Wear 376-377 (2017) 1830–1840

Contents lists available at ScienceDirect

Wear

journal homepage: www.elsevier.com/locate/wear

Lessons learned from the test-to-test variability of different types of wear data

P.J. Blau

Blau Tribology Consulting, P.O. Box 631, Enka, NC 28728, USA

ARTICLE INFO

Article history:
Received 8 September 2016
Received in revised form
15 November 2016
Accepted 16 November 2016
Available online 22 November 2016

Keywords:
wear
wear testing
repeatability
BOCLE
ASTM wear tests

ABSTRACT

Contrary to the established principles of the scientific method, a surprising number of experimentally-based papers submitted to tribology journals and conferences report only one test result for each material pair or set of applied conditions. However, like hardness, yield strength, fatigue life, and other material properties, wear data exhibit varying degrees of repeatability and reproducibility (R/R). Repeatability concerns the replication of experiments within the same laboratory using the same equipment and materials. Reproducibility concerns testing on different equipment, usually at a different location, but using the same lot of specimens and procedures. An important question is: How many replicate measurements are needed to validate trends in wear behavior or to relatively rank materials, surface treatments, or lubricants? Without repeatability information, it is virtually impossible to establish whether reported material rankings or the effects of variables are real or fall within normal data scatter. The purpose of this paper is to characterize and analyze the R/R of wear data that result from a variety of sources, including material homogeneity, choice of units of measure, and choice of experimental variables. Case studies compare R/R for different forms of wear and their test methods, including ASTM standards. Lessons learned are presented for five forms of wear: (1) cavitation erosion, (2) three-body abrasion, (3) solid particle erosion, (4) dry sliding wear, and (5) fuel lubricity using the ball-on-cylinder (BOCLE) test. Wear transitions can also affect R/R. These examples provide insights for validating wear models, deciding how many repeated tests to make, and when ranking wear-resistance.

1. Introduction

In 2016, a survey of the biggest problems facing science was conducted of 270 international scientists in biomedical and social sciences [1]. Results indicated that the third in a list of the top seven problems was: "Replicating results is crucial and rare." The study's authors, Belluz, et al., further state:

> "Testing, validating, retesting — it's all part of a slow and grinding process to arrive at some semblance of scientific truth. But this doesn't happen as often as it should, our respondents said. Scientists face few incentives to engage in the slog of replication. And even when they attempt to replicate a study, they often find they can't do so. Increasingly it's being called a "crisis of irreproducibility."

While the foregoing survey refers to fields other than tribology, the author's experience in editing journals and conference papers suggests that the same can also be said for tribology-related papers. That raises the question: Why do so many wear paper authors base general conclusions on only one or two tests per set of test conditions? In the author's experience, there are at least seven possible reasons for this shortcoming:

1) Investigators are inadequately trained in the scientific method or may have forgotten that repetition and verification of results are central components of good science.
2) Investigators ignore the fact that wear data, like other physical measurements, exhibit a degree of variability.
3) Experiments may be too costly or too time-consuming to repeat, even though the researcher acknowledges that repetition improves overall confidence in the data.
4) Only a small amount of experimental material is available for testing.
5) Wear is used as a metric to measure the effects of processing, material composition, or surface treatment variables. With emphasis shifted toward processing, there is inadequate attention paid to the true meaning and repeatability of the wear data itself.
6) Business or program deadlines must be met. These include conference paper deadlines, a sponsor's demanding results, project deliverable deadlines, the need to publish, to graduate the next class of students, and a need to move on to the next

E-mail address: weartester@msn.com

http://dx.doi.org/10.1016/j.wear.2016.11.012
0043-1648/© 2017 Published by Elsevier B.V.

project. Engineers are pressured to get "quick and dirty results" and to make rapid decisions on material or lubricant selection.

7) Design of experiments (DOE) approaches to wear measurement define a matrix of parameters like normal force, speed, and temperature, but researchers wrongly assume that only one test per combination of parameters accurately characterizes the net effects of each specific combination.

Physical measurements are subject to a degree of variability. Depending on the type of measurement, the sources of such variability can be attributed to the accuracy and calibration of the measuring instrument, human error, unrealized factors affecting results, influences from the surrounding environment (e.g., contamination, humidity, vibrations), or localized heterogeneities in the materials themselves. As an example of human factors and variability, the author participated in a standards-related, multi-laboratory study to study the variability of Knoop and Vickers microindentation hardness data. Several rounds of measurements were made. In a later round, four carefully prepared materials were indented on the same equipment and then sent to participants for light optical measurement of the impression lengths. It was found that the largest source of scatter in hardness numbers was not variations in the material homogeneity, preparation methods, or in the hardness apparatus, but rather a result of incorrect light-optical microscope magnification calibration and reader judgment of the length of impressions [2]. It was further realized, based on specimens of materials that varied in hardness, that the range of operator data scatter increased as the normal force decreased, especially below 100 g–f (100 mN). The lessons learned from that study was that the sources of scatter in micro-indentation hardness data include the operator, the equipment, and the material itself.

Recognition of the test-to-test variability in material wear behavior is embodied in roller bearing life prediction. In 1924, Palmgren introduced an equation that depicted the probability of 10% failure of a sample of roller bearings by rolling contact fatigue, RCF [3] After years of industry use, this approach was critically reviewed by Zaretsky [4,5]. It was reaffirmed that statistical variations are to be expected in wear life of a given material when tested by a consistent procedure, but more importantly, that factors other than contact fatigue can also affect wear and roller bearing life. Therefore, any approach that assumes that failure is exclusively by RCF (in contrast with corrosion, misalignment, and other possible root causes) can produce misleading conclusions. Microindentation hardness and RCF data are but two examples of unanticipated sources of variability in material behavior even when tests are conducted very carefully.

In order to understand the implications of scatter in wear data, two terms need to defined. The term "repeatability" refers to the replication of experiments within the same laboratory using the same procedures, materials, and operators. "Reproducibility" concerns testing on different equipment, usually at different locations, but using the same lot of materials and testing procedures. The reproducibility of wear data is important for those developing standard test methods, those whose jobs involve quality control of a manufactured product or material, and commercial firms who buy products on the basis of wear or durability requirements. In the latter case, the customer may test some fraction of the products he or she buys in order to ensure that the durability data from the supplier matches that of the delivered products when tested under similar conditions. Depending on the situation, repeatability and reproducibility (R/R) may serve different purposes in scientific studies, materials engineering, and manufacturing.

Tribology researchers who write archival-quality journal articles or submit laboratory test reports are usually more concerned with repeatability than with reproducibility. In other words, when attempting to establish certain relationships between wear results and applied parameters like normal force, speed of relative motion, temperature, or angle of erodant impingement, the conclusions drawn should consider test-to-test scatter. The lack of attention to repeatability, and attempting to draw broad, general conclusions from limited data is one reason why some papers are rejected from archival journals.

The primary R/R question becomes: "How many times should each experiment be repeated in order to properly validate the material's wear behavior? This remainder of this paper addresses that question for several forms of wear, based in particular on the development of wear and erosion standards and the lessons learned from those activities.

In an earlier paper, it was reported that different ASTM standard wear test methods exhibit different degrees of R/R for several types of wear [6]. In the current work, six case studies are presented illustrate how the wear test details (including the geometry, variables, and equipment design), the materials being examined, and the operator consistency can all affect the R/R to varying degrees. In order to understand these examples, some basic definitions and statistical terms must first be introduced.

2. Definitions of precision, bias, and a few statistical terms

All ASTM standard test methods are required to contain sections on "Precision and Bias". Basic statistics textbooks and handbooks provide formal definitions for these terms, but the concept of precision and bias can be summarized graphically. Fig. 1 utilizes shaded areas (not one unique value) to represent the wear data range expected for a given type material when subjected to specific contact conditions and measured properly. The average of that behavior is shown by a star shape in the center of the shaded areas. Individual test results are depicted as open circles. When more than one test is done, the average of those results (in this example, 6 tests) is shown by a cross. Case A shows an individual result that falls outside the true range of wear behavior. Case B shows when a result falls within it. Case C depicts six tests whose results are widely dispersed while Case D illustrates six test results that are closely clustered (more precise). Case E shows that while the results in Case D are precise, they are not accurate, and therefore, are considered to be biased. Experimentally, bias can be caused by a number of factors, such as an undetected error in an calibration of load or some other applied variable that skews test results. Even if a digital read-out or a data output is highly precise (many decimal places), its accuracy can be incorrect. Case F shows both high precision and high accuracy. Note that the average for the six experiments in Case F is very close to the average of the range in true wear behavior for that material. When only one wear test is done for a given set of imposed conditions and materials, it is not possible to determine its degree of precision or its bias.

As will be discussed later in this paper, the probability of observing a given value for the wear or wear rate is likely vary within some characteristic range for that tribosystem. In other words, each shaded circle in Fig. 1 represents a distribution of valid measurements, any of which could be considered an accurate measure of the behavior of the materials in that tribosystem.

ASTM has published a number of standards concerning the analysis and reporting of R/R data from interlaboratory (a.k.a. "round-robin") test programs. Basic statistical standards such as ASTM E691-15 [7] have been adapted to create ASTM G117-13 [8] for use in interlaboratory wear testing. These calculations are used to assess precision and bias in standards. Several important statistical parameters and nomenclature are given in Table 1. This paper does not purport to treat the rigorous subject of measurement uncertainty, but rather intends to illustrate how statistical data from specific wear tests can provide guidance to the engineer or modeler.

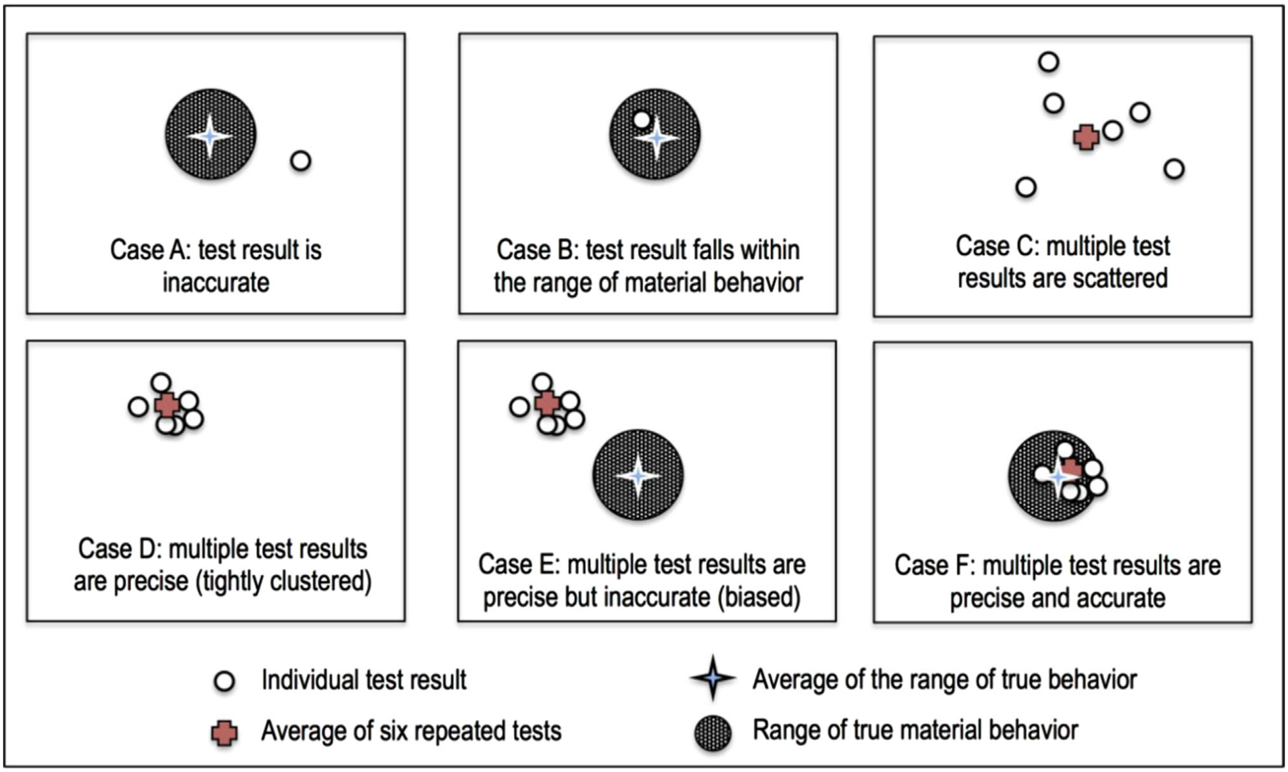

Fig. 1. Simple illustration of precision, accuracy, and bias of numerical data.

Table 1
Common measures used in determining repeatability.

Symbol	Name	Explanation
x_i	measurement *i*	An individual measurement
x_{ave}	average	Average of a set of measurements
n	number	Number of individual measurements made in the data set
X	average	Based on *n* individual measurements, this is the average or mean value
X_{lab}	within lab average	Average of all measurements made within a given laboratory
X_{all}	average of measurements by all labs	Average of measurements made by all laboratories participating in a study
d	deviation	The absolute value of the difference between a given measurement *x* and the average of all *n* measurements X, $d = \text{abs}(x\text{-}X)$
S	standard deviation	[see the footnote 1]; a measure of the spread of a set of *n* values of *x*
S_{lab}	within-lab standard deviation	The standard deviation for data obtained in the same laboratory
S_{all}	between-lab standard deviation	The standard deviation for data obtained by all participating laboratories
V	variance	the quantity inside the square root symbol in Eqn. (1)
CoV	coefficient of variation	ratio of the standard deviation of a set of *n* data to its mean value, commonly expressed in percent, viz. [CoV (%) = 100*(S/X)]
CoV_{lab}	within-lab coefficient of variation	The coefficient of variation of all measurements within a given laboratory
CoV_{all}	between-lab coefficient of variation	The coefficient of variation of measurements by all participating laboratories

Footnote 1: $S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - x_{ave})^2}$

3. Lessons learned from the repeatability and reproducibility of wear test data

Five case studies are presented here to illustrate lessons learned from the R/R of wear data. Several forms of wear have been chosen to show that different forms of wear exhibit different sensitivities to test methodology. Data to support these studies is drawn from the literature of standardization since repeatability and inter-laboratory tests are commonly a part of the standards development process. The forms of wear are:

Case Study 1: Cavitation wear
Case Study 2: Three-body abrasive wear
Case Study 3: Solid particle erosion
Case Study 4: Dry sliding wear
Case Study 6: Sliding wear as an indicator of fuel lubricity

Case *Study 1: Effects of the choice of wear criteria on ranking and confidence of cavitation erosion data.* Cavitation erosion is attributed to the creation and collapse of bubble fields adjacent to a surface. High, local pressure spikes due to the jetting effects of bubble collapse make cavitation erosion a serious concern in the wear of pumps, ship propellers, and pipes. Standard test method ASTM G32 was the first wear standard developed after the founding of ASTM G02 on erosion, the precursor of the ASTM Committee on Wear and Erosion. Standard G32 reads like a practical tutorial in cavitation

Table 2
Interlaboratory test results from cavitation erosion tests of Ni 200 alloy (source: [9]).

Lab no.	MER average (μm/h)	MER S (μm/h)	MER CoV (%)	IP average (min)	IP S (min)	IP CoV (%)	$t_{100\ \mu m}$, average (min)	$t_{100\ \mu m}$, S (min)	$t_{100\ \mu m}$, CoV (%)
1	29.6	0.88	3.0	29.7	6.8	22.9	234.	4.6	2.0
2	27.6	0.66	2.4	19.0	2,7	14.2	236.	4.5	1.9
3	23.5	0.14	0.6	18.3	2.5	13.7	275.	4.5	1.6
4	26.0	1.90	7.3	19.7	3.5	17.8	248.	24.7	10.1
Ave.	26.6	0.90	3.3	21.7	4.3	17.2	248.	9.8	3.9

testing and contains R/R data from several inter-laboratory studies [9]. The standard involves screwing a 15.9 ± 0.05 mm diameter button specimen onto the end of a shaft that rapidly vibrates (20 ± 0.5 kHz) with a 50 μm amplitude, and then periodically measuring weight loss and damage depth. A plot of wear loss versus time can be used to determine the maximum erosion rate (MER), the incubation period (IP) to the onset of erosive wear, and the time (t) to reach specific depths expressed in μm.

Results from an inter-laboratory study of commercially pure nickel, involving four participants, are shown in Table 2. Each laboratory conducted each tests three times and the data for three out of four participants turned out to be relatively consistent for a Ni200 reference alloy in terms of the CoV. Laboratory 4 data had significantly more scatter than that of the others. Other analyses such as the Student-T test [10] can be applied to determine if the extraordinary data fall within the same population, but those details are beyond the scope of this paper. Basically, the results illustrate how reproducibility data can be biased by consistency in only one of the participating laboratories, especially if a small number of participants like four is involved (ASTM statistical experts recommend a minimum of 6 laboratories for a valid inter-laboratory program). Note that the three different measures of erosion (i.e., MER, IP, or $t_{100\mu m}$) compared in Table 2 have different CoV's.

To better visualize this effect for two additional materials, Fig. 2 compares the mean erosion rate, the incubation period and the time to reach 100 μm depth from the ASTM G32-10 standard [9]

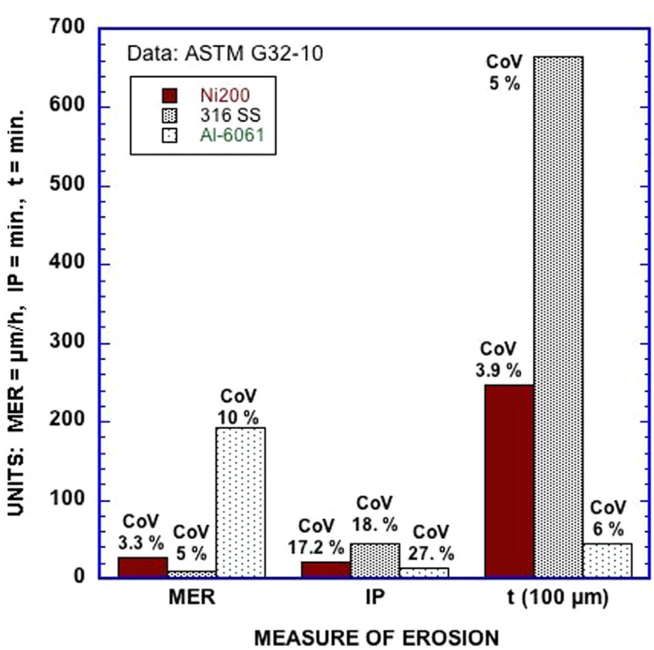

Fig. 2. Comparison of cavitation erosion of Ni200, AISI 316 and Al 6061-T6 using three different measures of erosive wear severity. The ranking of the materials changes depending on the criterion chosen to base the comparison and the test-to-test Coefficient of Variation (CoV) varies with the chosen measure as well. Data from ASTM [9].

for three alloys: Ni200, AISI 316 stainless steel, and Al alloy 6061-T6. Each alloy was tested in four different laboratories. The coefficient of variation of the data (%CoV) has been annotated on this plot. Note that (1) the metric chosen affects the relative ranking of erosion behavior between the materials, and that the uncertainty in the incubation period is considerably higher than that for the other quantities measured. Therefore, the number of repeated runs should be increased if the unit used to measure wear has innately a larger test-to-test variability.

Using other measures for cavitation may be more relevant to specific applications, so the foregoing illustration is not intended to be a complete comparison of cavitation methodology. It simply indicates that the scatter in some wear quantities is greater than others and that the relative ranking of wear behavior depends on the metric chosen, along with its inherent scatter.

We note that there are significant differences depending on which quantity is used to quantify wear in cavitation. The lesson from Case Study I then, is that the relative rankings of wear can be affected quantities used to benchmark wear. This lesson is not limited to cavitation erosion testing. Within the broader standards community, a range of different metrics for wear of various kinds have been used. The author's study of fifteen ASTM standards for different forms of wear indicated a preference for mass loss, but wear scar dimensions were commonly used as well [11]. German DIN standard 50321 lists 15 different measures for wear [12]. To the uninitiated, it may seem ironic that so many measurement options are available within a standards community that should be focusing on developing a commonality of approach and the ability to compare results using a set of standard metrics.

Case *Study 2: Comparison of the repeatability of three-body abrasive wear procedures using the same material*. ASTM G65 in dry sand/rubber wheel abrasive wear [13] is among the most popular wear testing standards. A rubber-covered wheel, 229 mm in diameter, rotates against a test specimen pressed against it from the side as sand is fed into the interface from the top. This standard contains five optional procedures, lettered A through E, having different normal forces and total number of wheel revolutions. Procedures are selected based on the type of material to be compared. In this example, consider Procedure A that uses 130 N and 6000 rev, Procedure B that uses 130 N and 2000 rev, and Procedure D that uses 45 N and 6000 rev. Wear is determined by weight loss converted to volume using the material density.

The amount of wear and the standard deviation of a set of data for an abraded 17-4 PH stainless steel was reported in the standard [13] using procedures A, B, and D. Fig. 3 shows how the average wear and standard deviation for the same material are affected by test procedure. For convenience, the CoV was also shown in the plot. Basically, the average wear volume scaled with the severity of the test and the CoV was similar for all three procedures. In terms of lessons learned, the selection of testing procedure did not significantly affect the repeatability of the test method even though the total wear increased with procedure severity.

Case *Study 3: Effects of test conditions on the repeatability and reproducibility of solid particle erosion results*. The ASTM G76 test [14] is used to measure the resistance of materials to impingement by airborne solid particles (50 μm alumina particles) normal to a

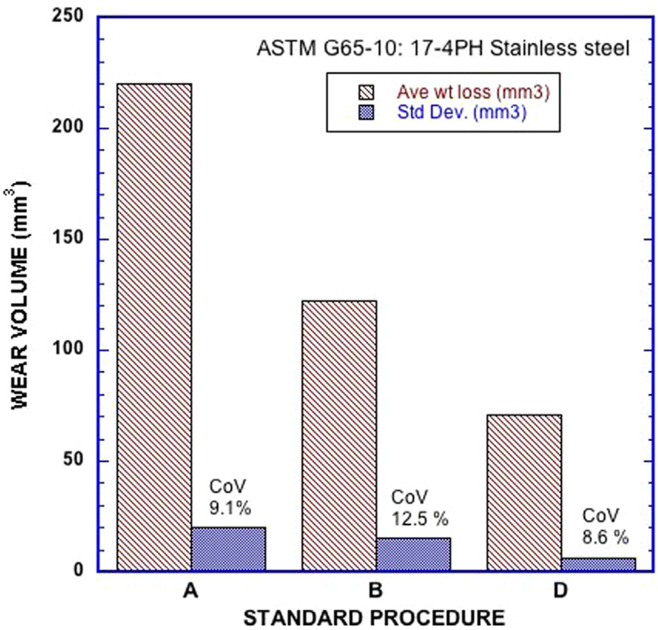

Fig. 3. Effect of testing procedure on the dry sand abrasive wear of 17-4 PH stainless steel. The coefficients of variation for each test procedure are similar, but the amount of wear loss differs. (Data from Ref. [13], Table X1.2).

Table 3
Repeatability and reproducibility of erosion test data for three steels. (data in units of 10^{-3} mm^3/g of erodent).

Material: Impingement velocity:	AISI 1020 30 m/s	AISI 1020 70 m/s	AISI 304 70 m/s
Erosive wear rate (0.001mm^3/g), all laboratories	2.734	28.16	32.60
Within-laboratory standard deviation (0.001mm^3/g) (repeatability)	0.468	0.969	1.597
Between laboratory standard deviation (0.001mm^3/g) (reproducibility)	0.807	4.786	6.786

surface at a velocity of 30 +/− 2 m/s. Those who use it routinely have commented that it is a particularly severe test and can erode most structural materials. Inter-laboratory erosion test data for three steels and from five participating laboratories (4–10 replicates per laboratory) are given in the standard. These are summarized in Table 3 where the units of wear and standard deviation are in 10^{-3} mm^3/g of erodant. Note that in two cases, the velocities are higher than the standard's recommended 30 m/s.

In all three sets of data, the reproducibility among the five laboratories was worse than was the average repeatability within each laboratory. That general trend is often observed when comparing within-laboratory to between-laboratory data. For the AISI 1020 at 30 m/s, reproducibility was about 1.7 times greater than its repeatability but when the velocity increased to 70 m/s, the difference between repeatability and reproducibility was nearly a factor of 5 worse. The first lesson from this is that the choice of wear testing conditions (such as impingement velocity in erosion) can significantly affect the degree of agreement between wear data obtained in different laboratories.

Increasing the velocity from 30 to 70 m/s for the 1020 steel increased its wear by roughly a factor of ten but roughly halved its repeatability (i.e., increased the scatter). On the other hand, increasing the velocity decreased the reproducibility by a factor of about six. A change in the imposed test conditions, velocity in this case, may have a smaller effect on the repeatability of the data than it has on the overall magnitude of the wear rate. The second lesson is that increasing the severity of imposed conditions does not necessarily raise the scatter in the results in the same proportion.

Comparing the AISI 1020 and AISI 304 stainless steel under the same erosion conditions (70 m/s), the erosion rate of the stainless steel was only about 16% greater, but the repeatability and reproducibility were worsened about 40 and 65%, respectively. The third lesson from Cast Study 3 is that the type of material tested can have a greater effect on inter-laboratory reproducibility than it has on repeatability within a laboratory.

Case *Study 4: Repeatability in sliding wear*. In addition to abrasive wear, dry or lubricated sliding wear is probably the most researched and tested forms of wear. Unfortunately, as shown in a previous paper from the proceedings of the 1999 Wear of Materials Conference [6], dry sliding wear data also displays among the greatest variability from test-to-test, and therefore, more duplicate runs are needed for sliding wear than for most other forms of wear. Two examples of sliding wear repeatability and reproducibility are provided here. The first once concerns the method used to measure wear in block-on-ring sliding and the second addresses some sources of scatter in common pin-on-disc testing.

The flat block-on-rotating ring geometry is a common one used in sliding wear testing. Results of a short study that compared three methods for measuring wear on block specimens were reported earlier [11], but can be used to illustrate differences in repeatability. A series of four wear tests were conducted using a CDA 638 bronze (95 wt% Cu, 2.8 wt% Al,1.8 wt%Si, 0.4 wt% Co) block sliding dry on an AISI 52100 test ring at 10 N load, 0.2 m/s sliding speed, and for 1 km sliding distance in air. Wear was measured independently in three ways: (i) by mass loss converted to volume using density, (ii) from measured scar width by light optical microscope, and (iii) from measured scar depth by profilometry, the latter two being converted to volume by assuming a cylindrical scar geometry with a radius of curvature equal to that of the test ring. Fig. 4 summarizes the results of these four tests and three methods of measurement. Estimates of volume obtained from weight loss were consistently lower than other measures used for wear volume. Were it not for one anomalous measurement, scar width and depth would be in very close agreement. In fact, three of the scar width-based measurements had a close grouping. The central question remains

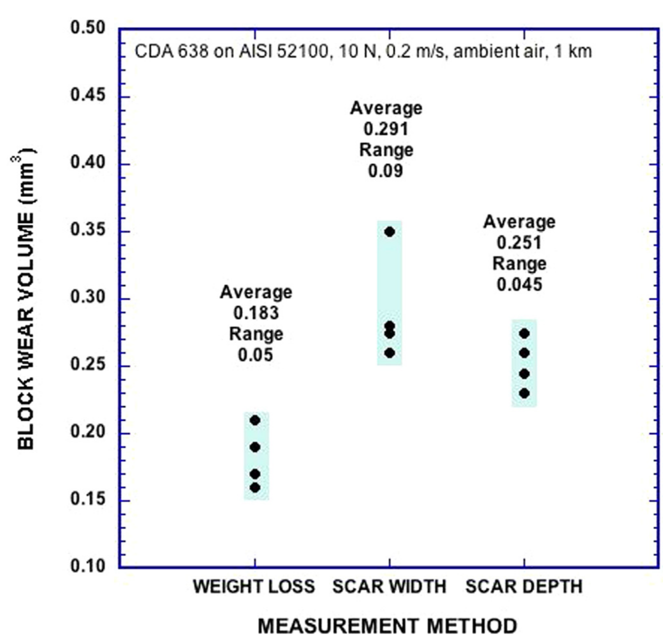

Fig. 4. Comparison of calculated wear volumes for four block on ring tests of Cu alloy CDA 638 on steel using three methods of block wear.

however: What is the accurate (correct) wear volume of each scar? Unfortunately, direct volume measurements using techniques like scanning mechanical profiling instruments and the like were not available, so a bias for the data obtained using these three methods could not be established. However, the lessons learned from this work are (i) that wear rates obtained using different methods to obtain volume in different studies should not be confidently compared, (ii) that weight loss data seem to given lower wear than dimensional measurements when converted to volume. Note that the weight conversion assumes that the density was correct for the block material, and that all the wear debris was that material (not oxides that might have a different density).

For better or worse, the pin-on-disc test method is one of the most common methods employed in sliding wear studies. A recent review of papers submitted to *Wear* during a six-week period in the spring of 2016 revealed that 31.5% of those papers involved use of a pin-on-disc test method [15]. Of those papers, nearly 37% of them claimed to be using ASTM standard test method G99 [16]. Unfortunately, a significant portion of those papers were declined because the authors failed to justify the reason for selecting that test method, did not justify the choice of the sliding partner material or applied testing parameters, or importantly for the current discussion, failed to conduct sufficient replicate tests to confirm their conclusions. Any archival experimentally-based paper should include consideration of the repeatability of the data used to justify its primary conclusions or any modeling efforts contained in that work.

There have been a number of studies published on the factors that influence pin-on-disc data repeatability. For example, the effects of using scar size versus mass loss was studied by Gee for dry sliding wear of self-mated alumina ceramics [17] Mass loss was the worst metric of the two, with a repeatability error of 42%, and wear scar size measurements were much more repeatable (5%).

When reporting normalized sliding wear factors such as K (mm^3/N-m), using wear volume (V), load (P) and distance slid (X) as follows:

$$K = \frac{V}{PX} \tag{1}$$

There are errors associated with each measurement: wear volume (ΔV), load (ΔP) and distance slid (ΔX). When multiplying errors as in the denominator of Eq. (1), one adds the relative errors, and when dividing, the relative error of the numerator minus that of the denominator. Further information on this is available from statistics texts, but the point is that measurement or equipment calibration errors can compound when computing sliding wear factors, and they degrade the repeatability of the results. When reporting only raw data such as depth of wear, one need contend with a smaller subset of possible errors.

Not only calibration, but the design of tribometers can affect repeatability and the magnitude of sliding wear measurements. Gee also studied the role of pin on disc loading system configuration and tester dynamics on the variability in wear and friction data for self-mated 95% dense alumina [18]. He reported that nearly two orders of magnitude difference in wear rate could result from the mechanics of the testing machine due to its influence on the operating wear mechanisms for the same pairing of materials. Significantly, with regard to reproducibility between laboratories, Gee states:

> "The results show that great care must be taken in the interpretation of the results of wear tests… the differences in the behaviour may come not from any intrinsic differences in the wear properties of the materials examined, but may simply be a function of the large differences in the design of the test systems used."

Furthermore, he adds:

> "The lesson must be taken from these experiments is that to derive a sufficient understanding of the wear behaviour of a set of materials it is important to perform enough tests to determine the domains of the dominant wear mechanisms."

Probably, the most comprehensive inter-laboratory test project for dry sliding wear was conducted under the auspices of the Versailles Agreement on Materials and Standards (VAMAS), and results were published in *Wear* [19]. Involving over 30 participating laboratories in seven countries, pin on disc test were conduced using both alumina and bearing steel (AISI type 52100). Combinations involved the four conditions of ball/disc as follows: steel/steel, ceramic/steel, steel/ceramic, and ceramic/ceramic using one set of sliding conditions (10 N, 0.1 m/s, 10 mm ball diameter, 1 km sliding distance, ambient temperature). From 3 to 5 tests were done for each set of materials and conditions in each laboratory. Wear scar width on the disc did not prove a satisfactory measure for all combinations, but ball scar size data can be compared. For this discussion, a subset of results for steel/steel (known as VAMAS Kit 1) and steel/ceramic (known as VAMAS Kit 3) are presented in Table 4. Table 5 summarizes the repeatability and reproducibility of the wear data. As in Case Study 3 for erosion data, the within laboratory variation for wear was less than the between laboratory results. One lesson learned here is that for these materials and conditions, measuring the system wear rate by the slope of the displacement versus sliding distance curve has much higher test-to-test variability than simply measuring wear scar diameter after the test. Like the example for cavitation provided in Case Study 1, the choice of sliding wear measurement metric affected the repeatability of the results.

Typically, investigators of sliding wear behavior rarely conduct a sufficient number of replicates to develop a statistical distribution of wear factors, but there was a relatively rare study presented by Wallbridge and Dowson at Wear of Materials 1987 [20] in which the distribution of wear factors (in units of mm^3/Nm) for multiple repeated experiments of the reciprocating sliding wear of ultra-high molecular weight polyethylene (UHMWPE) pins against stainless steel (AISI 316) plates was analyzed. Six different pin-on-flat sliding experiments were conducted under two dry conditions and four wet conditions (distilled water), respectively. The number of wear of wear factor measurements varied from 36 to 131 per experiment. It was found that the distribution of measured wear factors (k) could be well-represented by a log-normal distribution of the form:

Table 4
Selected Results from the VAMAS International Round-Robin Pin-on-Disc Testing Program [19].

Quantity	Steel ball on steel disc	Steel ball on alumina disc
Number of system wear measurements (pin displacement versus sliding distance)	47.	29.
Average system wear rate and standard dev. (pin displacement versus sliding distance)	70. ± 20. (μm/km)	81. ± 29. (μm/km)
Number of pin wear scar measurements	102.	60.
Average pin scar diameter and standard dev. (mm)	2.11 ± 0.27	2.08 ± 0.35

Table 5
Repeatability and Reproducibility of VAMAS International Round-Robin Wear Data [19].

Quantity	Range in repeatability (%)	Range in reproducibility (%)
System wear rate (μm/km)	± 14	± 29 to ± 38
Ball scar wear (in mm)	± 5 to ± 7	± 15 to ± 20

$$f(k) = \frac{1}{k\sigma\sqrt{2\pi}}exp\left[\frac{-1}{2\sigma^2}(\ln(k)-\mu)^2\right] \tag{2}$$

where σ is an empirical constant that defines the shape of the curve and μ is the geometric mean of the data. The arithmetic mean of the experimental data was compared to the arithmetic mean of the lognormal distribution to determine to what extent there was a different in simply taking average and when using a better-representative log-normal data fitting approach. Using selected data from Wallbridge and Dowson [20], the percent difference between using a simple average and using the more rigorous log-normal distribution is shown in Table 6. Fortunately, no significant error was introduced by simply averaging the wear data and compared with a more rigorous a log-normal analysis procedure. Yet, the conclusions of the work take some issue with that assumption and state: "Wear factors show considerable scatter, often over several orders of magnitude, even when experimental conditions are unchanged." Furthermore: "Scatter in wear factors for ceramic, polymers, and metals can be well-described by a log-normal distribution, which suggests a large number of low wear factors and a few higher ones."

If this latter assertion is true, then reporting wear factors based on measurements from only one or two sliding experiments does not adequately sample the behavior of a material pair. In fact, if most results tend to fall at the lower end of an asymmetrical distribution, but if larger values in the tail of distribution are still possible (what could happen, even if unlikely), then for this reason alone, tribo-component designers should include higher factors of safety when selecting materials based on experimentally-measured wear factors. Unfortunately, comprehensive studies of wear factor variability such as that of Wallbridge and Dowson are rare, and it cannot be assumed that all sliding wear data fits log-normal distribution. Therefore, that study only indicates that it is possible that wear data will exhibit other than a normal distribution, and it reinforces the need to conduct multiple repetitions.

During the early development of ASTM G65, a dry sand/rubber wheel abrasion test, the U.S. National Institute of Standards and Technology developed a heat-treated, tool steel wear calibration block and sells it as Standard Reference Material 1857. In view of the popularity of pin on disk testers, it might seem worthwhile to explore whether standard reference materials could be produced for them as well. However, in light of the VAMAS results [19], it is not clear that even highly homogeneous materials would produce similar wear data on pin on disc machines, especially those built by researchers with limited tribometer design experience. The best one can do then, is to characterize and calibrate the testing machine and materials as rigorously as possible, verifying repeatability with reference specimens from the same source, retained in the laboratory and suitably protected from corrosion and contamination.

Case *Study 5: Ball-on-cylinder, lubricant evaluation (BOCLE) test*. It can be said that the results of a wear or friction test cannot be truly understood without an understanding of the test method itself. The final example concerns the development of ASTM D5000 [21] a widely-used standard in which the diameter of a wear scar produced by a ball sliding on a rotating cylinder is used to evaluate the relative lubricity of aviation fuels. Introduced in the 1960s, the geometry and purpose of the ball-on-cylinder lubricity evaluator gives the test its acronym: BOCLE. One important concern is whether scar size measurements are sufficiently precise (sensitive) and repeatable to be capable of distinguishing the subtle effects of various fuel additives or external contaminants. A 2014 report by the non-profit Coordinating Research Council Ltd. (USA) [22] describes the BOCLE test and twelve methods that have been used to assess lubricity and a tendency for scuffing in fuel system components. Among the extensive results, there was a set of data from a survey of BOCLE test results from samples taken in four regions of the world. Results from this study led Rolls-Royce Ltd. to support following a British Defense Standard to establish testing norms for the lubricity of aviation fuels.

The average (X_{all}) and coefficient of variation (CoV_{all}) of wear scar measurements using Rolls-Royce data for a study done in 1984–1986 are shown in Fig. 5. The dashed line represents the average of data from all four regions, and the shaded box in the figure is ± 1*S* (standard deviation) for the entire data set. Several interesting trends can be observed. For example, the scatter (% CoV) varied considerably among regions, in which the USA had the lowest (albeit the number of data was also the smallest). No single reason should be inferred for the results in Fig. 5 because the method of sampling, the environment to which the fuel was exposed before collection, and the original source of the fuels must be considered as potential influences. Furthermore, the scatter inherent in the BOCLE method had to be considered, as did specimen preparation and chamber environment during testing. Therefore, it is worthwhile to consider further, some sensitivities that were revealed during the early development of the BOCLE method.

The early development and application to practice of the BOCLE test (ASTM D5001) included considerations of its repeatability and reproducibility as affected by the specimen preparation and the procedures. This work was described in studies by Biddle et al. [23] and later by Lacey [24]. The 1987 study by Biddle et al. [23] described running 180 experiments on sets of three test rings each from four production lots for those rings in five different fuels and fluids. Thirty-six data points were obtained for each fuel sample.

Table 6
Comparison of Arithmetic Mean of Experimental Wear Factor Data with the Arithmetic Mean of a Log-normal Wear Model. (UHMWPE pin sliding on AISI 316 Flat spec with various surface finishes, data from Ref. [20]).

Condition (surface finish, μm Ra)	Mean of experimental data (x10^9 mm^3/N-m)	Mean of log-normal fitting to the data (x10^9 mm^3/N-m)	Difference between data mean and log-normal (%)
wet (< 0.01)	142	139	2.1
wet (≤ 0.01 to < 0.05)	342	344	−0.6
wet (≤ 0.05 to < 0.1)	644	656	−1.9
wet (≤ 0.1 to < 0.32)	1460	1442	1.2
dry (≤ 0.01 to < 0.05)	126	127	0.8
dry (≤ 0.05 to < 0.1)	68	68	0

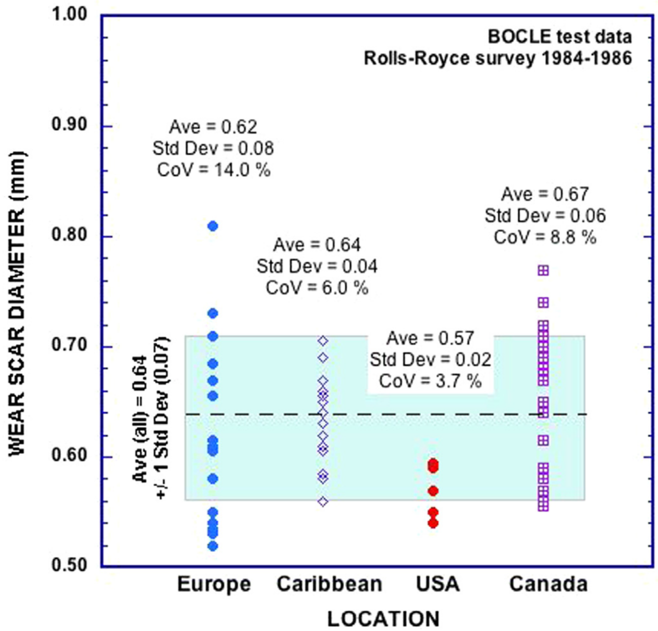

Fig. 5. Analysis of fuel lubricity data from a study BOCLE tests sampled in four regions of the globe. Data from Ref [20].

Surface finish began to emerge as a key factor in repeatability. A round-robin (interlaboratory testing) program, described by Lacey [24] later established that tests performed at 10% relative humidity (RH) in the test chamber had higher precision than those performed at 50%RH, and a comparison with fuel pump wear severity ratings could be made based on a critical BOCLE scar diameter of 0.62 mm for maximum acceptable pump wear.

The BOCLE test was later extended to measure the scuffing characteristics of the fuels. The test fluid was placed in a reservoir at 50% relative humidity with a half-immersed rotating polished AISI 8720 steel cylinder approximately 50 mm in diameter. A 12.7 mm diameter ball of AISI 52100 steel was pressed against it with increasing load until a change in friction (friction coefficient exceeded 0.175) and wear is observed. The minimum load to produce that transition was used as a guide to the scuffing load capacity in a fuel injection system. Like the earlier work by Biddle et al. [23], one lesson learned was that repeatability in scuffing load data was highly dependent on the surface roughness of the cylinder specimen (test ring). If the root mean square roughness (R_q) was much above 0.2 μm, there was little difference in results for three test fuels but when the ring was polished to about R_q=0.04 μm the difference in fuels was optimized. Using that method of specimen preparation, a series of BOCLE scuffing tests (ASTM D5001, Procedure D, 10% RH), was conducted to establish the repeatability within one laboratory versus the reproducibility between three laboratories. The lessons learned were that two factors (ring surface finish and test chamber relative humidity) needed to be controlled in order to improve the sensitivity, repeatability, and reproducibility of a sliding wear test method to usefully measure the lubricity and scuffing characteristics of steel in aviation fuel environments.

4. Discussion

Unlike simpler, more directly controlled mechanical tests, like indentation hardness tests and tensile tests, wear tests tend to be affected by more a more complex combination of tribosystem variables. Considering so many potential influences on results, a certain degree of test-to-test variability is not unexpected. In fact, there may not be one unique ("true") value of wear coefficient or wear rate to typify the system of interest; rather for each tribosystem, operating conditions, and set of materials, there is likely to be a range or distribution of characteristic wear responses. This typical range in wear data due to small, possibly unrecognized perturbations in the tribosystem and materials would be revealed after a sufficient repetition of experiments.

The foregoing case studies provide a series of lessons on the effects of test method, material heterogeneity, specimen preparation, humidity and the measurement metrics selected to measure wear. The need for repeating experiments in order to strengthen confidence in the results should be evident from this work, especially where the normal scatter of test-to-test data is significant. Furthermore, it should be evident that scatter varies for different forms of wear and is greatly affected not only by the test method selected but also by which metrics are used for wear.

Material heterogeneity is a concern in all foregoing examples of laboratory testing. It can be a major contribution to the variability in wear data if the microstructural features in the material are similar in size to the wear contact. For example, if a large pore, inclusion, recrystallized zone, or clump of hard particles comprises a significant fraction of the wear scar, then those features can disproportionately affect the wear within that area. The positioning of the specimen in the testing apparatus may expose such heterogeneities in the wear location. This concern is analogous in some ways to the Indentation Size Effect (ISE) in hardness testing [25]. When a hardness test is used on a composite material, the amount of material that is deformed by the indenter can be comparable to the sizes of features such as second phases, pores, and additive particles. Analogously, when one uses a small, laboratory-scale tribometer with a contact area size the order of a few mm^2, he or she in essence samples only a small volume of material relative to the bulk, and conclusions about wear resistance are dependent on only a tiny sample of material worn away. A comparable Wear Area Size Effect (WASE) can result in data scatter if there are microstructure and compositional variations within the material being worn away.

During non-conformal sliding wear testing, such as sliding a sphere on rotating disc or a flat block against a ring, the contact area expands more quickly at the beginning of a test (e.g., during running-in). The effects of this process in constant load testing is that nominal contact pressure decreases with time, but the area grows to contain a larger portion (sample) of the material being tested. Therefore, when testing heterogeneous materials, there may be more variability in the wear-in stage than in the steady-state stage because a smaller area is controlling the wear (and friction) behavior than later in the experiment.

Some wear testing situations involve larger sampling areas, making the WASE is less important. Examples include automotive disc brake pads or a wear plate in a jaw crusher in a quarry or mine. Then the effects of local variations in material tend to be averaged out, and other sources of variability like consistency of procedure or precision of the measurement method may become more important.

There has been a great deal published on specific factors that could affect the variability and scatter in wear data (e.g., [26–29]. Additional factors include specimen preparation methods, cross-contamination from adjacent equipment, calibration errors, technician training, human judgment, and many more. Therefore, the case studies discussed here were selected to support the case for repetition. Economic concerns about testing cost, publishing deadlines, and time pressures on engineering decisions have made it difficult to replicate wear tests sufficiently to establish the repeatability of wear data and to validate new wear models. Particularly in the case of research papers being submitted to journals in recent years, conclusions are too often based on only one or two

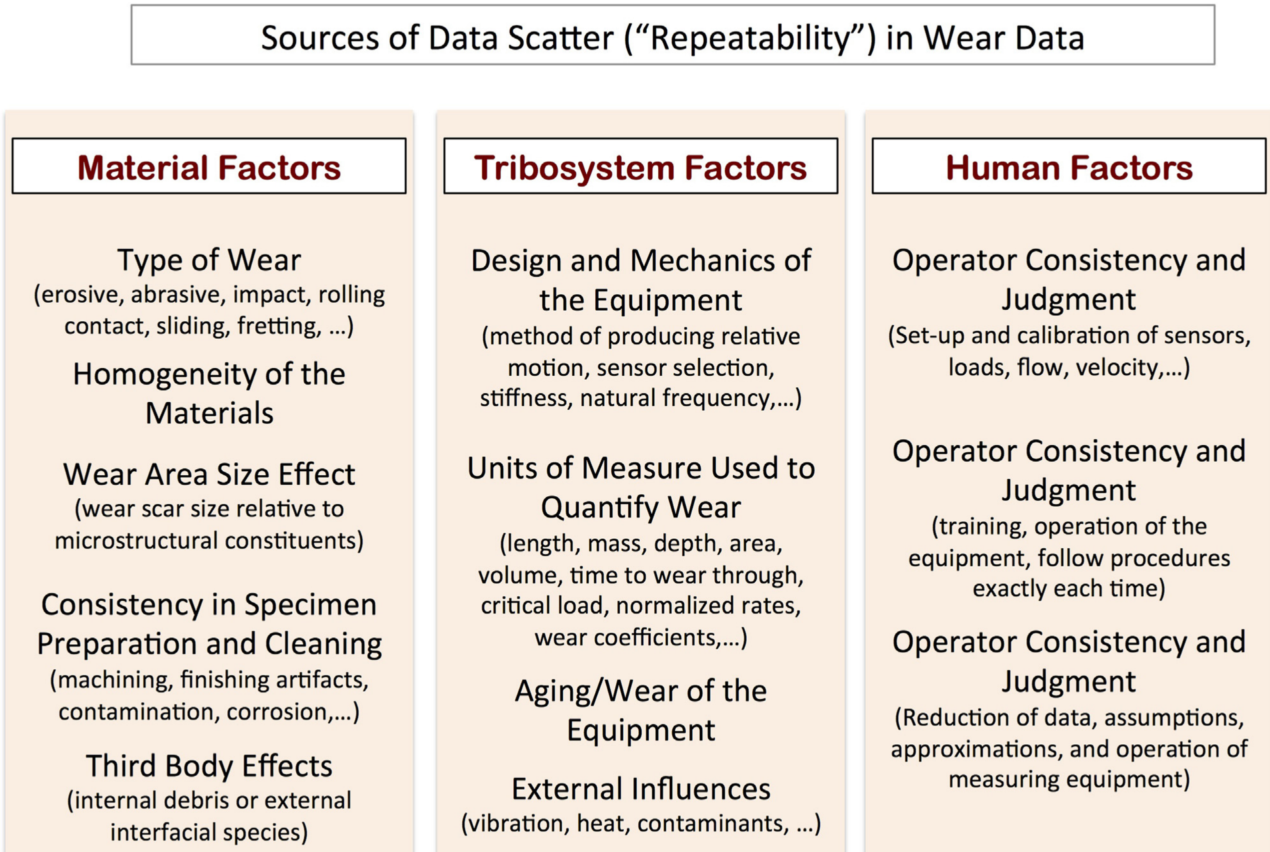

Fig. 6. Summary: sources of repeatability and reproducibility in wear testing.

test results. By contrast, private industry may conduct numerous product wear tests for quality control, but such extensive data are not usually submitted to technical journals. One notable exception to this is the publication of statistical life factors for rolling element bearings in commercial product catalogs.

To summarize, a depiction of the influences on repeatability and reproducibility in wear data, including those discussed earlier in this paper and mentioned in other studies, are represented in Fig. 6 and classified by material factors, tribosystem factors, and human factors. That is, the scatter observed in wear data are due in part to intrinsic differences in the materials or substances themselves, but also to limitations in the methodology used to detect, measure, and quantify behavior, and last in the limitations of design and the consistency of those conducting the experiments. The number of repetitions of wear results should ideally be sufficient to establish a statistical distribution of results for each set of conditions. Lacking that degree of time and resources, more than one test, preferably three or more, should be conducted along with an qualifying statement concerning the expected scatter for those results and its basis.

As a final point, the subjects of precision, bias, and accuracy of measurement must be put into the context of wear transitions. Transitions in the rate of wear may occur in some tribosystems when the contact conditions reach certain critical values (e.g., critical load, critical speeds, critical temperatures). Well-known transitions include running-in and changes in wear rate due to the wear-through of coatings. Erosive wear, with its incubation period is another example [9]. Changes in wear rate have been taken into account in multi-stage models (e.g., [30]) and were discussed at the 2015 Wear of Materials conference (e.g., [31]). Therefore, when measuring wear, one needs not only consider the repeatability of measurements, but also the repeatability of wear transitions. If transitions are known to occur, tests to determine nominal wear behavior should not be conducted too near the boundaries of those transitions in parameter space. An exception to this point is when the transitions themselves are the subjects of study.

The presence of wear transitions raises a further question: What are the effects on apparent wear rates from time-dependent transitions? A tribosystem that is prone to transitions must consider the period of time spent at each stage of wear life. As a simple example for sliding wear, consider a tribosystem operating at constant load P and velocity v in which the first period involves wearing through a coating (t_1) and the second, wearing down into the substrate (t_2). If the wear rate of periods 1 and 2 (W_1 and W_2, respectively) are assumed to be linear with time, then as a first approximation, the wear volumes for each stage (V_1, V_2) are additive, then:

$$V = V_1 + V_2 = k_1\mu_1 Pvt_1 + k_2\mu_2 Pvt_2 \tag{3}$$

where the wear factors (k_1, k_2) are expressed in units of volume lost per unit of frictional work (friction coefficient$=\mu$). For simplicity, consider two cases using the hypothetical values in Table 7. Three cases, all adding up to 3600 s (1 h) total time are shown and the results plotted in Fig. 7. The applied load, velocity, friction coefficients for each stage, and wear factors for each stage remain the same, but the transition point between stage 1 to stage 2 can cause a significant difference in the wear factor if it is calculated by dividing the total volume by the sliding distance and normal force. The difference produced by changing the period of the first stage from 30 s to 1200 s in this example resulted in nearly a 50% difference in the apparent wear factors (shown in Fig. 7 legend). Therefore the need to account for wear transitions can significant affect the repeatability and reproducibility of results.

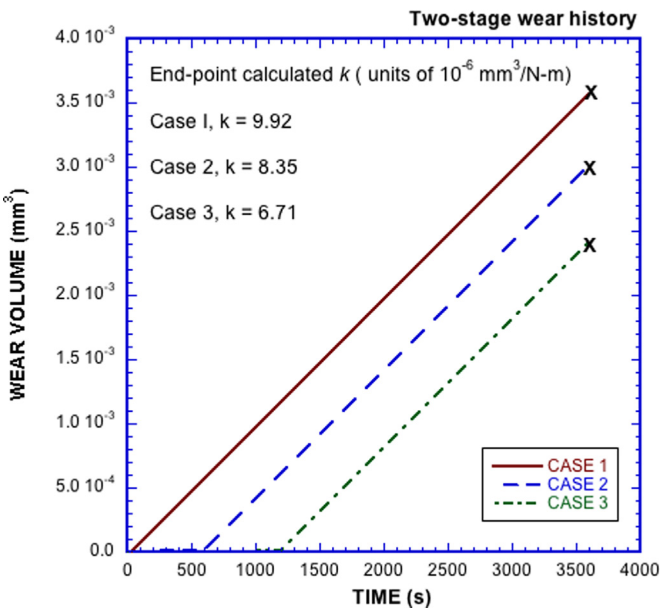

Fig. 7. Three hypothetical cases in which the time between Stage 1 and Stage 2 affects the final wear factor, which is calculated based only on the total wear volume (marked by an "X" on each curve) and the applied load and sliding distance.

5. Conclusions

Like other physical measurements, wear data exhibit a degree of variability. Repeatability and reproducibility (R/R) reflect the variability from test-to-test in a laboratory and from laboratory-to-laboratory respectively. Five case studies were used to illustrate the insights to be gained and the uses for R/R data. One can summarize the lessons learned from such studies are as follows:

1) *Cavitation*. In cavitation erosion testing of three metals using ASTM G32, the wear ranking and the test-to-test variability depended on the choice of wear metric used. Maximum erosion rate, for example, ranked materials in a different order of merit than did incubation period, but the incubation period had the highest percent coefficient of variation (CoV) than did other methods of measurement. The lesson here is that metrics used to measure wear can affect the ranking and repeatability of the results.
2) *Three-body abrasion*. When comparing different procedures for dry sand/rubber wheel abrasion testing of stainless steel using three different procedures in ASTM G65, there was similar CoV despite the fact that the cumulative wear volume differed by a factor of four between the three procedures. One lesson from this is that magnitude of wear produced by different test procedures did not significantly affect the repeatability of the data. Making tests more severe to help discriminate between material behavior can be done without compromising the data repeatability.
3) *Solid particle erosion*. In solid particle erosion tests (ASTM G76) of two different steels, repeatability was worse when the particle velocity used in testing was increased. For a mild steel (AISI 1020), increasing velocity by a factor of about 2.3 times, increased the wear rate by a factor of about 10 times; however, it only worsened the repeatability of the results by a little more than 2 times. Therefore, increasing test severity does not proportionately decrease repeatability.
4) *Dry sliding wear – block-on-ring*. When computing wear volume for four duplicate tests of a bronze block on a rotating steel ring based on (i) weight loss, (ii) wear depth, or (iii) scar width, weight loss consistently underestimated the wear volume compared with other methods. The precision was similar for weight loss and scar depth, but scar width had one anomalously high datum that prevented it from otherwise displaying quite good precision and comparing well with volume computed using scar depths.
5) *Dry sliding wear – pin-on-disc*. Measuring cumulative wear continuously by displacement sensors led to much higher variability than simply measuring the pin wear scar after testing. Repeatability in a large number of international pin-on-disc tests using steel/steel and steel/ceramic couples was +/− 14% but the reproducibility was as high as +/− 38%. Therefore, more repetition is needed for dry pin on disc tests to confidently measure the wear of materials. Furthermore, as pointed out in a comprehensive study of multiple pin-on-disc tests, if enough wear rate measurements are made, the wear rates tend to fit a log-normal distribution in their frequency of occurrence.
6) *Lubricated wear as a measure of fuel lubricity*. The ball-on-cylinder lubricity evaluation (ASTM D5001) test was introduced in the 1960s to measure lubricity of aviation fuels. Concerns of R/R were addressed over years of development and applied to conduct worldwide surveys of fuel lubricity. Effect of test chamber humidity and the preparation of specimens to an optimum roughness helped improve test consistency. Results produced international aviation fuel quality standards.

In addition, there were two additional considerations.

1) *Repeatability versus reproducibility*. For the wear types studied, the reproducibility of data between laboratories was nearly always greater that its repeatability. This has implications for standards development as well as agreeing on product acceptability test methods used by the producer and the user of the product. One can increase repeatability on one's laboratory without necessarily assuming that the results obtained will agree better with those from another laboratory.
2) *Wear transition effects*. A lack of repeatability of the transitions in wear as a function of contact duration can result in significant scatter in the wear rates calculated using the total wear volume measured at the end of the tests.

The R/R of different wear testing methods is dependent on many things: material hommogeneity, methodology, applied variables, metrics used to measure wear, consistency of testing, and external variables that may not be under the direct control of the individuals who perform the tests. Absent other information, conducting at least three tests per set of conditions is suggested. Since the factors that influence the wear of materials are complex, basing general conclusions on only one test result per set of variables is inappropriate. Renewed attention to R/R, precision, and bias of wear data will likely increase the quality of published wear research and wear model validation. It would also promote engineering confidence in wear studies whose primary goal is to compare or to select materials.

Acknowledgement

The author appreciates the information and suggestions provided by M. G. Gee of the National Physical Laboratory, U.K., and during the preparation of this article. The comments of the paper reviewers are also greatly appreciated.

References

[1] J. Belluz, B. Plumer, B. Resnick, The 7 biggest problems facing science, according to 270 scientists ⟨http://www.vox.com/2016/7/14/12016710/science-challeges-research-funding-peer-review-process⟩ posted on July 14 (2016).
[2] G. F. Vander Voort, Results of an ASTM E-4 round-robin on the precision and bias of measurements of microindentation hardness impressions, in: ASTM STP 1025, Factors that Affect the Precision of MechanicalTests, ed. R. Papirno and H.C. Weiss, ASTM International, W. Conshohocken, PA (1989) pp. 3–39.100 ASTM 691.

[3] A. Palmgren, The service life of ball bearings, Z. Ver. Deut. Ingr. 68 (14) (1924) 339–341.
[4] E.V. Zaretsky, How to determine bearing system life, Mach. Lubr. (2012) 30–33 (Nov.– Dec).
[5] E. V. Zaretsky, Rolling Bearing Life Prediction, Theory and Application, NASA Technical Publications NASA/TP-2013-215305, 2013, pp. 57.
[6] P.J. Blau, K.G. Budinski, Development and Use of ASTM Standards for Wear Testing, Wear 225-229 (1999) 1159–1170.
[7] ASTM E-691-15 Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method, ASTM Annual Book of Standards, Vol. 14.05 2015.
[8] ASTM G117-13, Standard guide for calculating and reporting measures of precision using data from interlaboratory wear or erosion tests, ASTM Annual Book of Standards, Vol. 03.02, ASTM International, West Conshohocken, PA USA (2014) pp. 524–529.
[9] ASTM G32-10, Standard test method for cavitation erosion using vibratory apparatus, ASTM Annual Book of Standards, Vol. 03.02, ASTM International, West Conshohocken, PA USA 2014 pp. 99–118.
[10] J.F. Box, Gosset, fisher, and the t distribution, Am. Stat. 35 (1981) 61–66.
[11] P.J. Blau, Needs and challenges in precision wear measurement, ASTM J. Test. Eval. 25 (2) (1997) 216–225.
[12] DIN Standard 50 321, Wear: Quantities for Use in Measuring Amounts of Wear, Deutsches Institut für Normen, Berlin, Germany, 1980.
[13] ASTM G65-10, Standard test method for measuring abrasion using the dry sand/rubber wheel apparatus, ASTM Annual Book of Standards, Vol. 03.02, ASTM International, West Conshohocken, PA USA 2014, pp. 257–268.
[14] ASTM G76-13, Standard test method for conducting erosion tests by solid particle impingement using gas jets, ASTM Annual Book of Standards, Vol. 03.02, ASTM International, West Conshohocken, PA USA, 2014, pp. 327–332.
[15] P. J. Blau, Wear Journal Editorial Records for Submissions Received Over a six-week Period in 2016.
[16] ASTM G99-2010, Standard test method for wear testing with a pin-on-disk apparatus, ASTM Annual Book of Standards, Vol. 03.02, ASTM International, West Conshohocken, PA USA, 2014 pp. 421–425.
[17] M. G. Gee, Results from a U.K. interlaboratory project on dry sliding wear of alumina. in: Wear Testing of Advanced Materials, eds. R. Divakar, P. J. Blau, ASTMSTP 167, ASTM International, W. Conshohocken, PA, 1992, pp. 129–150.
[18] M. G. Gee, Effect of Test Machine Dynamics on the Sliding Wear of Alumina." in Wear Testing of Advanced Materials, ed. R. Divakar, P. J. Blau, ASTMSTP 167, ASTM International, W. Conshohocken, PA, 1992, pp. 24–44.
[19] H. Czichos, S. Becker, J. Lexow, Multi-laboratory testing: results from the versailles advanced materials and standards programme on wear test methods, Wear 114 (1987) 109–130.
[20] N.C.Wallbridge and D.Downson, Distribution of wear rate data and a statistical approach to sliding wear theory, in: Proceedings of the International Conference on Wear of Materials, vol. 1, pp. 101–110.
[21] ASTM D5001-10 Standard test method for measurement of lubricity of aviation turbine fuels by the Ball-on-Cylinder Lubricity Evaluator (BOCLE), ASTM Annual Book of Standards 05.02 2014, ASTM International, West Conshohocken, PA USA.
[22] G. Rickard, P. Brook, Aviation fuel lubricity–a review, CRC Report AV-14-11, Coordinating Research Council, Inc., Alpharetta, Georgia, USA (2014), pp. 80.
[23] T.B. Biddle, R.J. Meehan, P.A. Warner, Standardization of lubricity test, Air Force Wright Aeronautical Laboratories Report. AFWAL-TR- 87-2014, 1987, pp. 40.
[24] P.I. Lacey, Wear mechanism evaluation and measurement in fuel-lubricated components, U.S. Army Contract Interim Report #BFLRF 286, Southwester Research Institute San Antonio, TX , 1994, p. 143.
[25] P.M. Sargent, Use of the indentation size effect on microhardness for materials characterization (ASTM STP 889), in: P.J. Blau, B.R. Lawn (Eds.), Microindentation Techniques in Materials Science and Engineering, ASTM International, West Conshohocken, PA, 1985, pp. 160–174.
[26] ASTM Annual Book of Standards, Vol. 02.02, on Corrosion of Metals; Wear and Erosion, ASTM International, W. Conshohocken, PA, 2014, 1020 pp.
[27] R. W. Bruce, ed., Handbook of Lubrication and Tribology, Vol. II, 2nd ed., Theory and Design, CRC/Taylor & Francis, Boca Raton, FL, 2012.
[28] P.J. Blau (Ed.), ASM Handbook, Volume 18: Friction, Lubrication, and Wear Technology, 18, ASM International, Materials Park, OH, 1992.
[29] G. Stachowiak, A. Batchelor, Experimental Methods in Tribology, 1st Ed, Elsevier Sci, The Netherlands 2004, p. 372.
[30] P.J. Blau, A multi-stage wear model for grid-to-rod fretting of nuclear fuel rods, Wear 313 (1-2) (2014) 89–96.
[31] P.J. Blau, How common is the steady state? The implications of wear transitions in material selection and design, Wear 330-331 (2015) 1120–1128.