Designation: G 117 – 02

Standard Guide for Calculating and Reporting Measures of Precision Using Data from Interlaboratory Wear or Erosion Tests[1]

This standard is issued under the fixed designation G 117; 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 guide offers direction on the handling of data from interlaboratory tests for wear or erosion. It describes a format for entering data and for subsequently reporting results on measures of precision in a Committee G02 standard. It indicates methods for calculation of the needed statistical quantities.

1.2 The document offers guidance based on a Committee G02 consensus, and exists for the purpose of emphasizing the need to use established statistical practices, and to introduce more uniformity in reporting interlaboratory test results in Committee G02 standards.

1.3 An example of how the methods described in this guide may be applied is available in personal computer format (DOS type system) on floppy disk as a spreadsheet (LOTUS, rel. 4) file. The purpose is to facilitate use of the methods in this guide. The example file contains all needed equations in the recommended format and can be edited to accept new data. ASTM Headquarters or the Chairman of G02 should be contacted for a copy of that computer file. The user must have spreadsheet software (for example, LOTUS or compatible) available.

1.4 The methods used in this document are consistent with Practices E 691 and E 177, and with the PC version of Practice E 691.[2]

2. Referenced Documents

2.1 *ASTM Standards:*

E 177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods[3]

E 691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method[3]

G 65 Test Method for Measuring Abrasion Using the Dry Sand/Rubber Wheel Apparatus[4]

G 76 Practice for Conducting Erosion Tests by Solid Particle Impingement Using Gas Jets[4]

G 77 Test Method for Ranking Resistance of Materials to Sliding Wear Using Block-on-Ring Wear Test[4]

3. Summary of Guide

3.1 Use of this guide in preparation of interlaboratory test results for inclusion in G02 standards involves a sequence of steps. First the raw data from the individual laboratories are entered into a table of any suitable form that permits calculation of average values and standard deviations for each laboratory. Then those two measures are entered, for each laboratory, into a table such as that shown in Fig. 1. Then the steps described in this guide are carried out, leading to calculation of the precision measures that are to be used in the standard being prepared.

4. Significance and Use

4.1 This guide is intended to assist in developing statements of precision and supporting data that will be used in Committee G02 standards. The methods and approach are drawn from Practice E 177 and E 691. It was felt that preparation of this guide and its use in Committee G02 would lead to appropriate statistical analyses and more uniformity in G02 standards regarding reporting of interlaboratory results and precision. The guide is not meant to substitute for possible use of Practices E 177 or E 691 in developing committee standards.

5. Procedure

5.1 An example of interlaboratory data analyzed and presented in the recommended format is shown in Fig. 1. The data were obtained from an interlaboratory series of solid particle erosion tests carried out in connection with Practice G 76. This table format can be used with either PC spreadsheet calculation or hand calculation.

5.2 Data tabulation and calculation can be carried out by use of a PC and numeric spreadsheet software (for example, LOTUS), as described in Table 1, or by any other appropriate means such as hand calculation (Table 2). The formulae were obtained from Practices E 177 or E 691 or from statistical analysis texts. Formulae that are used for calculation are given

[1] This guide is under the jurisdiction of ASTM Committee G02 on Wear and Erosion and is the direct responsibility of Subcommittee G02.20 on Computerization in Wear and Erosion.

Current edition approved Nov. 10, 2002. Published February 2003 Originally approved in 1993. Last previous edition approved in 1998 as G 117 – 98.

[2] Available from ASTM Headquarters. Order PCN 12-506910-34.

[3] *Annual Book of ASTM Standards*, Vol 14.02.

[4] *Annual Book of ASTM Standards*, Vol 03.02.

	A	B	C	D	E	F	G	H
1	**ASTM G-2 INTERLABORATORY TEST DATA - STATISTICAL ANALYSIS (G117_93 ver.2)**							
2								
3								
4					WITHIN-LAB REPEATABILITY		BETWEEN-LAB REPRODUCIBILITY	
5	TEST	LAB	NUMBER OF	AVERAGE	STD DEV	k STATISTIC	DEV FROM AVG	h STATISTIC
6	CONDITIONS	#	REPLICATES	(units)	(units)		(units)	
7		----	----------	------	------		------	
8	List key	1	3	9.800	0.500	1.100	1.100	0.434
9	information,..	2	3	10.500	0.100	0.220	1.800	0.710
10	.....	3	3	5.800	0.600	1.320	-2.900	1.144
11								
12		------	------	------	------		------	
13		3	3	8.700	0.455		2.563	
14		NUMBER	AVERAGE	AVERAGE	WITHIN-LAB		BETWEEN-LAB	
15					STD DEV		STD DEV (PROV)	
16								
17				C.O.V. (%) =	5.2		29.5	
18								
19				95 % LIMITS=	1.27		7.18	
20	** USE THE LARGER OF THE **				WITHIN-LAB		BETWEEN-LAB	
21	** 95% LIMITS FOR THE FINAL VALUE **							
22					k crit =	1.67	h crit =	1.15
23								
24					k and h values greater than k crit and h crit			
25	18-Nov-97				suggest those data should be examined for 'outliers'.			
26								
27								
28	*Recommended statement of precision*:				The average test value was 8.70(units)			
29					with a 95% repeatability limit (within-lab) of 1.27(units)			
30					and a 95% reproducibility limit (between-labs) of 7.18(units) .			
31								
32								

NOTE 1—Column and row labels A, B, . . . and 1, 2, . . . are not required.

FIG. 1 Example of Recommended Format for Data Analysis

TABLE 1 Formulae Used in PC Spreadsheet Shown in Fig. 1, in Notation Appropriate to Spreadsheet Software (for example, LOTUS)[A]

B13:	@COUNT(B8..B11)
C13:	@AVG(C8..C11)
D13:	@AVG(D8..D11)
E13:	@SQRT((@SUM(K8 . . K11))/B13)
G13:	@SQRT((@SUM(L8..L11))/(B13-1) + E13*E13*(C13-1)/C13)
	where:
	F8: +E8/ E13 — H8: @ABS(+G8/ L13)
	K8: +E8*E8 — L8: +G8*G8
	and so forth — and so forth
L13:	@SQRT((@SUM(L8..L11))/(B13-1)
E17:	100*E13/D13
G17:	100*G13/ D13
E19:	2.8*E13
G19:	2.8*G13

[A]Note—*N* is used as the divisor in (E12) to obtain the mean value of the variance, while *N*-1 is used as the divisor in calculating individual standard deviations (E7..E9) since they are estimates of population values. Practice E 691 should be consulted for further explanation.

TABLE 2 Formulae Used in Calculating Quantities for Fig. 1, Given in Usual Mathematical Notation

B13:	$N = \Sigma n$	Number of laboratories
C13:	$R = (1/N)\cdot\Sigma r$	Average number of replicates
D13:	$Q = (1/N)\cdot\Sigma q$	Average of the quantity measured
E13:	$W = [(1/N)\cdot\Sigma s^2]^{0.5}$	Within-laboratory standard deviation
G13:	$B = [(1/(N-1))\cdot\Sigma(q-Q)^2 + (1/N)\cdot\Sigma s^2\cdot(R-1)/R]^{0.5}$	Provisional between-laboratory standard deviation
F8:	s/W	*h*-statistic
H8:	d/s_x	*k*-statistic
K8:	s^2	cell standard deviation
L8:	d^2	cell deviation squared
L13:	$[(1/(N-1)\cdot\Sigma(q-Q)^2]^{0.5}$	standard deviation of cell averages
E17:	100·W/Q	Percent coefficient of variation, within-laboratory
G17:	100·B/Q	Percent coefficient of variation, between-laboratory
E19:	2.8·W	95 % confidence limits, within-laboratory
G19:	2.8·B	95 % confidence limits, between-laboratory

in Table 1 for spreadsheet calculation (for example, LOTUS) and in Table 2 for hand calculation.

5.3 The sequence of steps in assembling and handling the data is as follows (refer to the designated columns in Fig. 1):

5.3.1 Calculate the *average* value of the data for each of *N* laboratories. (Column D)

5.3.2 Calculate the *average* value *Q* of all the laboratory averages. (Cell D13)

5.3.3 Calculate the *standard deviation* values for each laboratory. Note that the quantity $(r - 1)$ is used as the divisor where *r* is the number of replicate results for each laboratory. (Column E)

5.3.4 Calculate the *within-laboratory standard deviation* value *W*. Note that this is the root-mean-square value of the laboratory standard deviations, using *N* as the divisor. This quantity is also called the repeatability standard deviation. (Cell E13)

5.3.5 Calculate the *within-laboratory coefficient of variation* in percent. (Cell E17)

5.3.6 Calculate the *k*-statistic values for each laboratory, by dividing each laboratory standard deviation by the within-laboratory standard deviation (Column F).

5.3.7 Calculate the *deviation* of the average for each laboratory from the average for all laboratories. (Column G)

5.3.8 Calculate the *between-laboratory standard deviation* value *B*. Note that this is the square root of the sum of the mean-square value of the deviations from the average, using $N - 1$ as the divisor, and the square of the within-laboratory

standard deviation multiplied by the quantity $(r - 1)/r$. This is also called the provisional reproducibility standard deviation. (Cell G13)

NOTE 1—It is termed provisional since the final reproducibility standard deviation will be the larger of the two calculated measures, the repeatability and the reproducibility standard deviations.

5.3.9 Calculate the *between-laboratory coefficient of variation* in percent. (Cell G17)

5.3.10 Calculate the *h*-statistic values for each laboratory, by dividing each laboratory deviation from average by the between-laboratory standard deviation (Column H).

5.3.11 Select the larger of the two quantities calculated in 5.3.4 and 5.3.8 for the (final) reproducibility standard deviation. An example is shown at the bottom of Fig. 1

5.3.12 Calculate the *95 % limits of repeatability and reproducibility* by multiplying the within-laboratory standard deviation and the (final) between-laboratory standard deviation, respectively, by the factor, 2.8×. (Cells E19 and G19)

NOTE 2—These limits are the maximum differences between two test results that can be expected to occur in 95 % of the cases.

5.3.13 Refer to Practice E 691, Table 12, and determine critical values of *k* and *h* for the number of laboratories and replicates involved. Examine the values in the *k*-statistic and *h*-statistic columns. Any values greater than the respective critical values indicate data outliers for that laboratory which should be inspected for validity. (cells F22 and H22)

6. Report

6.1 Examples of the recommended tabular format for the results of the calculations are shown in Fig. 2for three standards from Committee G02.

6.2 A recommended version of a statement of precision, drawn from Practice E 177, is as follows for the example shown in Fig. 1:

Average Test Value:	8.70 mm^3/g
95 % repeatability limit (within-lab)	1.27 mm^3/g
95 % reproducibility limit (between-labs)	7.18 mm^3/g

7. Keywords

7.1 erosion; precision; repeatability; reproducibility; wear

TEST CONDITIONS	LAB #	NUMBER OF REPLICATES	AVERAGE (mm3/g)	WITHIN-LAB REPEATABILITY STD DEV (mm3/g)	k STATISTIC	BETWEEN-LAB REPRODUCIBILITY DEV FROM AVG (mm3/g)	h STATISTIC
G-76; erosion;	1	5	31.500	1.100	1.135	3.340	0.711
1020 steel;	2	5	23.200	0.040	0.041	-4.960	1.055
70 m/s	3	5	22.900	0.900	0.929	-5.260	1.119
	4	5	32.400	0.650	0.671	4.240	0.902
	5	5	30.800	1.500	1.548	2.640	0.562
	5	5	28.160	0.969		4.780	
	NUMBER	AVERAGE	AVERAGE	WITHIN-LAB STD DEV		BETWEEN-LAB STD DEV (PROV)	
			C.O.V. (%) =	3.4		17.0	
			95 % LIMITS=	2.71		13.38	
				WITHIN-LAB		BETWEEN-LAB	
** USE THE LARGER OF THE ** ** 95% LIMITS FOR THE FINAL VALUE **				k crit =	1.71	h crit =	1.74

19-Nov-97

k and h values greater than k crit and h crit suggest those data should be examined for 'outliers'.

Recommended statement of precision: The average test value was 28.16(mm3/g) with a 95% repeatability limit (within-lab) of 2.71(mm3/g) and a 95% reproducibility limit (between-labs) of 13.38(mm3/g) .

TEST CONDITIONS	LAB #	NUMBER OF REPLICATES	AVERAGE (mm3)	WITHIN-LAB REPEATABILITY STD DEV (mm3)	k STATISTIC	BETWEEN-LAB REPRODUCIBILITY DEV FROM AVG (mm3)	h STATISTIC
G-65;dry sand;	1	6	34.830	1.530	1.083	-0.893	0.454
rubber wheel	2	3	32.900	1.040	0.736	-2.823	1.436
abrasion; D2	3	3	35.170	0.230	0.163	-0.553	0.281
steel; RR#7	4	4	35.950	2.170	1.536	0.227	0.115
6/26/80	5	6	38.750	1.660	1.175	3.027	1.540
	6	5	36.740	1.020	0.722	1.017	0.517
	6	5	35.723	1.413		2.327	
	NUMBER	AVERAGE	AVERAGE	WITHIN-LAB STD DEV		BETWEEN-LAB STD DEV (PROV)	
			C.O.V. (%) =	4.0		6.5	
			95 % LIMITS=	3.96		6.52	
				WITHIN-LAB		BETWEEN-LAB	
** USE THE LARGER OF THE ** ** 95% LIMITS FOR THE FINAL VALUE **				k crit =	1.75	h crit =	1.92

19-Nov-97

k and h values greater than k crit and h crit suggest those data should be examined for 'outliers'.

Recommended statement of precision: The average test value was 35.72(mm3) with a 95% repeatability limit (within-lab) of 3.96(mm3) and a 95% reproducibility limit (between-labs) of 6.52(mm3) .

TEST CONDITIONS	LAB #	NUMBER OF REPLICATES	AVERAGE (mm3)	WITHIN-LAB REPEATABILITY STD DEV (mm3)	k STATISTIC	BETWEEN-LAB REPRODUCIBILITY DEV FROM AVG (mm3)	h STATISTIC
G-77; block-	1	3	0.860	0.038	0.143	0.153	0.812
on-ring; H-60	2	3	0.515	0.196	0.738	-0.192	1.022
steel vs S-10	3	3	0.877	0.403	1.517	0.170	0.903
steel; RR#3	4	3	0.577	0.283	1.065	-0.130	0.693
	4	3	0.707	0.266		0.287	
	NUMBER	AVERAGE	AVERAGE	WITHIN-LAB STD DEV		BETWEEN-LAB STD DEV (PROV)	
			C.O.V. (%) =	37.6		40.6	
			95 % LIMITS=	0.74		0.80	
				WITHIN-LAB		BETWEEN-LAB	
** USE THE LARGER OF THE ** ** 95% LIMITS FOR THE FINAL VALUE **				k crit =	1.82	h crit =	1.49

19-Nov-97

k and h values greater than k crit and h crit suggest those data should be examined for 'outliers'.

Recommended statement of precision: The average test value was 0.71(mm3) with a 95% repeatability limit (within-lab) of 0.74(mm3) and a 95% reproducibility limit (between-labs) of 0.80(mm3) .

FIG. 2 Examples Using Data From Three Committee G02 Standards

APPENDIX

X1. GUIDELINES ASSOCIATED WITH PRACTICE E 691

X1.1 Introduction

X1.1.1 This Appendix will summarize certain guidelines found in Practice E 691. The purpose of this summary is to emphasize several key guidelines in any interlaboratory study (ILS) of wear and erosion. The reader is directed to Practice E 691 as the definitive document for more details and additional considerations.

X1.2 General Considerations

X1.2.1 Tests performed on presumably identical materials in presumably identical circumstances do not, in general, yield identical results. This is attributed to unavoidable random errors inherent in every test procedure; the factors that may influence the outcome of a test cannot all be completely controlled. The general term for expressing the closeness of test results to the "true" value or the accepted reference is *accuracy*. To be of practical value, standard procedures are required for determining the accuracy of a test method, both in terms of its bias and in terms of its precision. *Precision*, as discussed in Practice E 691, is expressed in terms of two measurement concepts: repeatability and reproducibility. Under repeatability conditions, the controlling factors are kept or remain reasonably constant and usually contribute only minimally to the variability. Under reproducibility conditions, the factors are generally different (that is, they change from laboratory to laboratory) and usually contribute appreciably to the variability of test results. To obtain reasonably estimates of repeatability and reproducibility precision, it is necessary in an interlaboratory study to guard against excessively sanitized data in the sense that only the uniquely best operators are involved or that a laboratory takes unusual steps to get "good" results. It is also important to recognize and consider how to treat "poor" results that may have unacceptable causes, for example, departures from the prescribed procedure.

X1.3 Number of Laboratories

X1.3.1 It is important that enough laboratories be included in the ILS to be a reasonable cross-section of the population of qualified laboratories, that the loss or poor performance of a few laboratories will not be fatal to the study, and that the ILS provides a reasonably satisfactory estimate of the reproducibility. According to Practice E 691, under no circumstances should the final statement of precision of a test method be based on acceptable test results for each material from fewer than 6 laboratories.

X1.3.2 This being said, it is often the case that test methods developed by G02 members are in use in only a few laboratories. In such cases, *provisional* interlaboratory testing may go forward involving as few as 3 laboratories, but no fewer. The responsible subcommittee must plan to conduct another ILS later that includes at least 6 laboratories, and then to use those results to replace the provisional data from the first ILS.

X1.4 Number of Materials

X1.4.1 An ILS of a test method should include at least three materials representing different test levels, and for development of broadly applicable precision statements, six or more materials should be included in the study, according to Practice E 691. The materials involved in any one ILS should differ primarily only in the level of the property measured by the test method. When it is known, or suspected, that different classes of materials will exhibit different levels of precision when tested by the test method, consideration should be given to conducting separate interlaboratory studies for each class of material. Each material in an ILS should be made to be or selected to be as homogeneous as possible prior to its subdivision into test units or test specimens.

X1.5 Number of Replicate Measurements

X1.5.1 It is generally sound to limit the number of test results on each material in each laboratory to a small number, such as three or four. The minimum number of test results per laboratory will normally be three or four for a physical test. This should apply to wear or erosion tests. As many as ten replicates may be needed when test results are apt to vary considerably. Generally, the time and effort invested in an ILS is better spent on examining more materials across more laboratories than on recording a large number of test results per material within a few laboratories.

X1.6 Consideration of Outliers

X1.6.1 If an investigation of the ILS data discloses no clerical, sampling, or procedural errors, any unusual data should be retained, and the precision statistics based on them should be published. If, on the other hand, a cause for unusual data was found during the investigation, the task group has several options to consider. If the laboratory clearly and seriously deviated from the test method, the test results for that laboratory must be removed from the ILS calculations. However, despite the danger of a questioned laboratory having prior knowledge, it may be appropriate to ask that laboratory to retest one or more materials following the correct procedure, and then include the new set of results as replacements in the ILS calculations. When a large number of laboratories have participated in the ILS and no cause for some unusual values have been found during the investigation, it may be appropriate to delete a laboratory from the study if all of the other laboratories are in substantial agreement. The number of laboratories that can be considered large enough to support deletion of data without an identified cause cannot be stated exactly. According to Practice E 691, any action which results

in discarding more than 5 % of the ILS data should not be taken, as it likely will lead to values of precision (primarily reproducibility) that the test method cannot deliver in routine application.

X1.6.2 This being said, it is often the case that test methods developed by G02 members are in use in only a few laboratories. In such cases, *provisional* interlaboratory testing results may result after a review that entails discarding more than 5 % of the data. The responsible subcommittee must plan in such a case to conduct another ILS later that includes more laboratories, and then to use those results to replace the provisional data from the first ILS. The final ILS data for the standard should reflect the criteria stated in Practice E 691.