Precision of Tests for ROOFING ASPHALT

If you're like me, at one time or another you've made a New Year's resolution to lose weight. The process for me begins on January 1st when I step on the scales to establish my baseline weight. Every week, I then return to the scales to determine my new weight and am sometimes confused by the results. Some weeks my weight will not have changed or I may have gained a pound or two, even though I worked out diligently and generally ate healthy foods. Other weeks, I'm traveling and have eaten poorly and not exercised, and yet my weight will stay the same or even drop by a pound.

Why does this happen? Well, fortunately as someone familiar with science, I know that each time I get on the scales constitutes a single test, the purpose of which is to generate a result (weight). Each test result (weight) is affected by a number of factors including whether or nor I've left my shoes on, or how much pizza I ate the night before. These are all examples of material variability, or variability that is inherent to the material (me) being tested.

A second source of variability may occur as a result of the equipment. Manufacturing tolerances for the load cell (or springs) of the scales can affect the measured result, as can the calibration of the scales. Another potential factor is the time of day when I get on the scales. I may have a different result if I weigh myself in the morning or in the evening. These are examples of testing variability. If I tightly control my test conditions (weigh immediately after getting out of bed in the morning and ensure that the scales have been recently calibrated within the expected weight range), then I can reduce the variability but not completely eliminate it.

Precisely Speaking

The American Society for Testing and Materials (ASTM) defines a test method as "…a definitive procedure for the identification, measurement, and evaluation of one or more qualities, characteristics, or properties of a material…that produces a test result." For any test method, we want to know how closely test results compare when repeated tests are conducted on the same material. Within-laboratory precision, which is often referred to as repeatability, is a comparison of test results conducted at separate times on the same material in the same lab. In our above example, this would be me stepping on the scales, obtaining a reading, and then repeating the process in a few minutes after I've brushed my teeth. Between-laboratory precision, which is often referred to as reproducibility, is a comparison of test results conducted at separate times on the same material in a different lab. This would be exemplified by me stepping on the scales, obtaining a reading at my house, driving to the gym, and then stepping on the scales at the gym and obtaining a reading.

For testing laboratories, we are concerned not only about repeatability within our lab, but also our reproducibility with other labs. In this instance, a statistical term called the difference two-sigma limit, or d2s for short, is used to provide an estimate of the acceptable range of test results from two labs that can be expected as a result of normal testing variability. It is determined by multiplying the calculated standard deviation (s) by $2\sqrt{2}$ or 2.83. In cases where the standard deviation is proportional to the average for different levels of the measured property, the d2s% is calculated as the d2s divided by the average of the two test results.

The d2s and d2s% limits are useful in evaluating test results because it lets us know how to respond to a set of test data in our lab compared to another lab. If the difference between the test results in Lab A and Lab B are within the d2s limits, then both results could be considered "correct" and the "actual" value could be somewhere between the two results. If, however, the difference between the test results in Lab A and Lab B exceeds the d2s limits, then a retest, and possibly re-evaluation of test equipment and procedures, may be needed. Needless to say, the smaller the d2s or d2s%, the more reproducible the test result.

Testing Variability of Roofing Asphalts

In 1999, Paramount Petroleum Corporation started a round robin program for evaluating the variability in some of the most common tests used in the Standard Specification for Asphalt Used in Roofing (ASTM D312). The purpose was to allow roofing asphalt suppliers and other testing labs the opportunity to compare their lab results to the average results from a larger group of labs. A repeated high variation from the average could be an indication of equipment or procedural problems within a lab.

BINDERS

By R. Michael Anderson, P.E.
and Steve Burhans

In 1999, Paramount supplied samples of a roofing asphalt binder (Type III) to 20 labs, with 18 labs supplying test data on seven tests in ASTM D312 and an additional test on rotational viscosity. In the first round robin, tests included:

ASTM	Test Method
D36	Softening Point (Ring and Ball)
D92	Flash Point (Cleveland Open Cup)
D5	Penetration @ 32ºF, 77ºF, and 115ºF
D113	Ductility @ 77ºF
D4402	Viscosity Determination using a Rotational Viscometer @ 400ºF

In later programs after 1999, the ductility test and penetration tests at 32ºF and 115ºF were discontinued because of poor response from the initial round robin labs. Less than half of the 1999 participants conducted these tests.

From its inception, the program continued to grow as samples were supplied and analyzed by Paramount Petroleum on approximately a six-month schedule. Figure 1 shows the participation in the program.

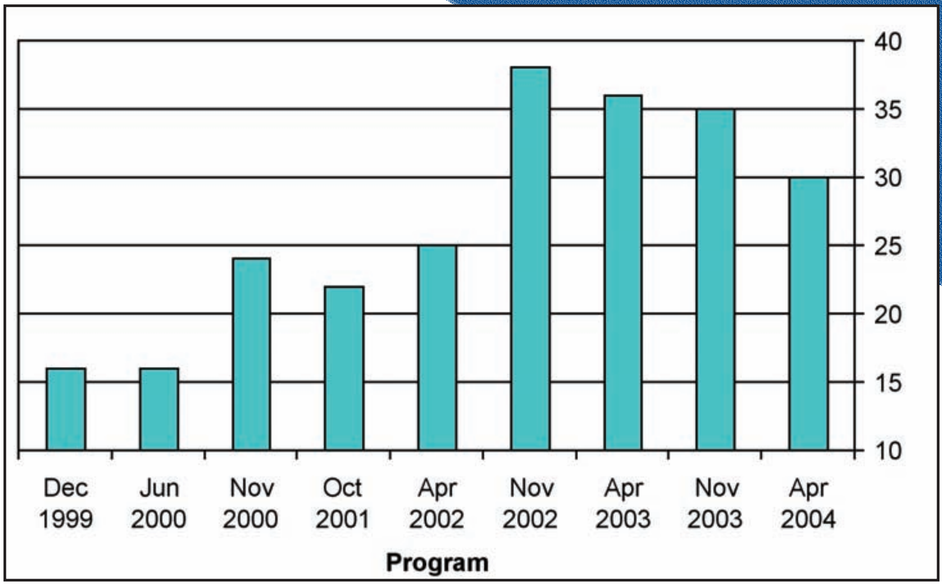

FIGURE 1: NUMBER OF PARTICIPATING LABS (PENETRATION TEST)

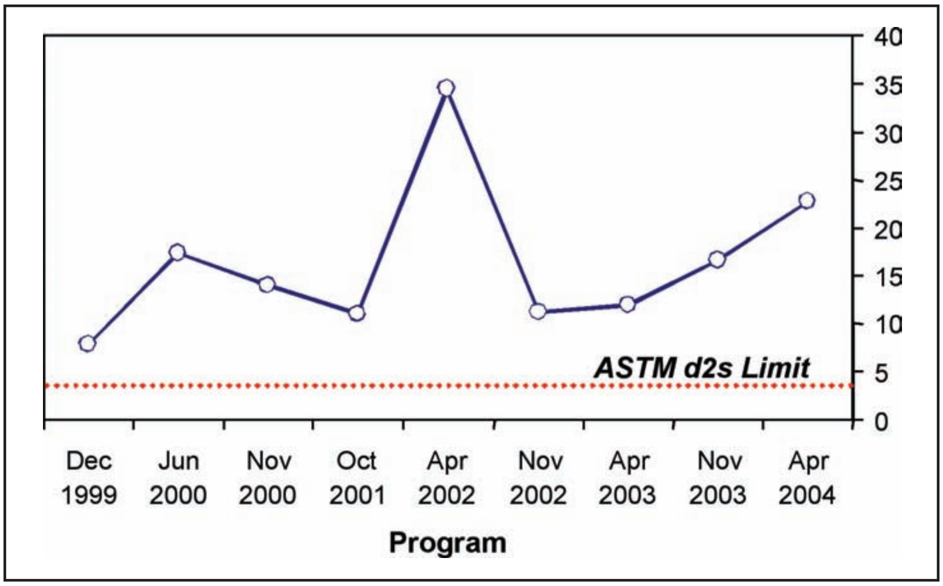

FIGURE 2: REPRODUCIBILITY LIMITS (d2s) FOR RING AND BALL SOFTENING POINT

The program also provided some interesting data regarding the reproducibility of the test results for the participating labs. While the results were, in general, more variable for all tests than the ASTM d2s limits, the variability was higher for some tests than others. For example, the ASTM d2s limit for the Ring and Ball Softening Point test is 3.5ºF. The average d2s limit for softening point for the round robin programs conducted from 1999 to 2004 was approximately four to five times greater (16.3ºF). The varying d2s calculated for each program is shown in Figure 2 along with the ASTM D36 limit. (It is interesting that the two highest d2s values, Apr 2002 and Apr 2004, occurred when polymer modified samples were tested.)

For other tests, such as penetration, the d2s calculated for each program (Figure 3) is much closer to the ASTM D5 limit. The average d2s limit for penetration for the round robin programs conducted from 1999 to 2004 was 7.9 dmm compared to the ASTM limit of 7.1 dmm.

The calculated d2s limits from the 1999-2004 round robin programs are shown in Table 1 along with the average d2s value (calculated from all 9 programs) and the ASTM d2s value. The higher variability indicates that some work is still needed within the participating labs to tighten down the between-lab variability of common tests for roofing asphalts. It is also worthwhile continuing the programs and analyses to determine if the testing variability is higher because of the material properties being tested (i.e., stiffer roofing asphalt binders compared to softer paving asphalt binders). If so, then a separate set of d2s limits for roofing asphalts may eventually be determined.

Raise the Roof

Beginning in 2005, the Asphalt Institute, with the cooperation of Paramount Petroleum, has taken the reins of the proficiency sample program. This year's program was started in June with reports being sent to the participants in early September (see *Lab Corner* on page 42 for more details). It is expected that the program will continue either annually or biannually depending on the interests of the participating labs.

Through increased knowledge of testing variability and its causes, we anticipate that the between-lab reproducibility will improve for the common roofing asphalt binder tests, as well as any future tests that may be used in characterizing the physical properties of roofing asphalts. To this end, we believe that the continued success of the Proficiency Sample Program for Roofing Asphalts will play a key role.

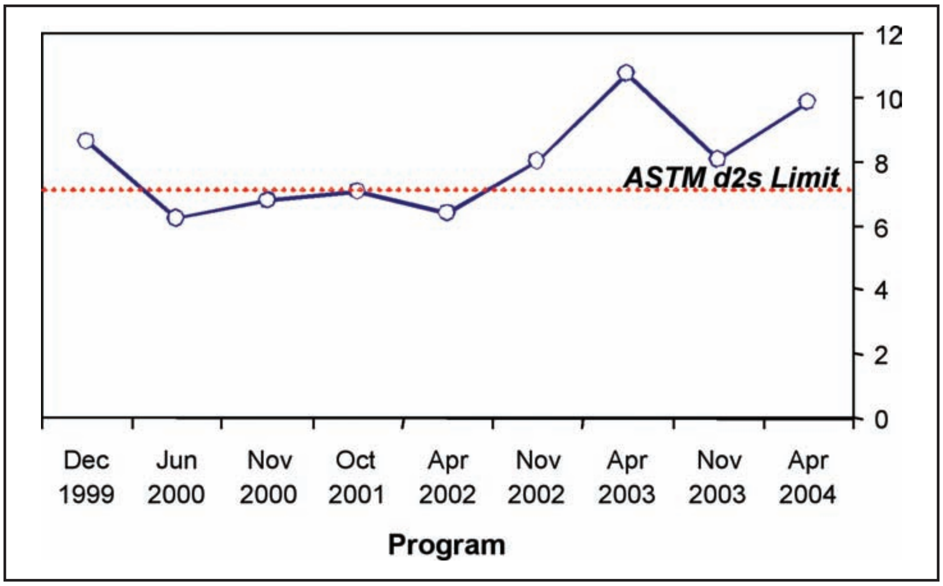

FIGURE 3: REPRODUCIBILITY LIMITS (d2s) FOR PENETRATION AT 77°F (25°C)

TABLE 1: MULTILABORATORY d2s LIMITS FOR ROOFING ASPHALT TESTS

Period	Softening Point, °F	Flash Point, °F	Penetration @ 77°F, dmm	Rot. Viscosity @ 400°F, cp	Rot. Viscosity
Apr-2004	22.7	52.2	9.8	131.2	42.9%
Nov-2003	16.6	58.3	8.0	171.1	45.2%
Apr-2003	11.9	69.4	10.7	68.1	40.9%
Nov-2002	11.1	90.8	8.0	131.9	43.0%
Apr-2002	34.5	51.7	6.4	45.4	23.4%
Oct-2001	11.0	66.4	7.1	95.7	38.2%
Nov-2000	13.9	55.0	6.7	124.3	65.7%
Jun-2000	17.2	50.5	6.2	210.0	74.3%
Dec-1999	7.7	66.5	8.6	90.8	44.0%
Average	16.3	62.3	7.9	118.7	46.4%
ASTM[1]	3.5	32.0	7.1		12.1%

1 The d2s limits for Rotational Viscosity come from the recently completed NCHRP 9-26 report, Web Document 71, entitled "Precision Estimates for AASHTO Test Method T308 and the Test Methods for Performance-Graded Asphalt Binder in AASHTO Specification M320".

2 Highlighted rows indicate polymer modified binder samples.

Mike Anderson is Director of Research and Laboratory Services for the Asphalt Institute. Steve Burhans is the Asphalt Technical Services Director for Paramount Petroleum.