C.N. McCowan,[1] C.M. Wang,[2] and D.P. Vigliotti[1]

Charpy Impact Verification Data (1994-1996): A Summary

Reference: McCowan, C. N., Wang, C. M., and Vigliotti, D. P., **"Charpy Impact Verification Data (1994-1996): A Summary,"** *Journal of Testing and Evaluation*, JTEVA, Vol. 27, No. 2, March 1999, pp. 89-99.

Abstract: We present a summary of Charpy impact verification test data that were evaluated by the National Institute of Standards and Technology from January 1994 to December 1996. The Charpy impact machines that met the verification requirements of ASTM Test Methods for Notched Bar Impact Testing of Metallic Materials (E 23) are broken down by year and by reference lot. Based on the data, a proposed verification rule that limits the range of the verification set has been examined. We also present the results for determining whether two energies (lower and upper ends of the machine capacity) or three (lower, middle, and upper) are needed to verify the performance of the large-capacity impact machines; currently E 23 requires three energies to be tested.

Keywords: Charpy V-notch, impact certification program, impact testing, notched-bar testing, pendulum impact machines, reference specimens

This report provides a summary of the Charpy impact verification data that were evaluated by the National Institute of Standards and Technology (NIST) from January 1994 through December 1996. An indirect verification program has been used to verify the performance of Charpy impact machines for more than 40 years [1,2], and in 1964 ASTM Standard E 23 [3] was revised to require verification testing. NIST has provided the verification specimens and administered the program since 1989. In this program, impact machines are verified annually to the requirements in E 23, and the verification data, which are generated by organizations that own test machines, are returned to NIST for evaluation.

The impact verification program can be divided into three basic parts: production and distribution of impact verification specimens; verification testing; and evaluation of the verification test data. Before verification testing, a reference value for the impact toughness of the verification specimens is determined, and the uncertainty associated with the reference value is confirmed to be below a limit that ensures the material homogeneity of the specimens. In practice, the reference value for the impact specimens is determined by testing a random sample of 75 specimens from a production lot, which normally contains 1200 heat-treated specimens.

The 75 specimens are divided into groups of 25 and tested on three machines that have been defined in E 23 as the reference impact machines for the United States. The impact toughness, defined as the energy absorbed in the test, is the average absorbed energy for the75 tests. If the

1National Institute of Standards and Technology, Materials Reliability Division, Boulder, CO 80305
[2] National Institute of Standards and Technology, Statistical Engineering Division, Boulder, CO 80305

lot meets the statistical requirements of ASTM Practice for Qualifying Charpy Verification Specimens of Heat-Treated Steel (E 1271), a reference value is assigned to the lot [4]. Once a lot is accepted, sets of five specimens, or verification sets, are sold to companies that want to verify the performance of their Charpy impact machines. The specimens are broken using the candidate machine, and the broken specimens, along with the absorbed energy results, are sent to NIST for analysis. If the results are within 5 or 1.4 J of the reference value and the markings left by the machine on the specimens indicate the machine is in good working condition, the candidate machine is certified by NIST to meet the requirements of E 23.

A database containing the results of verification tests has been collected that includes the serial number of the candidate machine, the capacity and the pendulum design of the machine, the energy obtained for each specimen tested, the reference energy for the specimens tested, and the date of the test. The principal use of these data is to track the performance of individual impact machines and to monitor the verification program. The data also provide an opportunity for cross-validation of the current acceptance criteria and evaluation of new criteria proposed for the verification of Charpy impact machines.

Data from the verification tests and pilot lot evaluations are presented here to provide a general overview of how the Charpy verification program works in practice. To do this, tables and graphics are employed to show the proportions of test results that meet the current verification requirements and to consider the influence of the verification specimen on the test results. The data are also used to examine a newly proposed range rule for verification tests. Finally, the data are used to evaluate the requirement in E 23 to test at three energies (lower, middle, and upper ends of the machine capacity) when verifying large impact machines over their full capacity. The three energies available for verification testing are referred to as low, high, and super-high energy specimens in this paper. There is an interest in determining whether only two energies (low and super-high) are sufficient to verify the performance of the large-capacity impact machines.

Materials and Procedures

Charpy impact verification specimens are sold by NIST at three energies: the low-energy specimens, with energies near 17 J, the high energy specimens, with energies near 100 J, and the super-high energy specimens with energies near 225 J. The low- and high-energy specimens are made from 4340 steel, which is heat treated to produce specimens at the appropriate absorbed energies. The super-high-energy specimens are made using a T-200 maraging steel. The data used here include the customer data from many different lots of low-, high-, and super-high-energy specimens. The total number of verification tests for the low-, high-, and super-high-energy levels are 2401, 2385, and 655, respectively.

If the five absorbed energy measurements from the Charpy machine being tested are denoted by e_1, e_2, e_3, e_4,, e_5, and E_c is the average; of e_i, i = 1, 2, ···, 5, then the low-energy certification criterion is given by:

$$-1.4 \leq E_c - E_r \leq 1.4$$

where E_r is the reference energy of the lot. To certify a candidate Charpy machine in the high- or super-high-energy ranges, the difference between E_c and E_r must be within 5 % of the reference value; that is, it requires:

$$-0.05 \leq (E_c - E_r)/E_r \leq 0.05$$

Results

Low Energy

Figure 1 plots the difference between the customer's average and the reference value, $E_c - E_r$, for the low-energy verification tests. The dotted lines indicate the 1.4 J pass or fail criteria. In the right margin, the density estimate for $E_c - E_r$ is also plotted. The distribution of $E_c - E_r$ is centered around 0 and has a longer "tail" on the positive side of $E_c - E_r$. Out of 2401 verification tests shown in **Fig. 1**, 283 tests (11.8 %) failed the indirect verification. **Figure 1** also shows that when a machine fails to meet the low-energy verification requirements, it fails more often with $E_c - E_r > 1.4$ (9.2 %) than with $E_c - E_r < -1.4$ (2.6 %). This is expected because most of the common factors, such as anvil radius, bearings, and mounting, that wear or loosen over time increase the energy absorbed by the machine during the test.

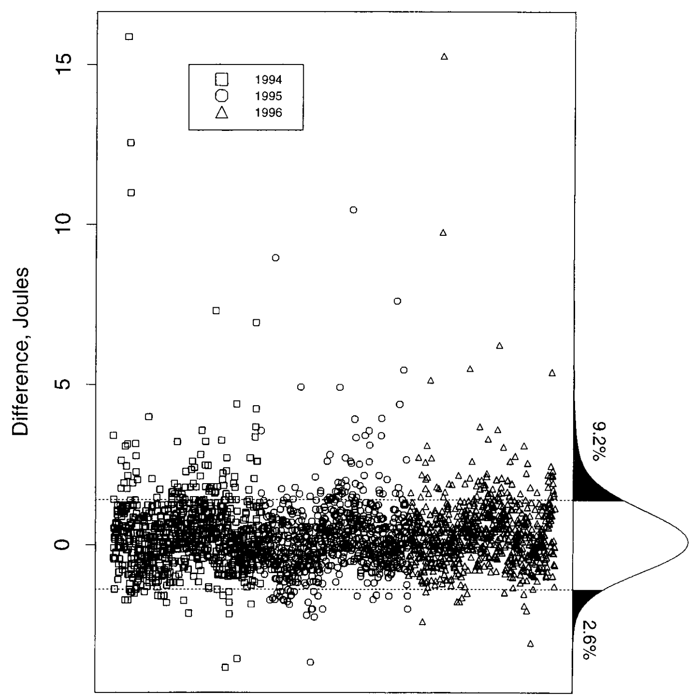

Figure 1. The difference, in J, between the customer's average for low energy verification tests and the reference value. The horizontal axis is the sequence of the test. The order within each year is not relevant. The dotted lines indicate the 1.4 J pass/fail criteria. The density estimate for the differences is plotted in the right margin.

Table 1. Pass/fail data for low energy verification tests (by year).

Year	Status	$E_c \leq E_r$	$E_c > E_r$	Total
1994	Pass	314 (39.3%)	388 (48.5%)	702 (87.8%)
	Fail	20 (2.5%)	78 (9.8%)	98 (12.2%)
1995	Pass	350 (43.6%)	359 (44.7%)	709 (88.3%)
	Fail	31 (3.9%)	63 (7.8%)	94 (11.7%)
1996	Pass	294 (36.8%)	413 (51.8%)	707 (88.6%)
	Fail	11 (1.4%)	80 (10.0%)	91 (11.4%)
Total	Pass	958 (39.9%)	1160 (48.3%)	2118 (88.2%)
	Fail	62 (2.6%)	221 (9.2%)	283 (11.8%)

In **Table 1**, the pass/fail data are broken down into test year. The number inside the parentheses is the percentage of the respective yearly total. The percentages of the row "Total" are with respect to the total number of the tests. The data show that the failure rate for the low-energy verification tests remains fairly constant (12.2 %, 11.7 %, and 11.4 %) from year to year.

In **Table 2**, the pass/fail data are broken down by the lot designation of verification specimens tested. The first column is the lot identification; the second column is the reference value of absorbed energy (J) of the lot; the third column is the standard deviation (J) of the pilot lot (calculated as the square root of the weighted mean of the variances of the three machines with weights equal to the number of observations, i.e., the pooled estimate); the fourth column is the lot size, or the number of machines tested using the verification specimens from that lot; the next two columns are, respectively, the number of machines that passed the verification test on the low and high sides; and the last two columns are the number of machines that failed the test and the failure rate. Only lots for which at least 100 verification tests were made are listed here. We present the specimen lot data in this way to evaluate the influence of the specimens on the outcome of the verification tests. The standard deviation of the pilot lot is one of the primary subtle trend for increasing failure over the small range in standard deviation present in the data. Unlike the yearly data, where the same population of machines is compared (since machines are tested annually), the failure rates of lots are expected to vary because the number of machines

Table 2. Pass/fail data for low energy verification tests (by lot).

Lot	Reference Energy	Standard Deviation	Lot Size	No. of Passes Low	No. of Passes High	No. of Failures	Failure Rate
44	17.2	0.62	132	46	70	16	12.1%
45	17.4	0.52	167	62	90	15	9.0%
46	17.4	0.58	165	90	69	6	3.6%
47	17.2	0.50	174	90	61	23	13.2%
48	17.6	0.72	162	79	67	16	9.9%
49	17.6	0.65	172	59	90	23	13.4%
51	16.9	0.71	148	61	64	23	15.5%
52	17.2	0.71	129	63	54	12	9.3%
53	16.7	0.57	130	49	70	11	8.5%
54	16.8	0.65	114	16	72	26	22.8%
55	17.5	0.84	117	48	54	15	12.8%
56	18.2	0.73	108	33	65	10	9.3%

tested for each lot is some different fraction of the machine population tested each year. So the acceptance criteria for the lots, and lower variation should result in less influence by the specimens on the outcome of the test. The fact that there is not a trend of increasing failures with increasing standard deviation in **Table 2** is not surprising, however, because much of the variation in the failure rates of the lots is probably due to sampling, which would obscure any standard deviation cannot be related to the failure rate for these data, and the failure rate alone should not be considered to have too much significance.

Probably the best indicator of specimen influence on the verification test we have is the distribution of the pass (low or high) data shown in **Table 2**. These data, monitored periodically as they are accumulated on each lot, show how many test results were higher or lower than the reference value assigned to the verification specimens being tested. A very skewed result here indicates that the reference energy assigned to the verification specimens could be very different from the average of the test results for the *good* machines. We would not expect the two averages to be identical, but if the *good* machines tested are a representative sample of the population of *good* machines, we typically find good agreement between the two averages. So, when the pass data are very skewed and the failure rate is high, as for Lot 54, we suspect the lot might be influencing the verification test results.

To more fully evaluate the data for Lot 54, the distribution in energy for the customer data is compared to the distribution of the pilot lot data in **Fig. 2**. It shows that the distribution of the customer data for Lot 54 is approximately Gaussian, with a slightly higher average energy than the pilot lot data (the average energy of the pass data for Lot 54 is 17.2 J). The peak around 18 J in the customer data is not considered to be part of the distribution of good machines. Rather, it defines a population of machines that failed the test and that differs from the population of *good* machines (both pass and fail data were included in the plot of the customer data). Although the apparent shift in average energy for the customer data may push some good machines across the pass/fail limit, most of the machines that failed appear to be representative of a population consisting of bad machines. Note also the almost bimodal shape near

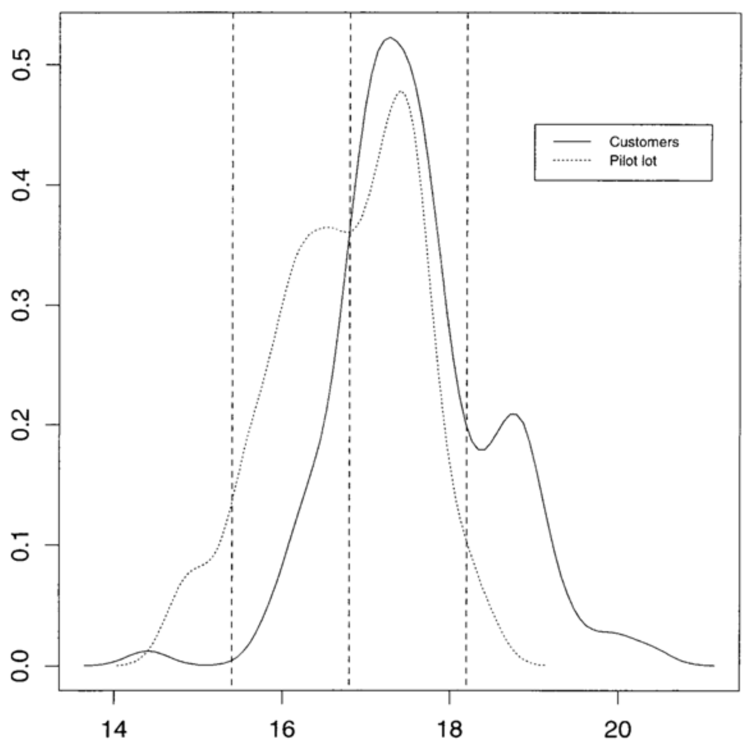

Figure 2. The density plots of the energy absorbed for the pilot lot data (from 75 individual specimens) associated with Lot 54. The three vertical dashed lines correspond to the reference energy (16.8 J) and the acceptance region (16.8 ± 1.4 J).

the peak of the distribution of the pilot lot data. This example appears somewhat extreme, but is not unexpected, because the pilot lot data are a compilation of data from three machines, each of which has characteristic differences and biases compared with the two others. For this pilot lot, two machines had very similar average energies (17.3 and 17.0 J), which differed from the average of the third machine (16.0 J). The combined data of the three reference machines serve as a good illustration of how the inclusion of different machines results in a balanced average reference energy for the specimens, but the distribution is broadened due to the bias between machines and does not provide a good measure of the inherent scatter of the specimens. To remove the machine bias and better estimate the scatter of the specimens, the pooled standard deviation of the three machines is used. For this pilot lot, the pooled standard deviation was 0.65 J, while the (not pooled) standard deviation for the (combined) data shown in **Fig. 2** was 0.87 J.

High Energy

Figure 3 plots the relative difference between the customer's average and the reference value, $(E_c - E_r)/E_r$, for the high-energy verification tests. The dotted lines indicate the 5 pass or fail criteria. In the right margin, the density estimate for $(E_c - E_r)/E_r$ is also plotted. The distribution of $(E_c - E_r)/E_r$ is centered around 0 and has a slightly longer "tail" on the positive side of $(E_c -$

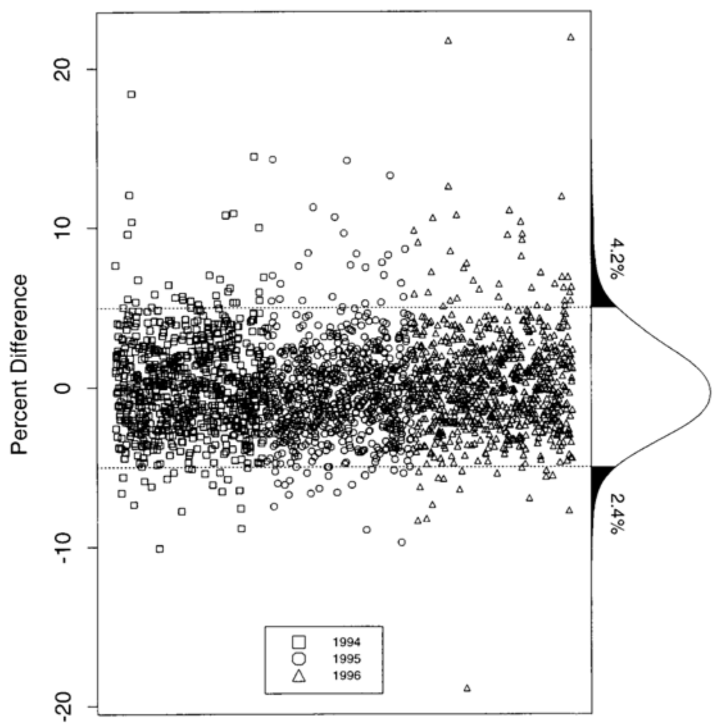

Figure 3. The relative difference between the customer's average for high-energy verification tests and the reference value. The horizontal axis is the sequence of the test. The order within each year is not relevant. The dotted lines indicate the 5% pass/fail criteria. The density estimate for the relative differences is plotted in the right margin.

Table 3. Pass/fail data for high-energy verification tests (by year).

Year	Status	$E_c \leq E_r$	$E_c > E_r$	Total
1994	Pass	363 (46.8%)	357 (46.1%)	720 (92.9%)
	Fail	21 (2.7%)	34 (4.4%)	55 (7.1%)
1995	Pass	418 (53.9%)	313 (40.3%)	731 (94.2%)
	Fail	19 (2.4%)	26 (3.3%)	45 (5.8%)
1996	Pass	397 (47.6%)	379 (45.4%)	776 (93.0%)
	Fail	17 (2.0%)	41 (4.9%)	58 (7.0%)
Total	Pass	1178 (49.4%)	1049 (44.0%)	2227 (93.4%)
	Fail	57 (2.4%)	101 (4.2%)	158 (6.6%)

Table 4. Pass/fail data for high energy verification tests (by lot).

Lot	Reference Energy	Standard Deviation	Lot Size	No. of Passes		No. of Failures	Failure Rate
				Low	High		
44	98.2	2.48	106	51	54	1	0.9%
45	99.1	2.62	160	76	79	5	3.1%
46	100.7	2.94	167	78	70	19	11.4%
47	108.6	3.42	169	125	19	25	14.8%
48	103.1	3.25	182	94	77	11	6.0%
49	102.6	3.16	183	65	106	12	6.6%
51	101.1	3.09	191	74	109	8	4.2%
52	102.2	2.96	123	49	70	4	3.3%
53	98.2	2.71	136	28	98	10	7.4%
54	99.8	2.70	136	70	62	4	2.9%
55	97.5	3.47	112	62	39	11	9.8%
56	106.7	3.58	121	65	41	15	12.4%

$E_r)/E_r$. Out of 2385 verification tests shown in **Fig. 3**, 158 tests (6.6 %) fail to pass the indirect verification requirements, with 4.2 % failing on the high side and 2.4 % on the low side.

In **Table 3**, the pass/fail data are broken down into test year. Like the low energy data, the failure rate for the high-energy verification tests remains fairly constant (7.1 %, 5.8 %, and 7.0 %) from year to year. The failure rate for the high-energy verification tests is consistently lower than that for the low-energy verification tests. It has long been recognized by E 23 that the very-high-strength, low-energy impact specimens show performance problems with machines that high-energy specimens do not. The difference in pass/fail percentages highlights this point.

Table 4 displays the pass/fail data by lot of verification specimens. Again, with the small number of machines tested under each lot, the failure rates vary from lot to lot. Considering the first three lots (44, 45, and 46), for example, the standard deviations and distributions of the pass data are similar, but the failure rates vary greatly. We assume that the explanation here is that a higher percentage of *bad* machines were tested using the Lot 46 specimens. However, for Lot 47, which has a very skewed pass data, we cannot necessarily attribute the high failure rate to sampling.

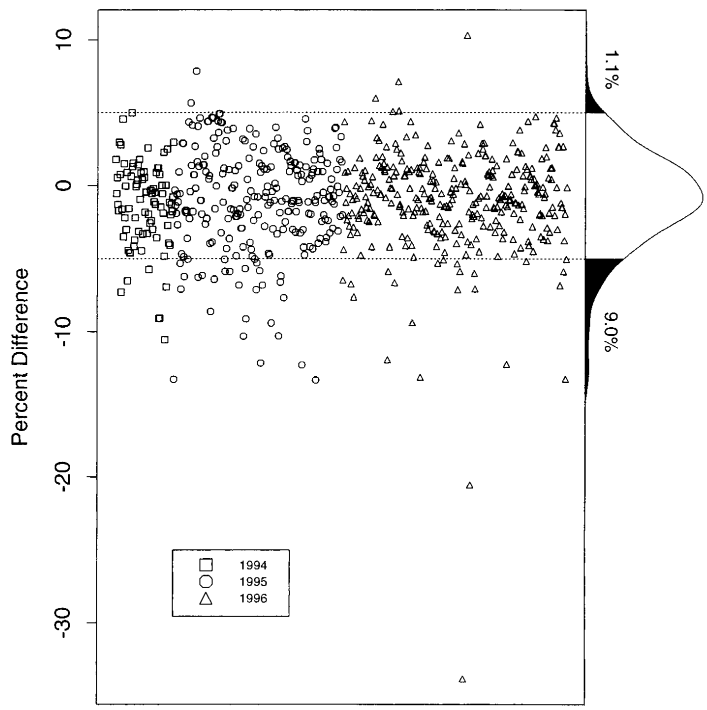

Figure 4. The relative difference between the customer's average for super-high-energy verification tests and the reference value. The horizontal axis is the sequence of the test. The order within each year is not relevant. The dotted lines indicate the 5% pass/fail criteria. The density estimate for the relative difference is plotted in the right margin.

Super-High Energy

Figure 4 plots the relative difference between the customer's average and the reference value, $(E_c - E_r)/E_r$, for the super-high energy verification tests. The dotted lines indicate the 5 % pass or fail criteria. In the right margin, the density estimate for $(E_c - E_r)/E_r$ is also plotted. The distribution of $(E_c - E_r)/E_r$ has a longer "tail" on the negative side of$(E_c - E_r)/E_r$. This implies when a machine fails the super-high verification test, it tends to have a low E_c value. This differs from that of the low- and high-energy verification data and is suspected to be a result of the different specimen-anvil (and striker) interaction for the super-high-energy test. These very ductile specimens are deeply brinelled by the anvils and wrap around the striker during the test. Overall, 10.1 % of the 652 cases evaluated failed to meet the 5 % verification criteria of E 23.

Table 5 contains the pass/fail data broken down into test year for the super-high verification tests. The failure rate varies more for these data than it does for the high- or low-energy verification data, but this is a new test. The number of machines tested from year to year is not as constant as in the low- or high-energy tests, and many machines are being tested for the first

Table 5. Pass/fail data for super-high-energy verification tests (by year).

Year	Status	$E_c \leq E_r$	$E_c > E_r$	Total
1994	Pass	47 (56.0%)	30 (35.7%)	77 (91.7%)
	Fail	7 (8.3%)	0 (0.0%)	7 (8.3%)
1995	Pass	117 (47.8%)	97 (39.6%)	214 (87.3%)
	Fail	29 (11.8%)	2 (0.8%)	31 (12.7%)
1996	Pass	182 (55.8%)	116 (35.6%)	298 (91.4%)
	Fail	23 (7.1%)	5 (1.5%)	28 (8.6%)
Total	Pass	346 (52.8%)	243 (37.1%)	589 (89.9%)
	Fail	59 (9.0%)	7 (1.1%)	66 (10.1%)

Table 6. Pass/fail data for super-high energy verification tests (by lot).

Lot	Reference Energy	Standard Deviation	Lot Size	No. of Passes		No. of Failures	Failure Rate
				Low	High		
3	229.9	7.25	135	81	33	21	15.6%
4	226.3	6.78	159	73	74	12	7.6%
5	222.5	8.12	169	81	77	11	6.5%
6	224.8	5.59	158	92	52	14	8.9%

time at these energies. For these reasons, we anticipate that the 10.1 average failure rate for the super-high energy tests will decrease slightly in the years to come.

Table 6 displays the pass/fail data by lot of verification specimens. Lot 3 has the highest failure rate among the lots for which we now have data, but we suspect that many of the failures associated with this first lot were due to not testing the specimens at room temperature (as required). The rest of the lots have comparable failure rates.

Range Summary

The range of the five measurements in a verification test is defined as

$$R = \max\{e_i\} - \min\{e_i\}$$

A new verification rule limiting the range of the absorbed energy measurements for the low- and high-energy tests is being balloted for ASTM E 23. The range test is designed to detect excessive variation, that is, to identify machines that have very high scatter in their measurements and just happen to have mean energy values that agree with the reference energy. Splett and Wang [5] also proposed an alternate certification procedure that accounts for the lot and machine variations.

The ranges being considered for limiting the low- and high-energy tests are 5 and 15 J, respectively. So, for the low-energy test, a candidate machine would fail if its range were greater than 5 J even it passed the ±1.4 J criteria, and for the high-energy test, a candidate machine would fail if its range were greater than 15 J even it passed the ±5% criteria.

Based on the data from 1994 to 1996, these range limits appear reasonable **Figs. 5 and 6** display the range of the five measurements for the low- and high-energy verification tests. The solid markers designate the machines that fail the ± 1.4 J criteria in the low-energy test (**Fig. 5**) and ±5% criteria in the high energy test (**Fig. 6**). There are 16 tests with $R > 5$ in **Fig. 5** and 17 tests with $R > 15$ in **Fig. 6**. If the range rule were in use, nine additional tests would fail in the low energy test, increasing the failure rate from 11.8 % to 12.2 %. Similarly, eight additional tests would fail in the high-energy test, increasing the failure rate from 6.6 to 7.0 %. In both cases, the range rules would increase the failure rate by 0.4 %.

An alternate range rule would be to use the normalized range

$$R_n = R/E_r$$

and fail the candidate machine if $R_n > r$, where r is some specified limit. The range rule based on R_n enables us to have the possibility of using one limit for all the three energies. It also provides a useful interpretation for the rule. It can be shown (e.g., see Ref 6) that for samples of five observations from a Gaussian distribution

$$R \approx 2.33\, S$$

where S is the standard deviation. Thus,

$$R_n \approx 2.33\, S/E_r$$

and the rejection criterion $R_n > r$ is approximately equivalent to the rejection criterion

$$S/E_r > r/2.33$$

which can be interpreted to mean that (in additional to the regular "difference" criteria) a machine would fail the test if its noise-to-signal ratio is greater than $r/2.33$. For example, with $r = 25$, the threshold would be 10.73 %.

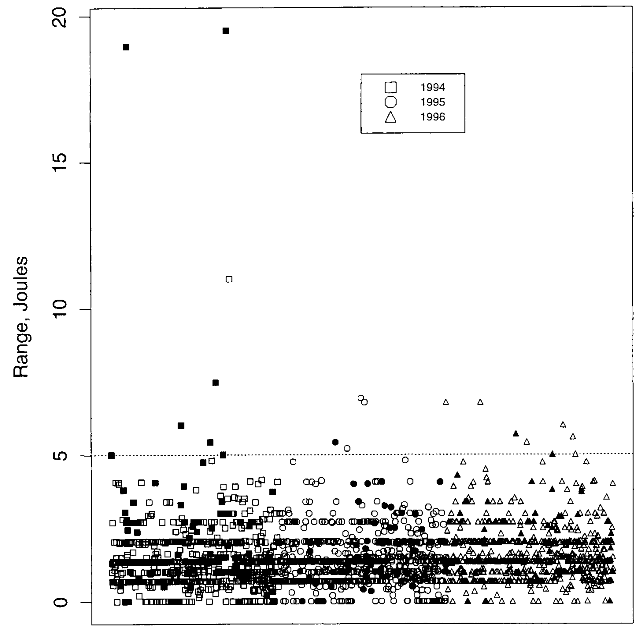

Figure 5. The range of the five specimens for the low-energy verification test. The horizontal axis is the sequence of the test. The order within each year is not relevant. The dotted line indicates the 5 J proposed rule. The solid markers designate the machines that fail the ±1.4 J criteria.

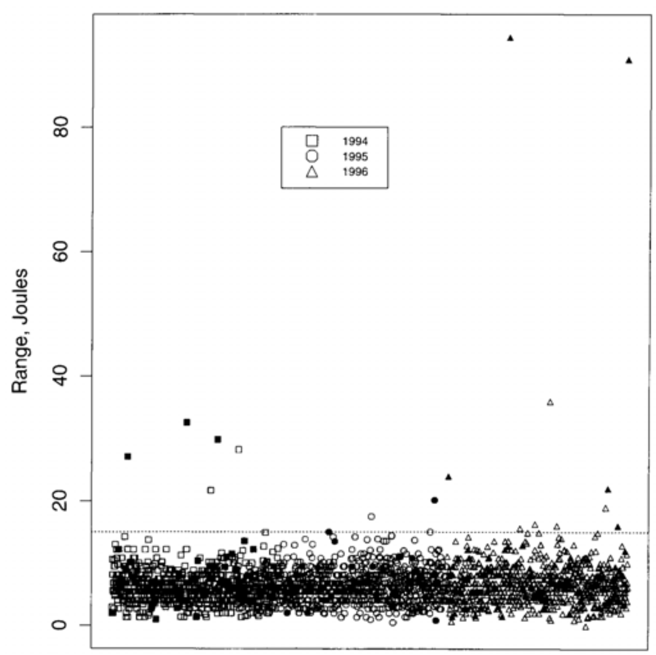

Figure 6. The range of the five specimens for the high-energy verification test. The horizontal axis is the sequence of the test. The order within each year is not relevant. The dotted line indicates the 15 J proposed rule. The solid markers designate the machines that fail the ±5% criteria.

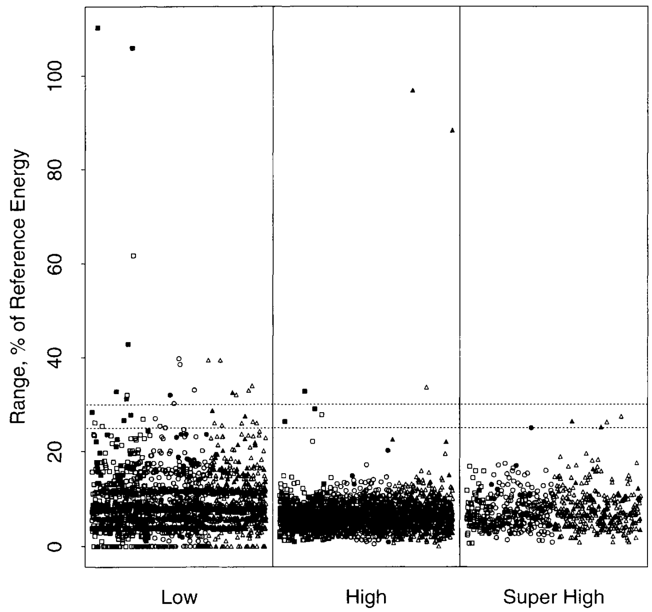

Figure 7. The normalized range of the five specimens for all the three energy verification tests. The solid markers designate the machines that fail the ±1.4 J or ±5% criteria. The dotted lines, at 25% and 30 %, are possible values of *r* to use in the range value.

Figure 7 plots R_n for all three energy verification tests. Again, the solid markers designate the machines that fail the ±1.4 J or ±5 % criteria. The two dotted lines indicate the two possible values of *r* to use: 25 and 30 %, which correspond to 10.73 and 12.88 % in the noise-to-signal-ratio scales. A 30 % R_n criterion would fail eleven additional tests (0.46 %) for the low energy, one additional test (0.04 %) for the high, and no additional test for the super-high energy.

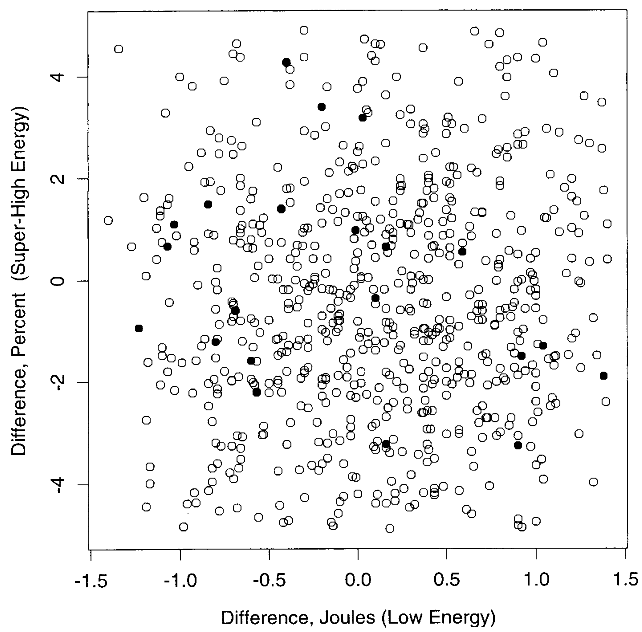

Figure 8. The scores of the low and super-high energy tests for cases that passed both tests. The "○" points are cases that also passes the high energy tests.

Energies Required for Verification Testing

Currently, the testing of the low, high, and super-high energies is required by ASTM Standard E 23 to certify candidate machines with capacities of greater than 289 J. However, the responsible ASTM subcommittee and task group has questioned whether the testing of the high energy is necessary in the certification of high-capacity machines. To address this question, we examine the verification tests from 1994 to 1996 for which the low-, high-, and super-high-energy tests were performed. **Figure 8** plots the values of $E_c - E_r$ (in the low-energy test) and $(E_c - E_r)/E_r$ (in the super-high test) for the cases that passed the low- and super-high-energy tests. The 698 "o" points are cases that also passed the high energy test. The 21 "●" points (2.92 %) are cases that failed the high energy test. Among the 21 cases that failed, 13 tests failed with $(E_c - E_r)/E_r$ between 5 and 6 %, and five tests failed with $(E_c - E_r)/E_r$ between 6 and 7 %. Thus, the successful result of the low- and super-high energy verification tests cannot ensure that a machine will pass the high-energy verification criteria of ±5 %, but in most cases, the results of the high-energy test are in good agreement with the reference energy.

Discussion

The failure rates for the low and high-energy verification tests are nearly constant, averaging 11.8 and 6.6 %, respectively. The failure rate for the super-high-energy test is more variable, in part due to the smaller and unequal number of tests performed in each year and the recent introduction of the test. Overall, the verification program appears to be functioning as described by those who originally implemented the program [7]. Data from a

1970 report [2] showed that initial failure rates for the impact verification program were 44 % (433 tests); but as machines were repaired and retested the failure rates began to decline sharply (11.5% was the lowest failure rate reported). Clearly the verification program has established and maintained a population of impact machines that can be reliably used for acceptance testing: more than 700 of the 800 machines tested annually in the program are within 1.4 J, or 5 % of the reference values, indicating that these machines differ from each other by less than 2.8 J, or 10 %.

The distributions of the data show significant numbers of machines near the limits of the pass/fail criteria, and we assume that some *good* machines failed the verification test. If the pass/fail criteria were widened to include more of these borderline machines, however, we suspect that the distribution of test results would broaden over time and a similar situation would develop near the new acceptance limits. In effect, it is the stringent pass/fail criteria adopted by E 23 many years ago that has resulted in the narrow distribution of impact test results in the program today. Frankly, any less stringent requirements would result in acceptance tests with little value, particularly when qualifying high-strength steels for severe environments. So accepting the current 1.4 J or 5 % pass/fail requirements as practical and necessary, we can strive only to decrease the probability of *good* machines failing the verification test and of *bad* machines passing the test.

The current acceptance criteria are based solely on the averages of the verification set and the pilot lot and do not take the variation of the data into account. The proposed range rule is a step in the right direction to help identify *bad* machines that would currently pass the verification test. Based on the low-energy data presented here, though, the probability of failing *good* machines with this rule because of one single outlying measurement is too high. We think it will be necessary to visually examine the specimens for jamming marks and evaluate the distribution of the five test results to more accurately identify machines with excessive variation. In addition, the proposed range rule can detect excessive variation only in the verification set and does not incorporate any of the information available on the pilot lot variation.

Further consideration of the pass/fail criteria and how they relate to the variations of the pilot lot data, our best indicator of variability, are needed. It has been proposed [5] that the candidate machine may be certified if $E_c - E_r$ is in interval (L, U) with

$$U = -L = d + 0.76 \cdot S \qquad (1)$$

where S is the pooled standard deviation of the pilot lot and d is a constant. The question is, what value should d be given. For illustration, we will use S = 0.7 J which is a typical standard deviation for the low-energy verification specimens (**Table 2**). With S = 0.7 J, U = 1.4 J, d is 0.9 J. Historically, low-energy lots have been accepted for distribution if the standard deviation of the pilot lot was 1 J or less [4], so when verification tests are conducted with specimens having a standard deviation of more than 0.7 J, say 1 J, these tests would have only to meet ±1.66 J criteria according to Eq 1. This is reasonable and would lower the probability of a *good* machine failing the verification test, but as we have already stated, to maintain a useful verification program we have accepted a maximum difference of 1.4 J for the low-energy range. So we need to reduce the maximum allowable standard deviation for the acceptance of specimen lots to increase the probability of certifying *good* machines and leave the pass/fail criteria at their present values. This is clearly the most direct and best method by which we can improve the impact verification program.

There is always uncertainty associated with the reference energy assigned to the lot. Overall, the average energies for pilot-lot data are in good agreement with the verification test results. It is, however, very difficult to evaluate the influence of the specimens on the verification data. In practice, we mitigate the influence of the specimens on the verification test by monitoring these data and visually inspecting the five returned specimens. For example, if a test fails to meet the 1.4 J or 5 % requirement and there are no markings on the specimens to indicate that the machine is in need of repair, we will retest the machine using different specimens if we have reasons to question the specimens used in the test. In practice this approach works well in minimizing the effect of the specimens on test results, but a better estimate of the mean would help avoid this problem.

To reduce the uncertainty of the reference energy assigned to verification specimens, the sample size can be increased and/or the sampling method can be improved. We think that increasing the sample size will improve our estimate only marginally. We are, however, considering changing from a random sampling method to a systematic sampling related to the positions of specimens in the heat-treating baskets. In addition, the control specimens placed in the sampling locations would be marked to identify their bar-stock origin (the ingot location of the bar stock is known, but we do not track the individual bars from which the samples are made). This type of sampling would allow us to include two variables of our processing in a consistent manner.

Generally, most machines that perform well at the upper and lower bounds of their capacities also perform as expected at mid-range energies. But three of machines failed the mid-range test, and **Fig. 8** shows that tests at any one energy alone provide limited predictability on how the machine will perform at the other energies. For example. **Fig. 8** shows the results for a machine that had nearly perfect performance (0 % difference) at both low and super-high energies, but the machine failed the high-energy test. This is due primarily to the fact that the specimens used to test at the three energy levels interact quite differently with the machine. For this reason we believe it is necessary to test at each energy level at which the machine will be used to have any certainty of the performance of the machine at that energy level.

Conclusions

Some conclusions based on the Charpy impact verification data that were evaluated by NIST from January 1994 through December 1996 are as follows:

1. The adoption of the stringent pass/fail criteria (1.4 J or 5 %), originally proposed for verifying Charpy impact machines in 1955, has produced a large population of impact machines that are suitable for acceptance testing.

2. The range rule now being considered for E 23 can detect excessive variation in measurements. More studies are needed before the rule is implemented.
3. Impact machines of large capacity should continue to be verified by testing at three energies.

Acknowledgment

This work is a contribution of the National Institute of Standards and Technology and is not subject to copyright in the United States.

References

[1] Driscoll, D. E., "Reproducibility of Charpy Impact Test," *Symposium on Impact Testing*, *ASTM STP 176*, American Society for Testing and Materials, Philadelphia, 1955, pp. 70-74.
[2] Fahey, N. H., "The Charpy Impact Test—Its Accuracy and Factors Affecting Test Results," *Impact Testing of Metals*, *ASTM STP 466*, American Society for Testing and Materials, Philadelphia, 1970, pp. 76-92.
[3] ASTM E 23-88, "Standard Test Methods for Notched Bar Impact Testing of Metallic Materials," American Society for Testing and Materials, Philadelphia.
[4] ASTM E 1271-88, "Standard Practice for Qualifying Charpy Verification Specimens of Heat-Treated Steel," American Society for Testing and Materials, Philadelphia.
[5] Splett, J. D. and Wang, J. C.-M., "Proposed Changes to Charpy V-Notch Machine Certification Requirements," Pendulum Impact Machines: Procedure and Specimens for Verification, *ASTM STP 1248*, American Society for Testing and Materials, Philadelphia, 1995, pp. 182-194.
[6] John, P. W. M., *Statistical Methods in Engineering and Quality Assurance*, John Wiley & Sons, Inc., New York, 1990.