A Case Study in Establishing Quality Assurance Limits for Automated Pavement Distress Data in North Carolina

Douglas J. Frith, P.E. (Corresponding Author)
Quality Engineering Solutions, Inc.
701 Jones Street
Reno, NV 89503

Shelley M. Stoffels, D.E.
Department of Civil and Environmental Engineering
The Pennsylvania State University
208 Sackett Building
University Park, PA 16802

Dennis A. Morian, P.E.
Quality Engineering Solutions, Inc.
405 Water Street, PO Box 3004
Conneaut Lake, PA 16316

Neil Mastin, P.E.
North Carolina Department of Transportation
4809 Beryl Road
Raleigh NC 27606

Submitted to the 9th International Conference on Managing Pavement Assets
Submission date: 12/08/2014

Word Count

Abstract:	165
Text:	3052
Figures:	1000 (4)
Tables:	750 (3)
Total:	4967 equivalent words

ABSTRACT
The North Carolina Department of Transportation (NCDOT) began collecting automated pavement distress data on a state-wide basis in 2012. Concurrently, they contracted the quality assurance reviews of the reported pavement distress to an independent source. This paper discusses the means and methods utilized by NCDOT and the quality assurance contractor to develop statistically valid quality assurance limits, which are also meaningful in terms of pavement management decision-making impacts, for the collected and processed asphalt pavement distress data. The paper describes the strategic selection of control sections to include a range and mix of the distresses with impact in the current decision trees, the data collection on the control sections, the rating methodology, and the rater pools and preparation to develop predicted limits for the control of the data. The paper discusses the consideration of multiple control indices, and the need to also reflect a range of values for those aggregate indices for multiple distresses, and presents the statistical analysis from the asphalt concrete control sites.

BACKGROUND

In 2012, the North Carolina Department of Transportation (NCDOT) began automated pavement distress data collection and processing for their statewide pavement program. Previously, pavement condition data was collected using windshield sampling techniques and methodologies. A data collection vendor (subsequently referred to as the vendor) was selected to collect, process, and deliver the pavement distress information on approximately 19,000 miles of roadway, including all Interstate and Primary routes. Concurrently, NCDOT selected another contractor (subsequently referred to as MP) to provide independent quality assurance (QA) efforts. One aspect of the quality assurance checks that will be discussed in this paper was the determination of quality limits for the QA checks to be performed throughout the duration of the data collection project. This paper presents a case study in the development and application of quality limits for asphalt pavement distress data.

The objective of this effort was to control the variability associated with the pavement distress identification so that the pavement management system recommendations for treatment will not be substantially affected by this variability. In order to determine the anticipated variability among the reported distresses for asphalt concrete pavements, the data from 14 control sites were evaluated. The 14 control sites were rated by a rater pool consisting of four MP raters, three NCDOT raters, and one vendor rating. All rating pool participants were familiar with both the high speed and windshield NCDOT rating guides.

These control site ratings were analyzed with respect to the decision trees used in the current NCDOT pavement management system (PMS) to determine the factors which had the most significant effect on the pavement treatment recommendations. MP then examined potential key elements to control the data, including specific individual distresses, the NCDOT Pavement Condition Rating (PCR), and other composite distress indices. Recommendations for improving and controlling the rating process were developed.

CONTROL SITE SELECTION

The selection of pavement distress control sites is critical to the process, as it is important to include pavement sections that have a variety of distress types and also a variety of extents and severities. The distress data collection manual was reviewed to identify the distress types and severities reported. As NCDOT was transitioning from their historical windshield method to high speed data collection, they were also implementing a high speed distress manual. For that reason, it was important to select control sites that met the criteria of both pavement distress guides. Guidelines for the control site selection were provided to NCDOT which included a matrix of distress types and severities. In addition, the sites were to be located within a day's travel from Raleigh, North Carolina. NCDOT staff reviewed past distress reports and other information, and identified 14 potential asphalt surfaced control sites located in seven different counties, representing four different Divisions of various lengths, as identified in Table 1.

Based upon the field review, matrices were developed (Tables 2 and 3), indicating what distress types and severities could be reported, and the corresponding control sites currently exhibiting those distresses. Table 2 shows the distresses valid for the windshield rating methodology *(1),* while Table 3 shows those for the high speed methodology *(2)*. Although not all cells in the matrix were covered with field occurrences, the most commonly occurring distresses were included.

TABLE 1 Potential Control Site Locations

OID	SURFACE TYPE	COUNTY NAME	COUNTY NUMBER	DIVISION	ROUTE	BEG_ MP	END_ MP	LENGTH (miles)
1	AC	Davidson	29	9	30000008	7.63	8.13	0.5
2	AC	Davidson	29	9	30000008	8.68	9.18	0.5
3	AC	Davidson	29	9	30000008	10.152	10.652	0.5
4	AC	Randolph	76	8	30000022	11.83	13.16	1.33
5	AC	Stokes	85	9	30000008	19.052	19.302	0.25
6	AC	Stokes	85	9	30000066	15.355	15.655	0.3
7	AC	Stokes	85	9	30000066	16.397	16.697	0.3
8	AC	Stokes	85	9	30000066	17.797	18.057	0.26
9	AC	Edgecombe	33	4	30000111	8.23	8.73	0.5
10	AC	Edgecombe	33	4	30000111	11.54	12.04	0.5
11	AC	Edgecombe	33	4	20000258	14.322	14.652	0.33
12	AC	Richmond	77	8	20000001	20.451	20.951	0.5
13	AC	Chatham	19	8	30000751	6.41	6.91	0.5
14	AC	Wake	92	5	30000054	9.573	10.073	0.5

TABLE 2 Matrix of Distress Occurrences on Asphalt Concrete Control Sites, Based upon the Windshield Distress Definitions

DISTRESS TYPE	SEVERITY		
	LIGHT	MODERATE	SEVERE
Alligator (Small Quantity)	4, 7, 10	2, 3	2, 11
Alligator (Large Quantity)	1, 5, 9, 12, 14	5, 11, 13, 14	
Transverse Cracking	4,5,6,9,10,11,12, 14	4, 5, 14	
Rutting	6, 13		
Raveling	6, 8, 13		
Bleeding	8		6, 7
Patching	5, 9		
Oxidation			

TABLE 3 Matrix of Distress Occurrences on Asphalt Concrete Control Sites, Based upon the High Speed Distress Definitions

DISTRESS TYPE	SEVERITY		
	LIGHT	MODERATE	SEVERE
Transverse	4,5,6, 13,14	4,5,7, 11,12,14	9,10,11
Longitudinal (Outside of WP)	12		
Longitudinal Lane Joint			
Alligator	1,2,3,4,5,7,9,10,11,12,14	2,3,4,5,9,11,12,13,14	2,11
Patching (WP)	5		
Patching (NWP)			
Delamination			
Bleeding	8		6,7
Rutting	6,12,13		
Raveling	6,8		
Transverse Reflective	12	12	
Longitudinal Reflective	12		

DATA COLLECTION

Data collection on the control sites was completed by a combined rating team consisting of two experienced pavement distress engineers from MP and at least two NCDOT Field Engineers. This team travelled to each location to perform a field review. The sections had been previously marked by District personnel. The rating team performed a pavement distress evaluation, conducted following the historical windshield methods. The windshield method consisted of driving the sections at a low rate of speed, qualitatively documenting the ride quality (as either low, moderate or high) and identifying the distresses observed. The distresses are reported in bins based upon distress type, severity, and quantity. Using the data bins combines severity levels and estimates quantities. After completing the windshield survey, the rating team would then stop and review the roadway from the shoulder. In many cases, additional distress was observed or multiple severities were noted that were not identified in the windshield pass. Although the windshield survey was not modified, extensive notes of these differences were reported.

The vendor's data collection vehicle travelled each of these 14 control sections over the next two months collecting the high speed data, which consisted of roadway geometrics, ride, rutting, and pavement images, collecting both range images and intensity images as part of their three-dimensional system. The images provided a permanent record of roadway conditions at the time of survey, isolating the effect of climatic changes from the data analysis process. The ride quality was collected as International Roughness Index (IRI) values for the sections, and then translated into low, moderate or high severity based upon vendor identified ranges. The vendor also provided distress ratings from the collected images using the their automated process.

Once the digital images were available, a rating pool was established, consisting of three NCDOT raters and four MP raters. These raters were provided with the distress rating manuals and training on the rating software. Each rater then performed an independent distress rating using the images from the 14 control sites. The data was summarized for each control section and then tabulated for further analysis.

ANALYSIS

The control site data analysis included three stages. First, the vendor submitted data was compared with the consensus field ratings. Second, the vendor data was compared with independent ratings completed by the seven members of the rater pool. Finally, statistical quality assurance limits were developed.

Comparison to Field Ratings

For the 14 asphalt surfaced control sites, the vendor identified more distress for 9 out of the 14 sites. A comparison of calculated PCR values is shown in the following chart (Figure 1). The PCR value used by NCDOT is calculated based upon the severity and quantity of the following distresses, using various weighting factors: Alligator Cracking, Transverse Cracking, Rutting, Raveling, Bleeding, Patching and Oxidation. Generally, the vendor identified larger percentages of alligator cracking at higher severities. They also tended to report more moderate ride quality, whereas the field raters always identified the ride as low roughness. On site 13, the vendor identified severe patching, whereas the field team identified no patching. The field raters tended to report low transverse cracking on a number of sections where no transverse cracking was reported from the vendor. Finally, there seems to be a difference in the severity of bleeding reported on sections 4 and 8 with both being reported by the vendor as severe bleeding.

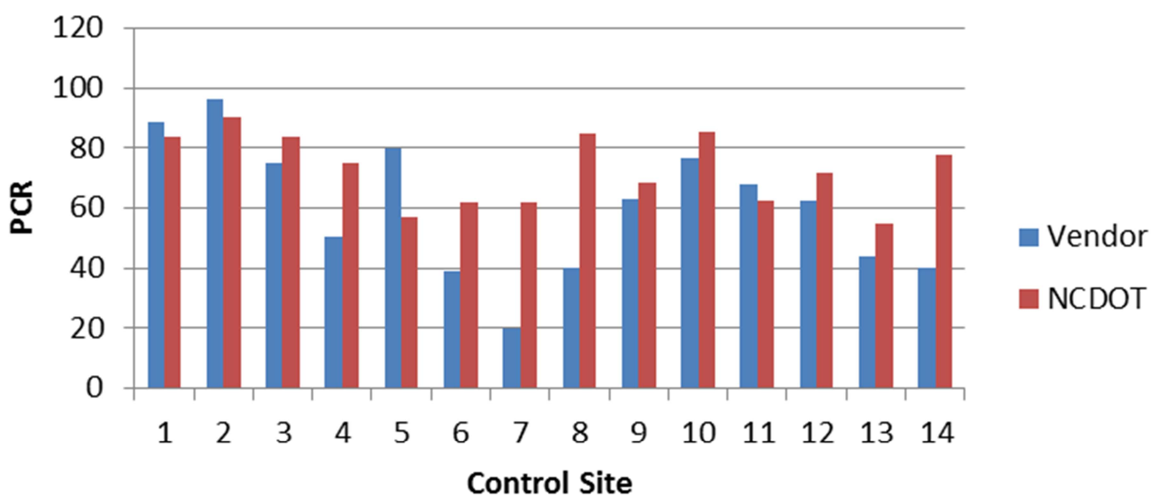

FIGURE 1 Comparison of windshield distress from vendor images versus field ratings.

As expected, the distress identified from the digital imaging system was generally more complete than that reported from the windshield surveys. However, the results from this data comparison identified some deficiencies in the distress definitions and rating process, which were corrected prior to production level data reporting. For example, the definition of bleeding was being misinterpreted, as well as the algorithm used to report transverse cracking.

Comparison to Image Ratings

The delivered control site data was then compared to the rater pool ratings, which had been completed by seven different raters (four from MP and three from NCDOT) by reviewing the digital images and providing a distress rating using the vendor provided rating software. These ratings were compared separately considering both the windshield and the high speed reporting definitions.

The comparison between the calculated PCR values based upon the windshield reporting method is shown in Figure 2. In this case, the vendor reported a lower distress rating (indicative of more distress) for 8 of the 14 control sites. The major differences between the ratings are that the rater pool generally rated higher quantities of low severity alligator cracking and the vendor generally rated higher quantities of bleeding. These differences led to clarification of some of the rating definitions and the algorithms used to summarize and report the data.

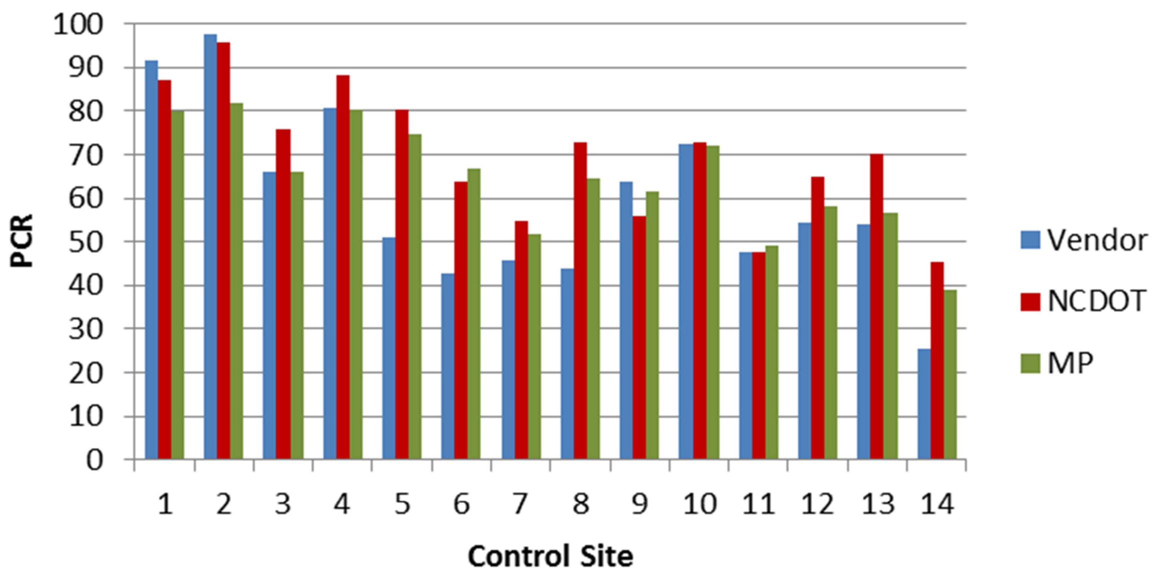

FIGURE 2 Results of windshield data summaries for the vendor and the rating pool.

The high speed data was also used to calculate two different index values, a load related distress index (LDR) and a non-load related distress index (NDR). These index values were calculated using definitions from VDOT. The LDR and NDR represent structural and environmental performance characteristics, respectively *(3)*. The LDR is associated with alligator cracking, patching, potholes and surface patching, while the NDR is associated with transverse cracking, longitudinal cracking, bleeding, raveling and block cracking. The results of this comparison are provided in Figures 3 and 4.

When comparing the high speed data, trends similar to those from the windshield comparison were evident. The rating pool identified more low severity alligator cracking, but a much smaller amount of longitudinal cracking than the vendor. This difference was improved by clearly delineating the two cracking types in the survey method. Regarding transverse cracking, the rater pool identified more cracking on almost every site. The other major difference was in the amount and severity of bleeding reported.

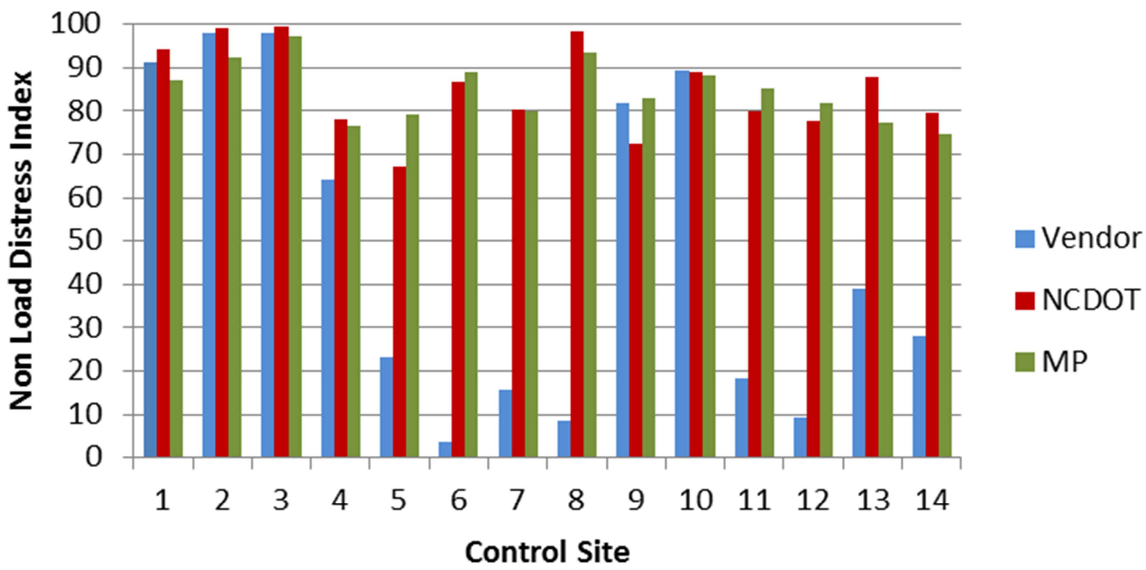

FIGURE 3 Comparison using the high speed data based upon the non-load distress index.

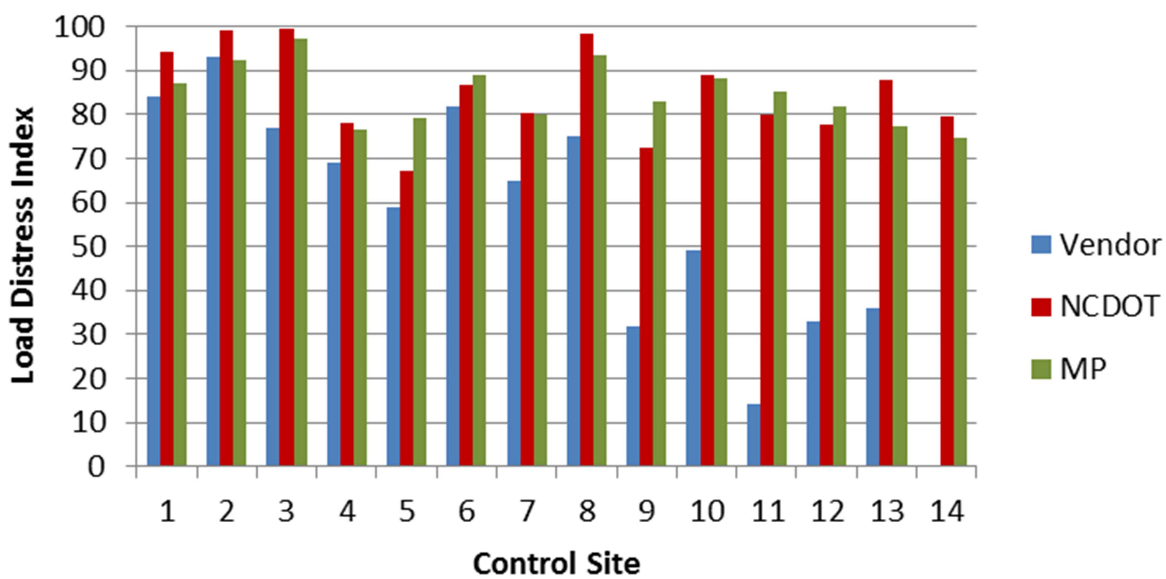

FIGURE 4 Comparison using the high speed data based upon the load distress index.

The control site comparisons resulted in the following recommendations from the quality assurance contractor (MP) to NCDOT and the vendor:

- The method of reporting transverse cracking must be better defined.
- The limits used by the vendor to determine ride quality rating needed to be reviewed.
- The rating/reporting of patching by the vendor needed further review.
- Differences in distress identification and classification existed between the vendor and the rating pool, especially in the areas of low alligator cracking, longitudinal cracking, patching and bleeding.
- Detailed distress rater training was recommended as it would likely result in lower data variability by assuring that all raters are using the same methods and definitions.

The vendor implemented several of these recommendations and applied the revised rating procedures and reporting algorithms to a much larger pilot set of data.

Determination of Quality Assurance Limits

The control site ratings were analyzed with respect to the decision trees used in the current NCDOT PMS to determine the factors which had the most significant effect on the pavement maintenance recommendations. MP then examined potential key elements to control the data, including individual distresses, combinations of distresses, and the NCDOT PCR. The NCDOT treatment decision process was carefully considered, and tested for various scenarios, to determine its sensitivity to variability in distress reporting. It was determined that using only the PCR for QA/QC would be inadequate for ensuring reliable recommendations.

In addition, the Virginia Department of Transportation (VDOT) indices of Load Distress Rating (LDR) and Non load Distress Rating (NDR) were computed from the high speed distress ratings *(3)*. The VDOT LDR and NDR were used as potential intermediate quality control parameters, falling somewhere between the detail of individual distresses and as single composite index, and have been successfully used as QA/QC parameters by VDOT *(4)*.

The statistical evaluation for distress items was based upon the ASTM methodologies for precision and bias *(5,6,7)*, recognizing that the control site evaluations represent a very small sample size. Using the analogy of rater groups to laboratories, the ASTM-defined d2s range has been applied to the distress rating results to determine the range of variability that can be considered acceptable for the NCDOT pavement condition ratings. The d2s is used for the direct one-to-one comparison of reported test results from two sources, not for the longitudinal comparison of results along a pavement. The application of this approach to pavement condition data has been documented in previous publications. *(4)*.

The "interlaboratory" d2s ranges (representing reproducibility of the process) have subsequently been utilized for the data quality assurance processes. Approximately 95% of all pairs of test results should be expected to differ in value by less than the d2s limit. During the quality assurance process, when comparing values reported from the production process to values from the MP raters, if more than 1 in 20 values exceed the established range limits, variability beyond that expected from a controlled process is being encountered.

A statistical evaluation was performed on the 14 control sites for each of the potential control variables to determine the allowable limit of variability between the vendor ratings and the rater pool ratings. This information was then applied to a pilot set of deliverable data to confirm the applicability of the control parameters and adequacy of the limits in consideration of the pavement treatment outcomes. Based upon this analysis, the following control limits have been applied by the QA contractor (MP) to the checked production data from the vendor:

- The difference in PCR values between the vendors reported data and MP determined data should not exceed the absolute value of 15.
- The difference in the total quantity of alligator cracking data reported (based upon the windshield reporting and summary method) should not exceed a value of 2.0.

When reported data exceeds these limits, detailed evaluations are conducted on the images and associated ratings to identify the source of specific difference between ratings. These differences can then be classified as either random errors or systematic errors. Systematic errors are those that seem to be the result of the process (for example, incorrect identification, classification, or summarization) and, therefore, could require entire data sets to be corrected. Random errors are those that do not occur in a consistent manner, but rather seem to apply in isolated instances. Explanations for random errors may be difficult to determine. Where random errors are noted, these section ratings must be corrected and resubmitted. When systematic errors are noted, large quantities of the data may need to be checked and resubmitted. Systematic errors may extend across NCDOT Divisions, requiring the reprocessing of large amounts of data.

RECOMMENDATIONS

Ultimately, it was determined that controlling the PCR and the quantity of alligator cracking (windshield method) should provide adequate control of the reported data. If the NCDOT treatment decision criteria are modified in the future, the sensitivity of the new criteria to variability in distress should again be assessed to ensure that additional controls are not required for reliable pavement asset management decisions.

As illustrated in this paper, the effort expended on the control sites proved extremely valuable in identifying differences in both the application of the distress definitions as well as the summarization and reporting of the distress data. The time and effort expended was minimal when compared to potential consequences of the delivery of entire data sets incorrectly. Utilizing a large rating pool aided in the discovery of interpretation and reporting issues that could be readily resolved through improved definitions and training. In addition, this information was used to develop statistically valid control and quality assurance limits that are meaningful for the NCDOT pavement management process.

REFERENCES

1. North Carolina Department of Transportation Pavement Condition Survey Manual, 2012.

2. NCDOT Digital Imagery Distress Evaluation Handbook, Version 1.0, North Carolina Department of Transportation, Pavement Management Unit, November 2011.

3. McGhee, K.H., Pavement Condition Indices for the Virginia Department of Transportation, Phase I, Flexible Pavements, September 2002.

4. Stoffels, S.M., Morian, D., Frith, D. and Larson, C., "Quality Analysis Methods for Pavement Distress Data," Proceedings of the Transportation Research Board Annual Meeting, January 2000.

5. Standard Practice for Preparing Precision and Bias Statements for Test Methods for Construction Materials, C670-10, ASTM International, West Conshohocken, PA.

6. Standard Practice for Use of the Terms Precision and Bias in ASTM Test Methods, E177-10, ASTM International, West Conshohocken, PA.

7. Standard Practice for Conducting an Interlaboratory Test Program to Determine the Precision of Test Methods for Construction Materials, C802-09a, ASTM International, West Conshohocken, PA.