Journal of ASTM International, April 2006, Vol. 3, No. 4
Paper ID JAI12869
Available online at www.astm.org

C. McCowan,[1] *G. Roebben,*[2] *Y. Yamaguchi,*[3] *S. Lefrançois,*[4] *J. Splett,*[5] *S. Takagi,*[6] *and A. Lamberty*[2]

International Comparison of Impact Reference Materials (2004)[1]

ABSTRACT: A three-year horizontal comparison has been completed between national laboratories that certify specimens for the indirect verification of Charpy impact test machines. The participants in this study were the Institute for Reference Materials and Measurements of the European Commission, the National Metrology Institute of Japan (NMIJ)[2], the National Institute of Standards and Technology in the United States, and the Laboratoire National d'Essais in France. The comparison was conducted to evaluate the impact of reference specimens over a three-year period. Sets of certified reference specimens, at low (15 J), medium (30 J), and high energy (100 J) levels were produced and distributed at the start of the study. Specimens were tested approximately every six months on each of the machines in the study. The results of the testing are presented and the stability of the various impact machines and specimens are discussed.

KEYWORDS: ASTM E-23, Charpy V-notch, EN 10045 impact testing, ISO 148 machine verification

Introduction

Charpy impact testing is often specified as an acceptance test for structural materials, and companies performing acceptance tests are typically required to verify the performance of their impact machine with certified reference materials. The laboratories in this comparison each have impact machines that play a role in the certification of reference materials for the verification of Charpy impact machines: (1) The European Commission Joint Research Centre, Institute for Reference Materials and Measurements (IRMM, Belgium), (2) Laboratoire National D'Essais (LNE, France), (3) The National Institute of Standards and Technology (NIST, USA), and (4) The National Metrology Institute of Japan (NMIJ, Japan). Annually, these four laboratories supply specimens to verify the performance of about 2000 impact machines around the world.

The purpose of this interlaboratory study is to determine the long-term stability of impact verification specimens and reference machines. We also examine the nominal differences (bias) among machines. Given the destructive nature of impact testing, and the lack of knowledge regarding the true breaking strength of the specimens, it is difficult to evaluate the absolute performance of Charpy impact machines. However, the relative performance of our machines can be examined. These types of horizontal comparisons help to define important similarities and differences between our impact machines and specimens, and the results allow us to target calibrations and changes to our respective programs that make them more transparent to the users.

Manuscript received December 21, 2004; accepted for publication October 21, 2005. Presented at ASTM Symposium on Pendulum Impact Machines: Procedures and Specimens on 8 November 2004 in Washington, DC ; T. A. Siewert, M. P. Manahan, , Guest Editors.C. N. McCowan, and D. Vigliotti, Guest Editors.
[1] Materials Research Engineer, NIST, Materials Reliability Division, 325 Broadway, Boulder, CO, 80305.
[2] IRMM, Reference Materials Unit, EC-JRC-IRMM, Retieseweg 111, B-2440-Geel, Belgium.
[3] NMIJ, Mass and Force Standard Section, AIST Tsukuba Central 3, 1-1-1, Umezono, Tsukuba, Ibarki, 305-8563, Japan.
[4] LNE, Mechanical and Equipment Testing, LNE, 5 Avenue Enrico Fermi, 78197, Trappes Cedex, France.
[5] Mathematical Statistician, Statistical Engineering Division, NIST, 325 Broadway, Boulder, CO, 80305.
[6] NMIJ, Vibration and Hardness Section, AIST Tsukuba Central 3, 1-1-1, Umezono, Tsukuba, Ibaraki, 305-8563, Japan.
[1]Contribution of NIST; not subject to copyright.
[2]The National Research Laboratory of Metrology (NRLM) was reorganized as NMIJ.

TABLE 1—*Details of machines use for impact testing.*

Certification program	ID	Machine capacity (J)	Pendulum design	Dial/Encoder	Striker radius (mm)
1	1	300	C	Dial	2
	2	300	C	Dial (until Oct 2003, then encoder)	2
2	3	350	U	Encoder	2
3	4	324	U	Encoder	8
	5	358	U	Encoder	8
	6	360	C	Encoder	8
4	7	500	C	Dial	2
	8	500	C	Dial	2

Materials and Procedures

Specimens

Verification specimens at three energy levels were used for testing. At each energy level samples came from a single batch. The specimens used for all of the verification specimens (low, medium, and high energy levels) were made using a heat treated AISI 4340 steel. The low energy specimens were heat treated to have a nominal energy of 15 J, when tested at −40°C. The medium energy specimens were heat treated to have a nominal energy of 30 J at 20°C. The high energy specimens were heat treated to have a certified energy near 100 J at −40°C.

Testing Details

The study was designed to test specimens twice a year over a three year period. On each test date, ten specimens at each of three energy levels were tested on each machine. Tests were performed at the temperatures for which the respective batches were produced (−40°C for low and high energy levels, 20°C for the medium energy level).

Machine Details

Eight pendulum impact machines were evaluated in this study, and some details concerning these machines are listed in Table 1. The grouping by *certification program* is of practical concern for this study, because there is interest in comparing the "verification systems" used in Europe, Japan, and the United States. The details for the machines and the average energy values determined for them in this study do not fully describe each of the certification systems [1]. In particular, Program 1 certifies Master Batches of impact specimens by use of an international intercomparison (with ten or more machines). To verify the performance of industrial pendulum impact machines, samples of so-called Secondary Batches are used. The Secondary specimens are compared with the Master specimens of the same nominal energy. These certification tests are done in repeatability conditions, on a single machine [2]. Until recently, this was machine 1, today this is done with machine 2. Therefore, a direct comparison of an impact machine from Program 1, with machines from Program 2, 3, or 4 is not a direct comparison of certification systems. Other details and interrelationships between machines and programs make direct comparisons difficult as well.

Fluctuations in the respective programs, due to machine repairs, part replacements, and other factors are expected to be apparent over the three-year period of this study. For example, the replacement of anvils might influence the energy value determined by a machine. So, correlations of machine performance with service records are considered.

Striker radii (2 and 8 mm) differ for the machines in this study. Although this is a real and identifiable variable for the machines, it is considered here as just another nonseparable machine variable or bias (2 and 8 mm results are directly compared). This approach is taken because the average differences in absorbed energy due to testing with 2 and 8 mm striker radii on these machines with AISI 4340 verification specimens is small considering the known magnitudes of machine bias [1]. The choice of striker was left up to the laboratory. The machines in program 1 and 2 always used a 2 mm striker radius, and the machines in program 3 always used an 8 mm striker radius. The machines in program 4 used the striker radii associated with the certified value of the specimens tested: 2 mm striker radii for medium energy level, and 8 mm striker radii for low and high energy specimens.

The maximum capacity of each machine is listed in Table 1. The capacities of machines 1 through 6 are similar, between 300 and 360 J. Machines 7 and 8 have the highest capacities used in the study, 500 J.

Results and Discussion

Specimen Stability

Before comparing relative machine performance, it is necessary to determine the stability of specimens over time. The seven tests performed in this study, over a three-year period (about every six months), allow a systematic investigation of the stability in time of the absorbed energy values of the batches. Since the data were collected at unequally spaced intervals, evaluations were based on actual measurement dates to obtain valid statistical tests as well as an accurate representation of the data over time. As shown in Fig. 1, a regression of the average energy for each machine and test date is made for each energy level, ignoring differences between machines. None of the regression slopes were significant at the 0.05 level.

A regression analysis of average energy versus test date was performed for each machine and energy level individually. Three slopes were found to be significant: The probability that the calculated slope would have occurred by chance if the "true" slope is zero for machine 3 at high energy was 0.002; For machine 6 at high energy the p value was 0.01; For machine 6 at low energy the p value was 0.03. The majority of machines do not display significant trends, and the trends noted are of magnitudes within the range of the overall (random) variation for several other machines. On average, the standard deviation of all the results on a single pendulum at a particular energy varies from 4 to 6 % (low energy), 5 to 7 % (medium energy), and 3 to 4 % (high energy)

We also analyzed the variance for each energy level based on machine and test date. The effect of the test date was not significant (at the 0.05 level) for any of the three energy levels even after accounting for differences between machines.

Conclusive evidence of specimen and machine stability is difficult to obtain. Because the Charpy test is destructive, it is difficult to separate drift in machines from drift in specimens with data obtained in this study. However, based on the results of the regression analyses using the combined data, it appears that the low, medium, and high energy level impact verification specimens were stable during the three-year period of the interlaboratory comparison. Assuming that specimens are stable, then there are two machines that may be drifting, machine numbers 3 and 6.

Further analyses were performed to determine whether the sample variance was stable across measurement occasions. Bartlett's test for equality of variance among measurement occasions was performed for each machine and energy level. Only one machine was found to have inhomogeneous variance across measurement occasions, machine 3 at high energy. This result indicates predictable behavior of machines over time with respect to variability, and is another indication of stability.

Estimates of Mean Energy

Table 2 displays means and standard deviations for each machine, test number, and energy level, as well as grand means and standard deviations based on the combined data. Considering the averages for each time point, as shown in Table 2, the differences between the grand mean and means of individual test numbers are small. The differences for low, medium, and high energy levels are within ±0.1 J, ±0.5 J, and ±1.8 J, respectively (less than 2 %). This variation in the estimates of mean energies for the specimens is small, but significant in context of the uncertainty that might be associated with certified values for verification specimens.

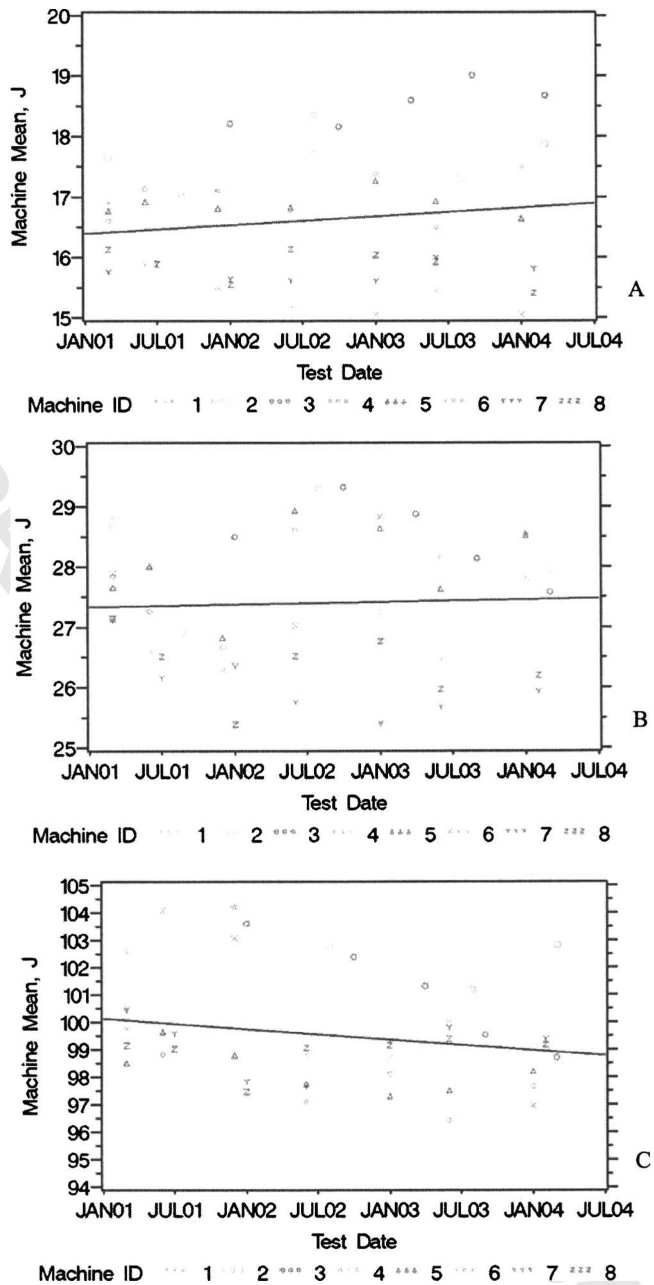

FIG. 1—*The (A) low* p=0.4*, (B) medium* p=0.8*, and (C) high energy* p=0.2 *plots and probabilities showing trends for specimen stability.*

Comparisons of Machines

If all machine means and variances were assumed to be stable we could combine measurement occasions and compare grand averages for each machines. However, machines may change over time, so we examined machines for each energy level and test number separately.

A one-way analysis of variance to determine the equality of machine averages was performed for each energy level and test number separately [2]. For all but two cases (test 2 at medium energy and test 1 at high energy) at least one machine was found to be statistically different from the other means at the 0.05 level of significance. Figure 2 shows energy means and associated variance bars for each machine and test number. The bars were computed as twice the standard deviation of the mean. Bars that overlap for any two machines indicate that the means for the machines are probably not significantly different. The figures represent a graphical confirmation of the analysis of variance results.

There is some evidence of a systematic offset between machines at low energy. Means for machines 6–8 are always lower than means observed for the other machines, while means for machines 2 and 3 are

TABLE 2—*Mean and standard deviation for individual tests and for combined (grand) values.*

Level	Test number	Mean energy (J)	Standard deviation (J)
H	1	100.3	1.3
H	2	99.3	3.0
H	3	100.8	3.1
H	4	99.4	2.3
H	5	99.0	1.4
H	6	99.1	1.6
H	7	99.0	1.9
H	Combined	99.5	2.1
M	1	27.9	0.7
M	2	26.9	0.7
M	3	26.7	1.0
M	4	27.9	1.5
M	5	27.7	1.3
M	6	27.0	1.1
M	7	27.5	1.0
M	Combined	27.5	1.1
L	1	16.5	0.7
L	2	16.5	0.6
L	3	16.5	1.1
L	4	16.7	1.2
L	5	16.7	1.3
L	6	16.7	1.2
L	7	16.7	1.4
L	Combined	16.6	1.0

typically higher than those observed for all other machines. Although there is some systematic difference among machines for medium energy (means for machines 6–8 are often lower than other machine means), there is no evidence of such an effect at high energy.

Another way to view the data is to plot energy means for each test number versus machine (Fig. 3). For low energy, the data indicate that means for each test number within a machine are fairly reproducible; however the separation of means among machines is quite large. For medium energy, the separation of means among machines is not quite as pronounced as for low energy. For high energy, means are fairly consistent among machines; however, the means for each test number within a machine are generally less repeatable.

Interlaboratory Comparisons

The evaluation of interlaboratory comparison data has been considered at length by international measurement laboratories, and working groups have been tasked with providing guidelines for these types of analyses. We can apply interlaboratory principles to the current data by assuming machines are laboratories. For example, Cox proposed an interlaboratory comparison procedure (Procedure A) for which nearly all the assumptions are satisfied [3]. The one assumption that may be violated specifies that measurements from all machines are independent (there may be some correlation among machines within a single laboratory). With this caveat in mind, we apply Procedure A to single test numbers and energy levels to provide a better reference for comparisons. The procedure used is as follows:

1. Determine the mean of all machines.
2. Determine the standard deviation of the mean.
3. Apply a chi-squared test to evaluate the overall consistency of the results.
4. If the consistency check does *not fail*, then we accept the mean as the reference value and calculate degrees of equivalence (or machine biases in our case).
5. If the consistency check *fails*, then an investigation would be implemented to resolve the inconsistencies.

Cox does not recommend computing a reference value unless the laboratories, or machines in our case, are consistent. For our machine comparison, only data from the test number one at high energy passed the

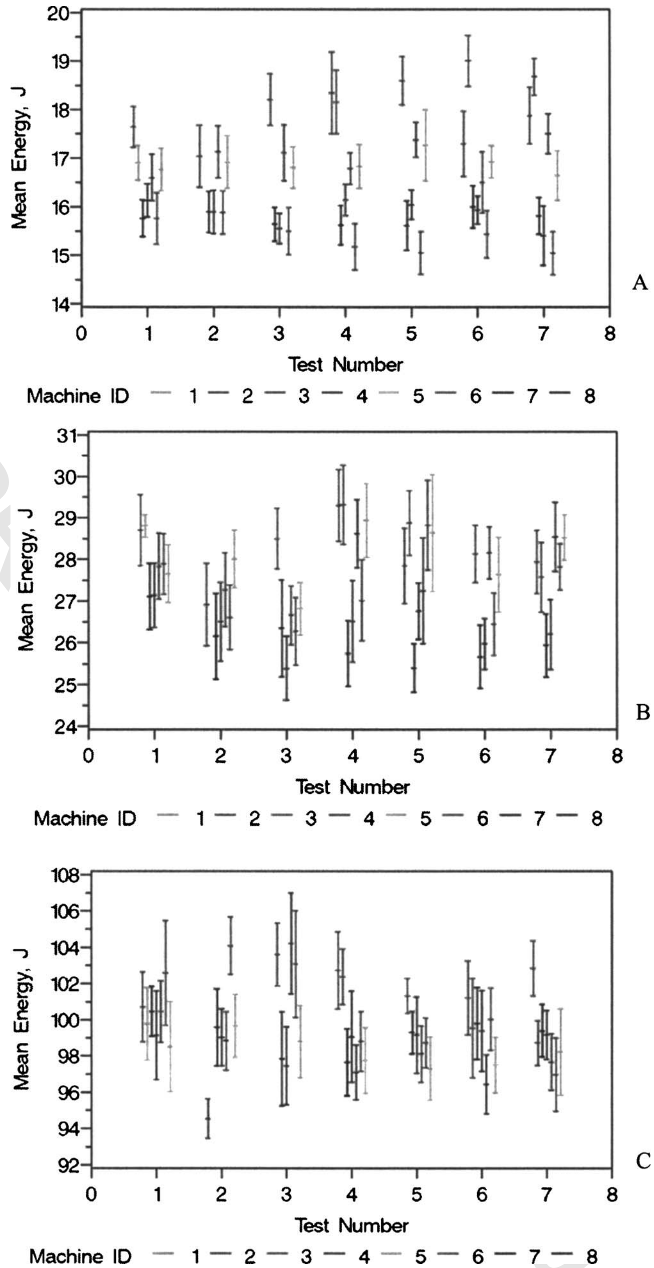

FIG. 2—*The average energy for each machine versus test number for: (A) low energy, (B) medium energy, and (C) high energy. Error bars are twice the standard deviation of the mean.*

consistency check, confirming the analysis of variance results discussed earlier. In general, the results of applying Procedure A indicate that more work is needed to develop and maintain a measurement of impact energy that is internationally consistent.

Contributions to Machine Bias

There are some recognized factors that might be expected to contribute to the bias between machines that are all in full compliance with direct verification requirements. These factors include striker radius, machine capacity, and pendulum design.

There have been numerous studies showing effect of 2 mm versus 8 mm striker radius designs on the measured energy of an impact test. [4–8] Clearly the effect of striker geometry is material dependent, and here only the effects relative to specimens made from type 4340 steel need to be considered. In this study the effect of striker geometry cannot be separated from other machine variables, but we can use data from a previous comparison for this purpose, and these data, shown in Fig. 4, include results for four of the machines used in this current comparison [1].

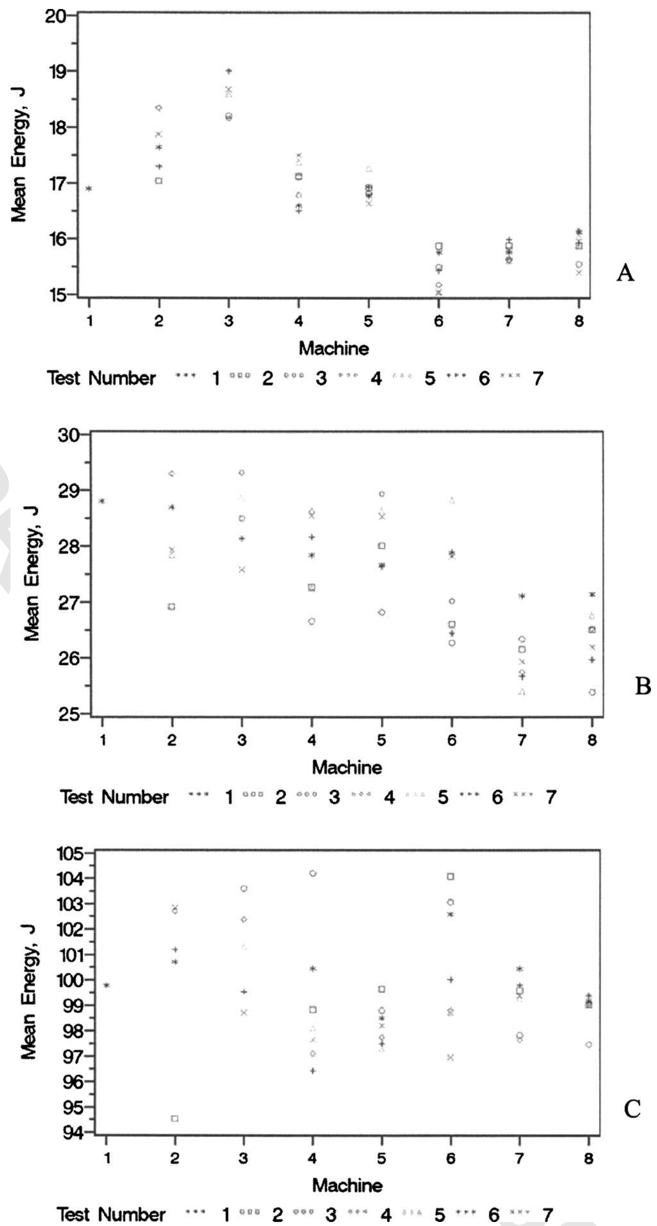

FIG. 3—*The average energy of (a) low, (b) medium, and (c) high energy specimens is shown for each machine and test. Here the grouping of data for a given machine is most apparent. At low energy (a) the relative difference between machines and the lack of overlap for the data indicated clear differences due to the machines.*

The data were generated by testing 15 specimens with a 2 mm striker and ten specimens with an 8 mm striker on each machine for nine sets of specimens (between 15 and 120 J). The average percent differences between the 8 and 2 mm radius strikers for these data at the nominal energy levels of 16, 25, 70, and 100 J are, respectively, about −3 % (−0.44 J), −0.2 % (−0.06 J), −1 % (−0.7 J), and −1 % (−0.9 J). Details in Fig. 4(*b*) show the average values for specimens of very low energy are strongly influenced by individual machines (or tests), so the average difference of −3 % at 16 J may be somewhat misleading. This is supported by the three sets of data near 25 J, for which no significant effect is shown. At 60 J and above, however, there is a trend that is reflected well by the average values. The machines tend to get higher energy results from a striker with a 2 mm radius, compared with results for 8 mm strikers. The magnitude of this effect (average of −1 %) is reasonably convincing because it is independent of the impact machine used for the test and on the origin of the specimens tested. Overall, a magnitude of 1 to 2 % seems like a reasonable approximation for the magnitude of the contribution of striker geometry to the machine bias in this study, and at low energies the effect may be much smaller. This last point is in

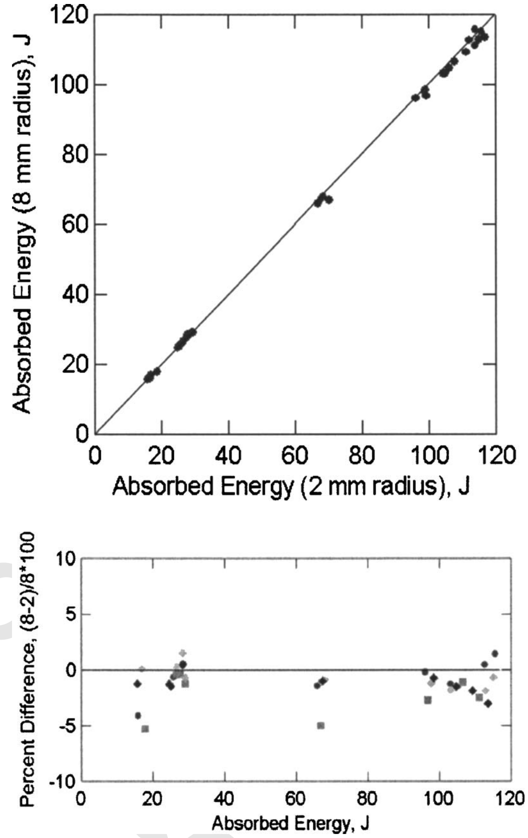

FIG. 4—*A comparison of data for 8 and 2 mm striker geometries gathered by use of reference impact pendulums and type 4340 verification specimens. The upper plot (a) shows a general comparison, and the lower plot (b) shows the percent difference of the 8 and 2 mm striker data gathered on each machine (colors), for each set of verification specimens tested (the 2 mm data was subtracted from the 8 mm data, divided by the 8 mm data, and multiplied by 100).*

agreement with the speculation that strikers with a smaller radius penetrate deeper into softer materials (hardness of 4340 steel specimens decreases as energy increases), which absorbs energy and results in higher measured impact energy [5].

The small effect of striker geometry does not help to explain the differences in machines observed at low energy, where machine differences were largest and best defined. However, it can be argued that at the high energy level, the effect of striker geometry should increase the energy measured by the machines with the 2 mm strikers by about 1 % compared with the machines using 8 mm strikers (machines 4, 5, and 6). There is not a clear trend for this argument, since consistent differences between machines at the high energy level due to striker radii are not apparent for the data taken in this comparison. For example, machines 6, 7, and 8 [Fig. 3(*c*)] have very similar designs, and the results for machine 6 (8 mm striker) are often higher than the results for machines 7 and 8 (2 mm strikers). Also, the differences between machines are too large in many cases to be attributed solely to striker geometry alone, and are likely confounded by other variables.

Machine capacity is not expected to be a significant variable here. The range in capacities for the machines is approximately 300 to 500 J and this range is too small to investigate the influence of capacity. Both of the 500 J machines (machines 7 and 8) tend to perform in a very similar manner, but this is due to overall machine design rather than capacity. We base this on the fact that machine 6 has very similar design and performance to machines 7 and 8, but it has a different capacity (360 J).

The pendulum design, C-type or U-type, alone cannot be identified as contributing to lower or higher energy values. Machines 6–8, which are C-type designs, often performed conservatively compared with the other machines. But machines 1 and 2, which are also C-type designs, often produced energy values higher than the grand mean values.

Machine Maintenance Effects

During the three-year period the participating machines underwent regular (typically annual) direct verifications. At such occasions, deviations from desired machine parameters can lead to replacement or adjustment of particular parts of the pendulum. Such actions can also be purely preventive. An overview of the major maintenance actions (mainly replacement of anvils, supports, or striker) failed to reveal a correlation with the measured values.

Summary and Closing Remarks

Currently, the verification programs associated with IRMM, LNE, NIST, and NMIJ all have machines that are performing within expected and reasonable bounds, and each program can consistently assign certified energies that are stable and suitable relative to their respective user groups. However, the bias between machines makes it difficult for the laboratories to independently provide measures of impact energy for the international community. Providing an internationally defined target for impact energy will require further cooperation between the laboratories and the implementation of a robust and traceable certification process.

Conclusions

The average energies measured for the three levels of impact verification specimens were stable over the three year duration of the study. This indicates that reasonable shelf life can be expected for properly heat treated type 4340 steel impact verification specimens.

Overall, the machines, and groupings of machines by program, appear to be stable over the three-year test period.

The grand average over all machines at each of the three energy levels seems sufficiently stable for the production and maintenance of an international reference value. However, the consistent differences between machines are larger than desirable for this "International Master Batch" approach.

References

[1] McCowan, C. N., Pauwels, J., Revise, G., and Nakano, H., "International Comparison of Impact Verification Programs," *Pendulum Impact Testing, A Century of Progress: STP 1380*, edited by T. A. Siewert and M. P. Manahan, ASTM, PA, 2000.

[2] Montgomery, D. C., *Design and Analysis of Experiments*, John Wiley & Sons, New York, 1984.

[3] Cox, M. G., "The Evaluation of Key Comparison Data," *Metrologia* 39, 589-595 (2002).

[4] Ruth, E. A., "Striker Geometry and its Effect on Absorbed Energy," *Pendulum Impact Machines: Procedures and Specimens for verification, STP 1248*, edited by T. A. Siewert and A. K. Schmider, ASTM International, PA, 1995.

[5] Nanstad, R. K. and Sololov, M. A., "Charpy Test Results on Five Materials and NIST Verification Specimens Using Instrumented 2 and 8 mm Strikers," *Pendulum Impact Machines: Procedures and Specimens for Verification, STP 1248*, edited by T. A. Siewert and A. K. Schmider, ASTM International, PA, 1995.

[6] Revise, G., "Influence of Dimensional Parameter of an Impact Test Machine on the Results of a Test," *Charpy Impact Test: Factors and Variables, STP 1072*, edited by J. M. Holt, ASTM International, PA, 1990.

[7] Siewert, T. A. and Vigliotti, D. P., "The Effect of Charpy V-Notch Striker Radii on the Absorbed Energy," *Pendulum Impact Machines: Procedures and Specimens for Verification STP 1248*, edited by T. A. Siewert and A. K. Schmider, ASTM International, PA, 1995.

[8] Yanaka, M. et al., "Effects of Striking Radius and Asymmetrical Strikes on Charpy Impact Test Results," *Pendulum Impact Machines: Procedures and Specimens for Verification, STP 1248*, edited by T. A. Siewert and A. K. Schmider, ASTM International, PA, 1995.

[9] Roebben, G., Lamberty, A., and Pauwels, J., "Certification of Charpy V-notch Reference Test Pieces at IRMM," *presented at the Second Symposium on Pendulum Impact Testing: Procedures and Specimens*, November 8–9, 2004, Washington DC.