status and lab are significant sources of variability, a two-way analysis was also conducted to examine their joint effect. See Table X5.5 and Fig. X5.4. (A lab by radiographic status interaction term was included in the ANOVA model and found to be significant, but there is no non-parametric analog for this analysis. This interaction's significance suggests that the difference between radiographically acceptable and rejected samples differs from lab to lab.)

X5.3.4.2 *Parametric Analysis*—The parametric analysis of the effect of lab was implemented via a two-way ANOVA using the number of cycles as the dependent variable with lab and radiographic status as the independent variables. The results clearly indicate that there is a significant difference in the number of cycles by both lab and radiographic status ($F_{11,176} = 11.39$, $p < 0.0001$). See Table X5.6. For parametric analysis on log-transformed cycles ($F_{11,176} = 11.39$, $p < 0.0001$), see Table X5.7.

TABLE X5.2 Descriptive Statistics by Lab

Lab	N	Mean	Median	Standard Deviation	Coefficient of Variation	Minimum	Maximum
Lab 1	52	146505.0	97242.5	157046.6	107.2	238	1001717
Lab 2	30	34507.7	29084	27765.9	80.5	102	121124
Lab 3	23	69004.3	64314	45415.3	65.8	16797	206635
Lab 4	35	185852.4	103750	242465.3	130.5	2500	1037238
Lab 5	19	17578.5	12429	18503.8	105.3	149	71174
Lab 6	29	13441.2	12728	10016.8	74.5	644	36178

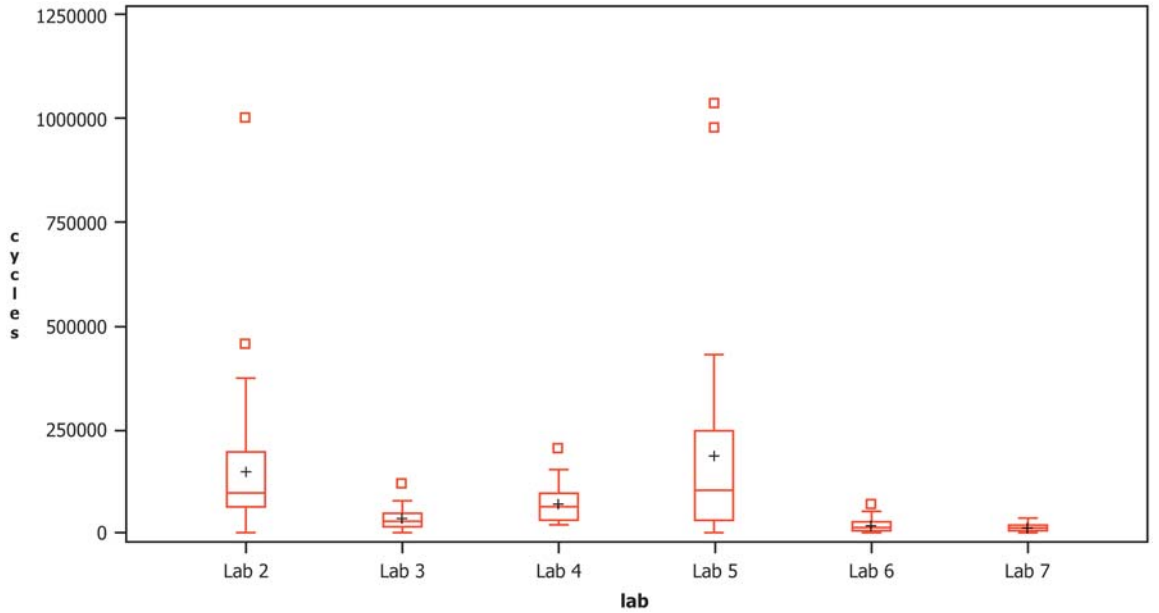


FIG. X5.1 Box Plots of Number of Cycles by Lab

TABLE X5.3 Descriptive Statistics by Radiographic Status

Status	N	Mean	Median	Standard Deviation	Coefficient of Variation	Minimum	Maximum
Reject	80	48088.6	25577	69506.8	144.5	102	375642
Accept	108	126130.8	69793.5	181632.3	144.0	3432	1037238

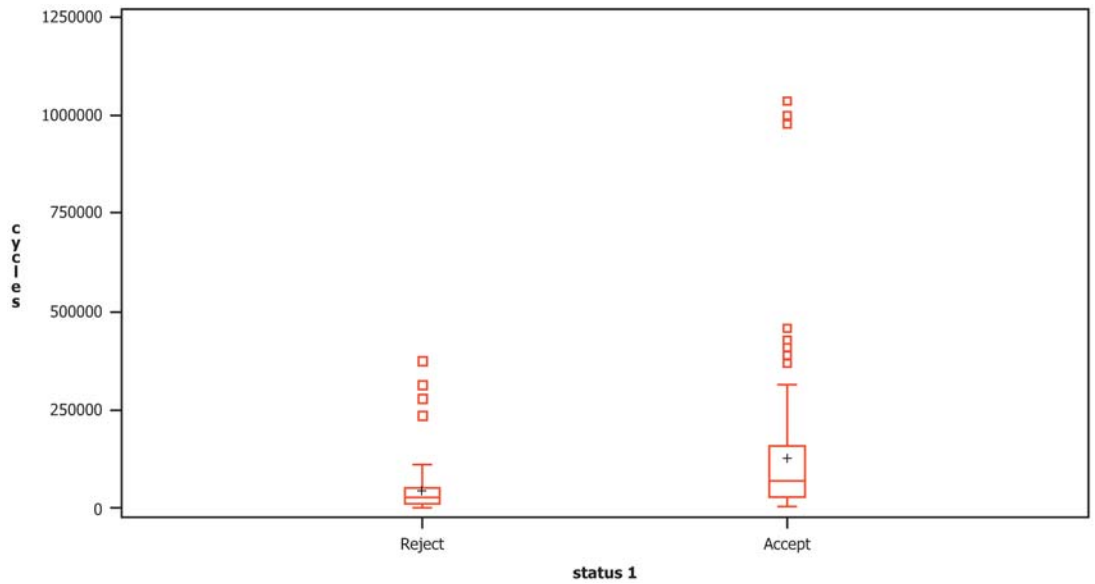


FIG. X5.2 Box Plots of Number of Cycles by Radiographic Status

X5.3.4.3 *Non-parametric Analysis*—Friedman's test was used to examine the joint effect of radiographic status and lab. Friedman's test is a non-parametric test which accounts for the effect of one variable (here, lab) after stratifying for the effect of another (here, radiographic status). The results suggest a clear effect of both variables ($\chi_5^{\ 2} = 103.84$, $p < 0.0001$).

X5.3.5 *Is it better to analyze on non-transformed or log-transformed scale?*

X5.3.5.1 Given the significant skew present in the data, analyzing on log transformed data is often recommended as a solution.

X5.3.5.2 *Parametric Analysis*—While the log-transformed values lead to more approximately normally distributed values, the results remain significantly different across site by both the parametric and non-parametric analyses. See Table X5.8 and Fig. X5.5. The one-way ANOVA results on the log-transformed

TABLE X5.4 Descriptive Statistics by Radiographic Status (log-transformed)

Radiographic Status	N	Mean	Median	Standard Deviation	Coefficient of Variation	Minimum	Maximum
Reject	80	4.29	4.41	0.73	16.96	2.01	5.57
Accept	108	4.78	4.84	0.56	11.76	3.54	6.02

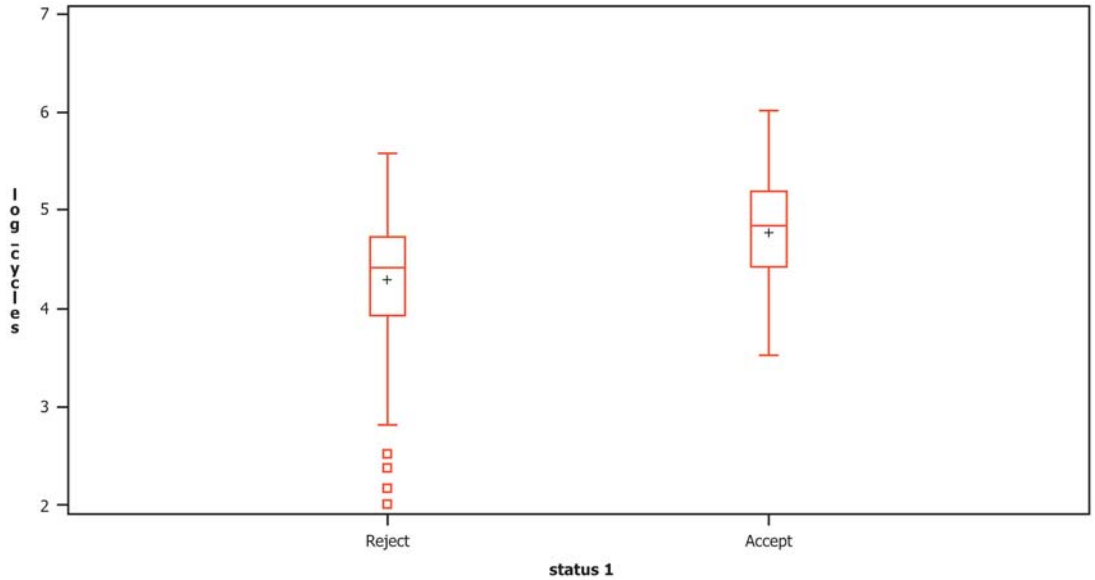


FIG. X5.3 Box Plots of Log-Transformed Number of Cycles by Radiographic Status

TABLE X5.5 Descriptive Statistics by Lab and Radiographic Status

Lab and Radiographic Status	N	Mean	Median	Standard Deviation	Coefficient of Variation	Minimum	Maximum
Lab 2 accept	32	180455.8	134825.5	177018.6	98.1	32483	1001717
Lab 2 reject	20	92183.8	702340	100142.3	108.6	238	375642
Lab 3 accept	15	51692.6	37334	27793.2	53.8	15554	121124
Lab 3 reject	15	17322.8	14478	13844.6	79.9	102	47053
Lab 4 accept	15	77385.1	75513	51862.2	67.0	16797	206635
Lab 4 reject	8	53290.1	54556.5	25887.9	48.6	23821	96713
Lab 5 accept	15	357861.5	246887	282834.3	79.0	117802	1037238
Lab 5 reject	20	56845.6	35167.5	73357.2	129.1	2500	278640
Lab 6 accept	15	18503.4	13036	17154.8	92.7	3432	71174
Lab 6 reject	4	14110.3	1844.5	25703.9	182.2	149	52603
Lab 7 accept	16	16618.56	13685	10673.52	64.23	5024	36178
Lab 7 reject	13	9530.54	7958	7854.04	82.41	644	25214

scale support a statistically significant difference between labs ($F_{5,182} = 22.24$, $p < 0.0001$).

X5.3.5.3 *Non-parametric Analysis*—As the log transformation is monotonic, the non-parametric analysis is unchanged if performed on the log scale (Kruskal-Wallis $\chi_5^2 = 88.47$, $p < 0.0001$).

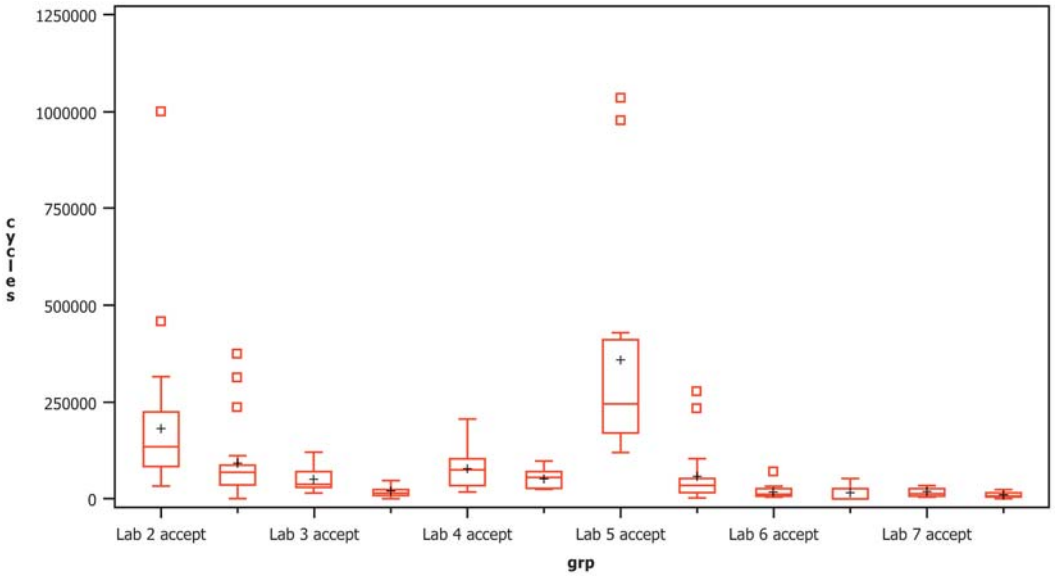


FIG. X5.4 Box Plots of Number of Cycles by Lab and Radiographic Status

TABLE X5.6 ANOVA Table for Effect of Lab, Radiographic Status, and Interaction

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Lab	5	915708971239	183141794248	13.16	<0.0001
Radiographic status	1	215418143539	215418143539	15.48	0.0001
Lab*radiographic status	5	499431046772	99886209354	7.18	<0.0001

TABLE X5.7 ANOVA Table for Effect of Lab, Radiographic Status, and Interaction (log-transformed)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Lab	5	34.94091387	6.98818277	33.80	<0.0001
Radiographic status	1	12.13659831	12.13659831	58.70	<0.0001
Lab*radiographic status	5	3.45752705	0.69150541	3.34	0.0065

TABLE X5.8 Descriptive Statistics by Lab (log-transformed)

Lab	N	Mean	Median	Standard Deviation	Coefficient of Variation	Minimum	Maximum
Lab 1	52	4.97	4.99	0.51	10.25	2.38	6.00
Lab 2	30	4.32	4.46	0.60	13.89	2.01	5.08
Lab 3	23	4.75	4.81	0.29	6.01	4.23	5.32
Lab 4	35	4.90	5.02	0.66	13.38	3.40	6.02
Lab 5	19	3.94	4.09	0.68	17.24	2.17	4.85
Lab 6	29	3.97	4.10	0.44	11.07	2.81	4.56

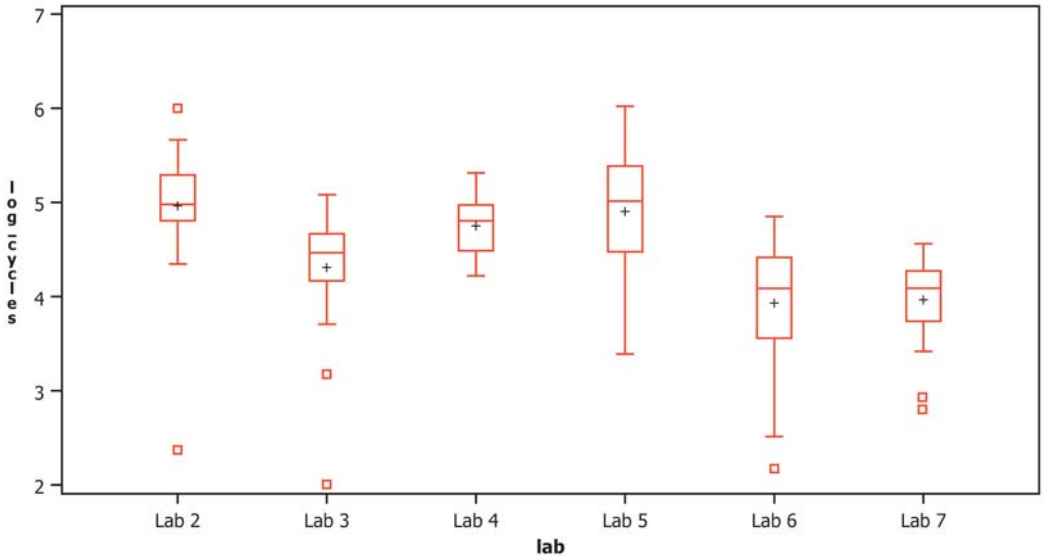


FIG. X5.5 Box Plots of Log-Transformed Number of Cycles by Lab

REFERENCES

(1) Collins, J. A., *Failure of Materials in Mechanical Design*, John Wiley & Sons, New York, 1981.

(2) Sheskin, D. J., *Handbook of Parametric and Nonparametric Statistical Procedures*, CRC Press, Boca Raton, FL, 1997.

(3) Wall, F. J., *Statistical Data Analysis Handbook*, McGraw-Hill, New York, 1986.

(4) Shapiro, S. S. and Wilk, M. B., "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, 52, 1965, pp. 591–611.

(5) Hollander, M. and Wolfe, D. A., *Nonparametric Statistical Methods*, 2nd ed., Wiley, 1999.

(6) Kusy, R. P.: "Characterization of Self-Curing Acrylic Bone Cements," *Journal of Biomedical Materials Research*, Vol 12, 1978, pp. 271–305.

(7) Johnson, J. A., "Fatigue of Acrylic Bone Cement Effect of Frequency and Environment," *Journal of Biomedical Materials Research*, Vol 23, 1989, pp. 819–831.

(8) Johnson, J. A., Provan, J. W., Krygier, J. J., Chan, K. H., and Miller, J., "Fatigue of Acrylic Bone Cement–Effect of Frequency and Environment," *Journal of Biomedical Materials Research*, Vol 23, 1989, pp. 819–831.

(9) Verdonschot, N. and Huiskes, R., "Creep Behavior of Hand-Mixed Simplex P Bone Cement Under Cyclic Tensile Loading," *Journal of Applied Biomaterials*, Vol 5, 1994, pp. 235–243.

(10) Harrigan, T. P. and Harris, W. H., "A Three-Dimensional Non-Linear Finite Element Study of the Effect of Cement-Prosthesis Debonding in Cemented Femoral Total Hip Components," *Journal of Biomechanics*, Vol 24, 1991, pp. 1047–1058.

(11) Crowinshield, R. D., Brand, R. A., Johnston, R. C., and Milroy, J. C., "The Effect of Femoral Stem Cross-Sectional Geometry on Cement Stresses in Total Hip Reconstruction," *Clinical Orthopaedics and Related Research*, No. 147, 1980, pp. 71–77.

(12) Demian, H. and McDermott, K., "Regulatory Perspective on Characterization and Testing of Orthopedic Bone Cement," *Biomaterials*, Vol 19, 1998, pp. 1607–1618.

(13) Lewis, G., "Properties of Acrylic Bone Cement: State of the Art Review," *Journal of Biomedical Materials Research (Applied Biomaterials)*, Vol 38, 1997, pp. 155–182.

(14) Topoleski, L. D. T., Ducheyne, P., and Cuckler, J. M., "A Fractographic Analysis of *In Vivo* Poly(methyl methacrylate) Bone Cement Failure Mechanisms," *Journal of Biomedical Materials Research*, Vol 24, 1990, pp. 135–154.

(15) Noether, G. E., "Sample Size Determination for Some Common Nonparametric Tests," *Journal of the American Statistical Association*, 82(398), 1987, pp. 645–647.

(16) Fan, X., Felsovaly, A., Sivo, S. A., and Keenan, S. C., *SAS for Monte Carlo Studies*, SAS Press, Cary, NC, 2001.

(17) Lewis, G. and Austin, E., "Mechanical Properties of Vacuum-Mixed Acrylic Bone Cement," *Journal of Applied Biomaterials*, Vol 5, 1994, pp. 307–314.

(18) Wixson, R. I., Lautenschlager, E. P., and Novak, M. A., "Vacuum Mixing of Acrylic Bone Cement," *Journal of Arthroplasty*, Vol 2, No. 2, 1987, pp. 141–149.

(19) Lidgren, L., Bodelind, B., and Moller, J., "Bone Cement Improved by Vacuum Mixing and Chilling," *Acta Orthopaedic Scandinavica*, Vol 57, 1987, pp. 27–32.

(20) Linden, U., "Fatigue Properties of Bone Cement: A Comparison of Mixing Techniques," *Acta Orthopaedic Scandinavica*, Vol 60, No. 4, 1989, pp. 431–433.

(21) Davies, J. P., Jasty, M., O'Connor, D. O., Burke, D. W., Harrigan, T. P., and Harris, W. H., "The Effect of Centrifuging Bone Cement," *Journal of Bone and Joint Surgery [Br]*, Vol 71B, No. 1, 1989, pp. 39–42.

(22) Casella, G. and Berger, R. L., *Statistical Inference*, 2nd ed., Duxbury, 2002.

(23) Little, R. E. and Jebe, E. H., *Manual on Statistical Planning and Analysis for Fatigue Experiments*, STP 588, American Society of Testing and Materials, 1975.

(24) Conway, J. B. and Sjodahl, L. H., *Analysis and Representation of Fatigue Data*, ASM International, Materials Park, OH, 1991.