PB2002-102150

Measurement and Specification of Construction Quality, Volume I

PUBLICATION NO. FHWA-RD-98-077

MAY 1998

U.S. Department of Transportation
Federal Highway Administration

Research and Development
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, VA 22101-2296

FOREWORD

This report, the first of a two-volume set of reports, presents the results of a study to better understand pavement construction quality. The three portland cement concrete pavements and three hot-mix asphalt pavements described in the report were subjected to extensive quality control/quality assurance sampling and testing during construction. Thus, knowing the actual quality level achieved on the projects, the researchers were able to evaluate current methods used by highway agencies to specify and estimate quality. The research findings and recommendations should help highway agencies make improvements in their quality assurance acceptance plans and specifications. The report will be of interest to engineers concerned with quality assurance, specifications, and pavement construction.

Sufficient copies of this report are being distributed to provide a minimum of three copies to each FHWA regional and division office, and 10 copies to each State highway agency. Direct distribution is being made to the division offices. Additional copies for the public are available from the National Technical Information Service (NTIS), 5285 Port Royal Road, Springfield, Virginia 22161.

Charles J. Nemmers, P.E.
Director
Office of Engineering
Research and Development

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-RD-98-077		

4. Title and Subtitle	5. Report Date
MEASUREMENT AND SPECIFICATION OF CONSTRUCTION QUALITY, VOLUME I: FINAL REPORT	
	6. Performing Organization Code
7. Author(s) C.S. Hughes, A.L. Simpson, R. Cominsky, O.J. Pendleton, R.M. Weed, T. Wilson	**8. Performing Organization Report No.**
9. Performing Organization Name and Address Brent Rauhut Engineering Inc. 8240 Mopac, Suite 220 Austin, Texas 78759	**10. Work Unit No. (TRAIS)** 3C5D **11. Contract or Grant No.** DTFH61-94-C-00078
12. Sponsoring Agency Name and Address Federal Highway Administration Office of Engineering Research and Development 6300 Georgetown Pike McLean, Virginia 22101-2296	**13. Type of Report and Period Covered** Final Report October 1994 - December 1997 **14. Sponsoring Agency Code**

15. Supplementary Notes

FHWA Contracting Officer's Technical Representative (COTR): Peter Kopac, HNR-30

Subcontractors: ERES Consultants, Inc., Champaign, IL; Infrasense, Inc., Arlington, MA; Law Engineering, Atlanta, GA

16. Abstract

This study consisted of testing six projects: three hot-mix asphalt concrete (HMAC) and three portland cement concrete (PCC). The primary objectives were to: (1) determine how current quality control test results vary in construction projects and how this variability affects pavement performance, (2) assess the suitability of current methods of quantifying materials and construction quality and quality variability, and (3) develop and improve methods that minimize current shortcomings.

Volume I of this report contains information about the data analysis. Volume II provides information about the data collected and use of the ground penetrating radar to determine thickness of the pavement. Volume II also contains univariate data on each of the tests performed on samples taken from the six projects in this study.

17. Key Words	18. Distribution Statement
Quality Assurance, Asphalt Concrete Pavement, Portland Cement Concrete Pavement, Ground Penetrating Radar	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	183	

Reproduction of completed page authorized

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yards	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: Volumes greater than 1000 l shall be shown in m^3.				
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5(F-32)/9 or (F-32)/1.8	Celcius temperature	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.71	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact)		
°C	Celcius temperature	1.8C + 32	Fahrenheit temperature	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

* SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.

(Revised September 1993)

TABLE OF CONTENTS

Chapter | Page

CHAPTER 1. INTRODUCTION 1
- **HISTORICAL BACKGROUND** 1
  - **Quality** 1
  - **The Problem** 2

CHAPTER 2. WORKING PLAN FOR CONDUCTING RESEARCH 7
- **WORKING PLAN** 7
  - **Task A - Conduct Literature Survey** 7
  - **Task B - Select Projects** 7
    - B.1. Coordination and Selection of Six Projects 8
    - B.2. Develop an Experimental Plan 8
    - B.3. Prepare Report Documenting Detailed Experimental Plan 9
  - **Task C - Conduct Testing** 9
  - **Task D - Perform Data Analysis** 9
    - D.1. Analyze the Data Collected in Tasks A & C 9
    - D.2. Assess the Implication of D.1 Findings on the Definition, Specification, and Measurement of Materials and Construction Quality 10
    - D.3. Perform Other Necessary Analyses in Order to Make Recommendations 11

CHAPTER 3. LITERATURE SEARCH 13
- **INTRODUCTION** 13
- **LEVEL OF QUALITY IN TODAY'S CONSTRUCTION** 14
  - **State Specifications** 14
    - Asphalt Concrete 15
    - Portland Cement Concrete 16
      - New Jersey 16
      - Oklahoma 17
      - South Dakota 18
      - Illinois 18
      - Michigan 18
      - West Virginia 19
      - Washington 19
      - Indiana 19
      - Minnesota 19
      - AASHTO 20
  - **Construction Quality** 20
    - Asphalt Concrete Pavement 21
    - PCC Pavements 22
- **QUALITY CHARACTERISTIC DISTRIBUTIONS** 24

TABLE OF CONTENTS (continued)

Chapter Page

Assumption of Normal Distribution . . . 24
Asphalt Concrete Pavements . . . 25
PCC Pavements . . . 26
QUALITY ASSURANCE SPECIFICATIONS . . . 32
Quality Assurance Risks . . . 33
Selection of Lot Sizes . . . 35
INFLUENCE OF SAMPLE SIZE ON QUALITY MEASURES . . . 38
TYPICAL PAVEMENT VARIABILITY . . . 40
Asphalt Concrete Pavement Variabilities . . . 40
PCC Material Variabilities . . . 41
RELATIONSHIPS BETWEEN QUALITY CHARACTERISTICS AND PERFORMANCE . . . 43
Performance Relationships . . . 44
Asphalt Concrete . . . 44
PCC . . . 50
SUMMARY . . . 57

CHAPTER 4. DESCRIPTION OF FIELD PROJECTS . . . 59
HMAC PROJECTS . . . 59
Duncan, Oklahoma Site No. 1 . . . 59
Duncan, Oklahoma Site No. 2 . . . 61
Baton Rouge, Louisiana . . . 61
PCC PROJECTS . . . 64
Mankato, Minnesota . . . 64
Rochelle, Illinois . . . 66
Bellefontaine, Ohio . . . 68

CHAPTER 5. SAMPLING AND TESTING . . . 71
SAMPLING AND TESTING FOR HMAC PAVEMENTS . . . 71
Test Sections (Lot Description) . . . 71
Bulk HMAC Samples . . . 71
Field Sampling and Testing . . . 73
Field Sample Code Designation . . . 73
Density . . . 73
Thickness . . . 73
SAMPLING AND TESTING FOR PCC PAVEMENTS . . . 73
Air Content . . . 74
Thickness . . . 74
Strength . . . 75
Smoothness . . . 75
Density . . . 75

TABLE OF CONTENTS (continued)

Chapter Page

FIELD TESTING . . . 75
FWD Testing . . . 75
GPR Testing . . . 76

CHAPTER 6. DATA ANALYSIS . . . 79
INTRODUCTION . . . 79
Acceptance Plans . . . 80
TESTS FOR NORMALITY . . . 80
Histograms . . . 82
Shapiro-Wilk Test, Skewness, and Kurtosis . . . 83
HMAC Projects . . . 86
PCC Projects . . . 87
SKEWBIAS Program . . . 87
Results of SKEWBIAS Program . . . 93
LOT SIZE AND BIAS RESULTS . . . 93
Precision, Accuracy, and Bias . . . 93
Selection of Lot Size . . . 95
Selection of Sample Size . . . 95
HMAC RESULTS . . . 98
HMAC Mixture Result Analysis . . . 98
Bulk Specific Gravity . . . 100
Maximum Theoretical Specific Gravity . . . 102
Air Voids . . . 103
Asphalt Content . . . 104
Aggregate Gradation . . . 107
19-mm Sieve . . . 107
13-mm Sieve . . . 109
10-mm Sieve . . . 110
5-mm Sieve . . . 111
2.36-mm Sieve . . . 113
1.18-mm Sieve . . . 114
0.30-mm Sieve . . . 116
0.15-mm Sieve . . . 117
0.075-mm Sieve . . . 118
HMAC Field Result Analysis . . . 120
Density . . . 120
GPR Thickness . . . 121
Core Thickness . . . 124
CORRELATION OF INTERRELATED HMAC PROPERTIES . . . 127
CONFORMAL INDICES FOR HMAC PROPERTIES . . . 128
PCC DATA . . . 129

TABLE OF CONTENTS (continued)

Chapter Page

PCC Mixture Result Analysis 129
Plastic Air Content 130
Compressive Strength 132
Unit Weight 133
PCC Field Result Analysis 135
Core Compressive Strength 135
Core Thickness 137
GPR Thickness 138
CONFORMAL INDICES FOR PCC PROPERTIES 144
CORRELATION OF INTERRELATED PROPERTIES OF PCC PROJECTS 149
RELATIONSHIP OF PAVEMENT VARIATION WITH PERFORMANCE . . 150
PROPOSED PROCEDURE FOR SEQUENTIAL ADDITION TO LOTS 152

CHAPTER 7. CONCLUSIONS AND RECOMMENDATIONS 155

REFERENCES 159

BIBLIOGRAPHY 171

LIST OF FIGURES

Figure		Page
1.	Decrease in Coefficient of Variation for Percent Passing the 4.75-mm Sieve	4
2.	Distribution of PCC Compressive Strength	27
3.	Distribution of PCC Core Thickness	28
4.	Histogram of 7-day Concrete Compressive Strength	28
5.	Histogram of 28-day Compressive Strength	29
6.	Histogram of Concrete Flexural Strength	29
7.	Second Histogram of Concrete Flexural Strength	30
8.	Histogram of Concrete Slab Thickness	30
9.	PCC Patching Amounts Versus Loading	51
10.	Pay Factor Versus COV of Compressive Strength	52
11.	Expected Pay Factor Versus COV of Slab Thickness	52
12.	Location of HMAC Test Sections, Duncan, OK, Test Site No. 1	60
13.	Location of HMAC Test Sections, Duncan, OK, Test Site No. 2	62
14.	Location of HMAC Test Sections, Baton Rouge, LA	63
15.	Location of PCC Test Sections, Mankato, MN	65
16.	Location of PCC Test Sections, Rochelle, IL	67
17.	Location of PCC Test Sections, Bellefontaine, OH	69
18.	Typical Statistical Characterization of One Quality Property from One Project	81
19.	Typical Display of Skewed Distribution by Program SKEWBIAS	90
20.	Linear Regression of GPR Thickness Versus Core Thickness for Project OK1	126
21.	Linear Regression of GPR Thickness Versus Core Thickness for All PCC Projects	145
22.	Linear Regression of GPR Thickness Versus Core Thickness for Project IL	146
23.	Linear Regression of GPR Thickness Versus Core Thickness for Project MN	147
24.	Linear Regression of GPR Thickness Versus Core Thickness for Project OH	148

LIST OF TABLES

Table — Page

1. Annual Average HMAC Densities in Virginia, % MTSG 5
2. Predicted Vs. Actual Samples within 1,2,3 σ Limits 26
3. Orders of Significance for Independent Variables, All Models for Rutting of HMAC Pavements 46
4. Orders of Significance for Independent Variables, All Models for Change in Roughness of HMAC Pavement 47
5. Orders of Significance for Independent Variables, All Models for Transverse Cracking in HMAC Pavements 48
6. Effects of Variables on HMAC Performance 49
7. Relationships Between Material Properties and Distress Development for PCC Pavements 53
8. Shapiro-Wilk (W), Skewness, and Kurtosis Results for HMAC Projects 84
9. Shapiro-Wilk (W), Skewness, and Kurtosis Results for PCC Projects 85
10. Typical Output of Program SKEWBIAS 91
11. Bias of Single-Tailed Percent Defective (PD) Estimates for Selected Values of Sample Size and Skew Coefficient 94
12. Descriptive Statistics for Bulk Specific Gravity 100
13. Required Sample Size to Attain Specific Bias Values for Bulk Specific Gravity 101
14. Descriptive Statistics for Maximum Theoretical Specific Gravity 102
15. Required Sample Size to Attain Specific Bias Values for Maximum Theoretical Specific Gravity 103
16. Descriptive Statistics for Air Voids, % (75-blow Marshall Compaction) 104
17. Required Sample Size to Attain Specific Bias Values for Air Voids, % 105
18. Descriptive Statistics for Asphalt Content, % 105
19. Required Sample Size to Attain Specific Bias Values for Asphalt Content, % 106
20. Descriptive Statistics for Asphalt Content From State and Study Tests 107
21. Descriptive Statistics for Percent Passing 19-mm Sieve 107
22. Required Sample Size to Attain Specific Bias Values for Percent Passing 19-mm Sieve 108
23. Descriptive Statistics for Percent Passing 13-mm Sieve 109
24. Required Sample Size to Attain Specific Bias Values for Percent Passing 13-mm Sieve 110
25. Descriptive Statistics for Percent Passing 10-mm Sieve 110
26. Required Sample Size to Attain Specific Bias Values for Percent Passing 10-mm Sieve 111
27. Descriptive Statistics for Percent Passing 5-mm Sieve 112
28. Required Sample Size to Attain Specific Bias Values for Percent Passing 5-mm Sieve 113
29. Descriptive Statistics for Percent Passing 2.36-mm Sieve 114
30. Required Sample Size to Attain Specific Bias Values for Percent Passing 2.36-mm Sieve 114

LIST OF TABLES (continued)

Table Page

31. Descriptive Statistics for Percent Passing 1.18-mm Sieve 115
32. Required Sample Size to Attain Specific Bias Values for Percent Passing 1.18-mm Sieve 116
33. Descriptive Statistics for Percent Passing 0.30-mm Sieve 116
34. Required Sample Size to Attain Specific Bias Values for Percent Passing 0.30-mm Sieve 117
35. Descriptive Statistics for Percent Passing 0.15-mm Sieve 118
36. Required Sample Size to Attain Specific Bias Values for Percent Passing 0.15-mm Sieve 119
37. Descriptive Statistics for Percent Passing 0.075-mm Sieve 119
38. Required Sample Size to Attain Specific Bias Values for Percent Passing 0.075-mm Sieve 120
39. Descriptive Statistics for Density Tests, % MTSG 121
40. Descriptive Statistics for GPR Thickness, mm 121
41. Required Sample Size to Attain Specific Bias Values for GPR Thickness, mm 123
42. Analysis of Variance Results from Linear Regression of GPR Thickness for HMAC Projects 124
43. Results for Parameters Used in Linear Regression of GPR Thickness for HMAC Projects 124
44. Descriptive Statistics for Core Thickness, mm 124
45. Results From Comparisons of GPR and Core Thickness Data for OK1 125
46. Significant Correlation Coefficient Values of HMAC Mixture Variables 127
47. Conformal Indices for HMAC Projects 129
48. Descriptive Statistics for Plastic Air Content, % 131
49. Required Sample Size to Attain Specific Bias Values for Plastic Air Content, % 131
50. Bias Values for Plastic Air Content, % 132
51. Descriptive Statistics for Compressive Strength, kPa 132
52. Required Sample Size to Attain Specific Bias Values for Compressive Strength, kPa 133
53. Bias Values for Compressive Strength, kPa 134
54. Descriptive Statistics for Unit Weight, kg/m^3 134
55. Required Sample Size to Attain Specific Bias Values for Unit Weight, kg/m^3 135
56. Bias Values for Unit Weight, kg/m^3 135
57. Descriptive Statistics for Core Compressive Strength, kPa 136
58. Required Sample Size to Attain Specific Bias Values for Core Compressive Strength, kPa 136
59. Bias Values for Core Compressive Strength, kPa 137
60. Descriptive Statistics for Core Thickness, mm 137
61. Required Sample Size to Attain Specific Bias Values for Core Thickness, mm 138
62. Bias Values for Core Thickness, mm 138
63. Descriptive Statistics for GPR Thickness, mm 139

LIST OF TABLES (continued)

Table | Page

64. Required Sample Size to Attain Specific Bias Values for GPR Thickness, mm 140
65. Bias Values for GPR Thickness, mm 141
66. Analysis of Variance Results from Linear Regression of GPR Thickness for PCC Projects 141
67. Results for Parameters Used in Linear Regression of GPR Thickness for PCC Projects 142
68. Results from Comparisons of GPR Thickness and Core Thickness for PCC Projects 143
69. Conformal Indices for PCC Projects 149
70. Significant Correlation Coefficient Values of PCC Variables 150
71. Example of Proposed Procedure for Approximating Normal Distribution by Sublot 153

CHAPTER 1. INTRODUCTION

HISTORICAL BACKGROUND

The topic of "quality" is currently receiving much attention within the highway community. This attention is demonstrated through several initiatives, a primary one being the National Quality Initiative (NQI), which is supported by such organizations as the Federal Highway Administration (FHWA), American Association of State Highway and Transportation Officials (AASHTO), American Road and Transportation Builder's Association (ARTBA), National Ready-Mixed Concrete Association (NRMCA), National Asphalt Pavement Association (NAPA), American Public Works Association (APWA), American Concrete Pavement Association (ACPA), American Consulting Engineers Council (ACEC), Associated General Contractors of America (AGC), the National Stone Association (NSA), and The Asphalt Institute (TAI). The NQI has been instrumental in supporting programs oriented toward the promotion of quality in AASHTO and State highway agencies (SHA's). Examples of this support are the recently issued *AASHTO Quality Assurance Guide Specification* and its companion document *Implementation Guide for Quality Assurance*, both dated February 1996,[1,2] and FHWA Demonstration Project 89, which has sponsored, among other efforts, a large number of quality assurance courses and workshops for SHA's, ACPA, and NAPA.

Quality

To make cost-effective improvements in highway construction, one must have a definition and a thorough understanding of the term "quality." Unfortunately, there has been no clear consensus of the definition for quality within the highway or transportation community. Although definitions, or perceptions of the definition, exist, they vary from organization to organization and are dependent on the organization's role in producing the final product. For this report, the definitions contained in Transportation Research Circular Number 457, *Glossary of Highway Quality Assurance Terms,* are used for consistency.[3] This document defines quality as: (1) the degree or grade of excellence of a product or service, (2) the degree to which a product or service satisfies the needs of a specific customer, and (3) the degree to which a product or service conforms with a given requirement. Using these definitions, most industries manufacturing raw materials consider quality in the following three broad areas:

- Quality of Design - defines the stringency of the specification (design requirements) for manufacture of the product.

- Quality of Conformance to Design - defines how well the manufactured product conforms to the original design requirements. Conformance quality is closely associated with the more standard term "quality control."

- Quality of Performance - defines how well the product works or performs.

Quality of performance is obviously the most important and is dependent on both the quality of design and quality of conformance. In other words, the best possible design can be used, but poor conformance controls can result in poor performance. Conversely, the best conformance controls cannot ensure good performance if the design is incorrect. Of the three, quality of design and quality of performance are the areas least understood by the highway construction industry; the quality of conformance to design has consistently received greater attention.

The Problem

The truth of the matter is that enough is not yet known about construction quality as related to performance. One reason is that, with the exception of a few SHA's, good systems to measure quality either have been nonexistent or only recently developed. This lack of good quality measurement systems has prevented the determination of whether quality of recent construction has changed from past years. Also, documentation of costs to provide a high quality product either does not exist or is buried so deeply that it has not been readily accessible. Lastly, it appears that sample sizes often are not large enough for accurate quality estimates in constructed projects.

Specific but limited studies have shown a reduction in material variability with time. For example, one report showed, through the accumulation of quality control and acceptance data, that the variability of hot-mix asphalt concrete (HMAC) materials decreased as contractors and SHA personnel began to become familiar with the operational characteristics of drum mix plants.[4] This reduction in material variability is illustrated in figure 1. Specifications and production standards gradually have been tightened as product quality improved after both contractor and SHA personnel improved skills as a result of long-term training initiatives. That reduction in mix variability of the gradation and other production parameters did result in a minor overall improvement in performance.[4]

Virginia tracked the effect changing its construction specifications had on asphalt concrete density levels being attained on the roadway.[5] The average density levels and variations over the observation period are summarized in table 1. The table shows the 1976 data on which the specification was based and a gradual increase in the average density, a gradual reduction in standard deviation, and resulting increases in pay factors for the initial years the specification was implements (1978 to 1983).

Another group of studies involving extraction measurements of asphalt content have shown a similar decrease in variability over time.[6] The variability of other properties does not seem to have changed, and few variabilities, if any, have been found to increase. Where the variability has been found to decrease, "it is likely that the decrease in variability of processes can be attributed to one or more of the following: contractor quality control, specifications that require a measurement of variability, improved industrial technology (e.g., computer driven plants), and improved test methods."[6]

But, the question remains, does a more consistent product always mean better performance? The answer may be no. If an inappropriate design parameter is used, or if the most important design parameter is completely omitted or ignored, the more consistent product may still perform poorly. Some also consider that properties requiring only a lower specification limit may benefit by having a higher variability.[7]

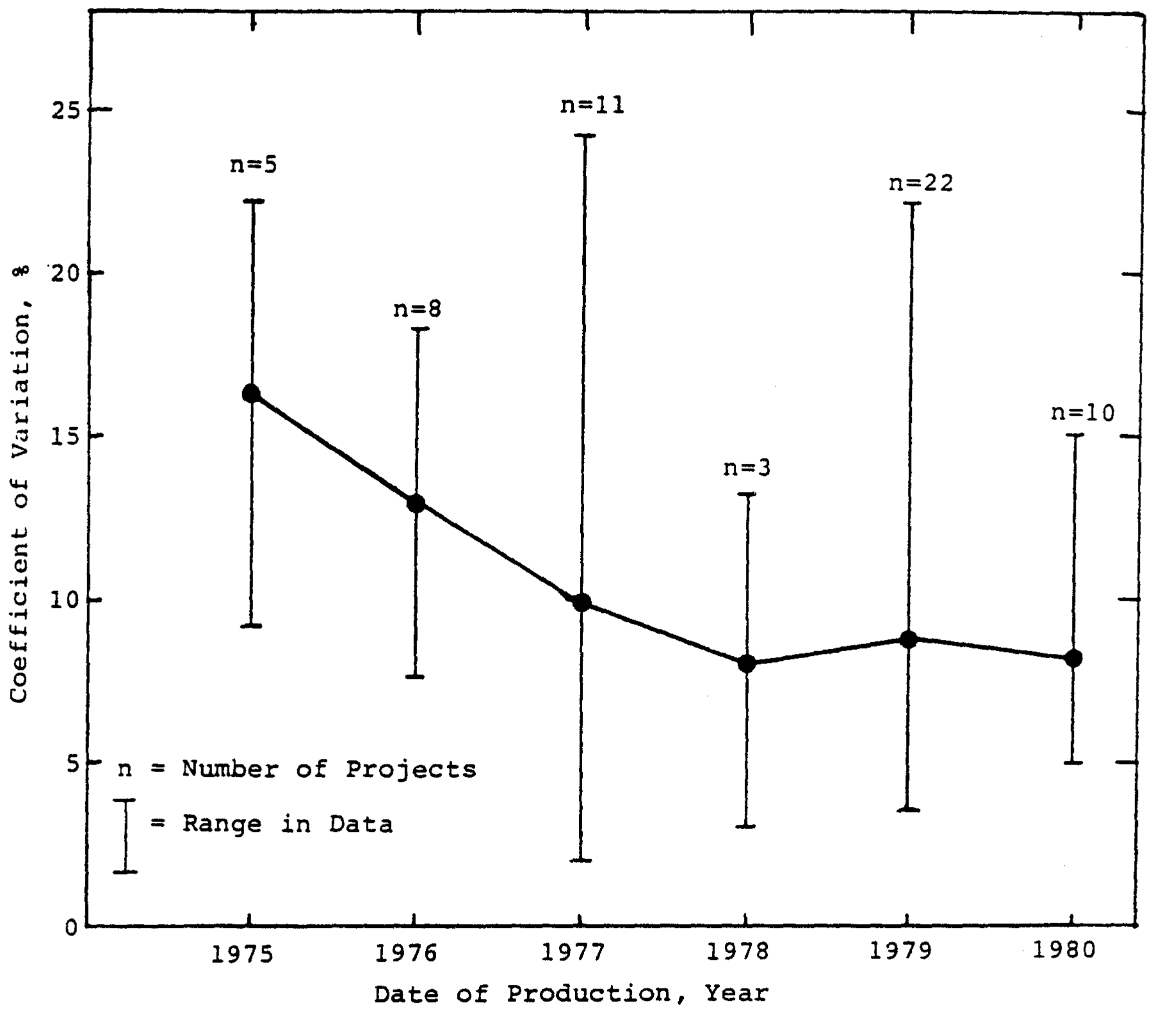


Figure 1. Decrease in Coefficient of Variation for Percent Passing the 4.75-mm Sieve[4]

Table 1. Annual Average HMAC Densities in Virginia, % MTSG[5]

	1976	1978	1979	1980	1981	1982	1983
Average Density (%)	91.3	91.6	92.0	92.6	92.7	93.1	93.1
Standard Deviation (%)	1.3	1.6	1.3	1.3	1.2	1.2	1.1
Pay Factor		97.3	98.9	98.9	99.7	100.4	100.4

To date, however, there is no clear consensus among SHA's on how to answer the following three fundamental questions that are a key to any quality assurance program:

1. What do we want?
2. How do we order it?
3. How do we know we got what we ordered?

These are the same questions that have been addressed previously, most recently in a major National Cooperative Highway Research Program (NCHRP) study.[8] Considerable uncertainty also is encountered when addressing the even more basic question: "What do we have?" To answer these questions there must be a direct tie between the design and control parameters and mixture or pavement performance. As stated in the *AASHTO Implementation Manual for Quality Assurance*, "...minor incremental improvements in the quality and durability of individual highway projects can translate into large system-wide improvements through increased performance."[2] Clearly, a need exists for a better fundamental understanding and measurement of quality in highway materials and construction.

This research effort examined the quality of highway construction in an attempt to advance the industry's ability to define, specify, and measure quality, which should lead to cost-effective improvements in construction quality.

This study consisted of testing six projects, three HMAC and three portland cement concrete (PCC). The primary objectives were to:

1. Determine how current quality control test results vary in construction projects.

2. Assess the suitability of current methods of quantifying materials and construction quality and quality variability.
3. Develop improved methods that minimize current shortcomings.

CHAPTER 2. WORKING PLAN FOR CONDUCTING RESEARCH

The detailed work plan for this project included seven tasks. Each of these tasks is discussed briefly below.

WORKING PLAN

Task A - Conduct Literature Survey

Task A began with a coordination meeting among the project staff to review and evaluate all available documents and projects ongoing in this subject area, including information that has been previously reported in the literature. The literature survey concentrated on the following eight issues:

- The level of quality in today's construction.
- Which quality measurements can be assumed to follow a normal distribution.
- The presence and quantification of longitudinal, transverse, and vertical variability in construction.
- The effect, on acceptance decisions, of erroneously assuming the existence of normally distributed measurements.
- How lot sizes are selected to ensure measurements are normally distributed.
- What acceptance sample sizes are large enough to provide sufficiently accurate measures of quality.
- The effect of variability on pavement performance.
- Interrelationships among quality characteristics.

Each of the eight areas were investigated through a Transportation Research Information Service (TRIS) literature search.

Task B - Select Projects

Task B had three subtasks: (1) to coordinate with one or more SHA's to select six construction projects to undergo increased testing, (2) to develop a detailed experimental plan for the sampling and testing to be done on those selected projects, and (3) to prepare a report documenting the detailed work plan.

B.1. Coordination and Selection of Six Projects

Three rigid construction projects [one each in Illinois (IL), Minnesota (MN), and Ohio (OH)] and three flexible construction projects [two in Oklahoma (OK1 and OK2) and one in Louisiana (LA)] were used to collect sufficient data to accurately identify the existing population of the measurements of interest. These were projects under construction (either overlays and/or new construction). The projects selected were those in which the contractor had an active quality control program in place. The following summarizes some of the items that were considered in the selection of these projects.

- Projects located in different States in order to include a variety of specifications and contractors.
- Projects with at least two 1.6-km sections (of uniform cross-section).
- Projects scheduled for completion within the 1995 construction season.
- Minimal number of pavement layers, preferably a single layer.
- Paving contractor with an active quality control program in place and willing to share the process control data.
- SHA's that specify and/or use the quality characteristics of interest in their quality control and acceptance program.
- Laboratory and field sampling and testing programs performed in accordance with AASHTO standards.
- Availability of structural design and mixture design details for review.

B.2. Develop an Experimental Plan

The experiment design defined the sample size, lot size, number of tests, and other variables needed to clearly identify the components of variance needed to develop a quality assurance plan.

B.3. Prepare Report Documenting Detailed Experimental Plan

A report was prepared documenting the detailed experimental plan. Copies of the experimental plan were submitted to the FHWA for approval.

Task C - Conduct Testing

Each of the SHA's and paving contractors were contacted to schedule all field work and sampling activities. At the beginning of each project, all field sampling and testing requirements were discussed with the SHA and contractors. The testing within this task was divided into three parts: on-site nondestructive tests, on-site laboratory tests, and off-site laboratory tests. During the paving operations, bulk samples of the paving mixture components were taken. The specific quality characteristics that were measured are discussed under Task D.

Nondestructive tests consisted of deflections measured with a Dynatest falling weight deflectometer (FWD) in accordance with the Strategic Highway Research Program (SHRP) requirements, thicknesses measured or estimated with ground penetrating radar (GPR), and densities measured with a nuclear density gauge. It was intended that smoothness data be obtained for each project by the SHA; however, such data were obtained on only one PCC project.

All the data results are included in volume II of the final report.

Task D - Perform Data Analysis

Task D had three subtasks: (1) analyze collected data; (2) assess the implication of analyses on the definition, specification, and measurement of materials and construction quality; and (3) investigate and perform other analyses to make recommendations.

D.1. Analyze the Data Collected in Tasks A & C

This subtask concentrated on the identification of those quality characteristic measurements that were believed to be normally distributed and the types of distributions followed by the other measurements. It was realized that the identification of normally distributed characteristics would take some judgment. Many natural and manufactured items are approximately normal; perfect symmetry in a distribution is highly unusual. The question that had to be answered in this task is "How normal is the distribution?" or, conversely "How much bias or skewness can be tolerated before declaring the distribution non-normal?"

Measurements made under task C were analyzed to evaluate the existence of normal distributions; to determine the population mean (μ) and standard deviation (σ) of each quality characteristic for each project; to estimate the number of samples necessary to estimate the properties of a sublot, lot, or project; and to determine which properties were interrelated.

From the controlled experiments, the following quality characteristics were measured:

- Portland Cement Concrete
  - Air content (plastic and hardened)
  - Strength (compressive and flexural)
  - Thickness (GPR and core)
  - Density
  - Falling weight deflectometer
  - Smoothness (only on Ohio)
- Asphalt Concrete
  - Asphalt content
  - Gradation of aggregate
  - Thickness (GPR on all projects and cores on OK1 only)
  - Density (nuclear on all projects and air voids on OK1 only)
  - Falling weight deflectometer

D.2. Assess the Implication of D.1 Findings on the Definition, Specification, and Measurement of Materials and Construction Quality

An attempt was made to assess the effects of applying conventional statistical acceptance procedures to populations that are not normally distributed. The statistical analyses allowed estimates of variability and identification of excessive variability.

The variabilities found in this research were compared with those reported elsewhere to provide a measure of reasonableness. While this comparison cannot be directly related to performance, it does provide an indication as to whether this variability is typical of other work or is larger or smaller.

D.3. Perform Other Necessary Analyses in Order to Make Recommendations

Conceptually, methods were considered for measuring quality for comparative analyses purposes, e.g., comparisons of quality among projects, contractors, States, etc. The literature indicated that conformal index (CI) values for the majority of quality characteristics are normally distributed.[9,10] CI values equal to zero meet the target value. CI values that are large indicate that the target was not met and/or that the variability is high.

Since CI values are normalized to a given target, direct comparison may be made among data for contractors, projects, or States for that quality characteristic. This procedure may be used for one-sided or two-sided specification acceptance.

The CI, like the standard deviation, is a measure of variation. However, the standard deviation is the root mean square of differences from the arithmetic average, or central value, while the CI is the root mean square of the differences from a target such as the job mix formula (JMF) value. In other words, the standard deviation is a measure of precision, while the CI is a measure of exactness (accuracy) or degree of conformance with the target.

In equation form:

$$\sigma=\sqrt{\frac{\Sigma(x-\bar{x})^2}{(n-1)}} \qquad CI=\sqrt{\frac{\Sigma(x-T)^2}{n}} \qquad \textbf{(1,2)}$$

where:

σ = standard deviation

x = each point in the distribution

$\bar{x}$ = mean value

n = sample size

CI = conformal index

T = target

The value "T" in the CI equation refers to the target value (JMF, design thickness, design density, etc.). The CI statistic may be used directly with both percent within limits/percent defective and the loss function approaches. The attractiveness of this statistic is that it focuses in on the target value and it is this target value that is defining the quality level.

Other analyses that were performed were:

- correlations among the various quality characteristics, within projects, to determine which, if any, were interrelated,
- comparisons between standard deviations found in the study and those reported in various references,
- comparisons between State results and those obtained by the study team when possible,
- precursory analysis of a proposed procedure to allow the accumulation of test results so as to increase the sample size of a lot, and
- comparisons between measurement procedures to identify similarities.

CHAPTER 3. LITERATURE SEARCH

INTRODUCTION

A literature survey was conducted to investigate measurement and specification of construction quality, including levels of variability in pavement construction and the effect that this variability has on the performance of the pavement. Records were found using the TRIS and EI Compendex® services. The intent of the literature survey was to identify information related to the following eight topics of interest:

- The level of quality in today's construction.
- Which quality measurements can be assumed to follow a normal distribution.
- The presence and quantification of longitudinal, transverse, and vertical variability in construction.
- The effect, on acceptance decisions, of erroneously assuming the existence of normally distributed measurements.
- How lot sizes are selected to ensure measurements are normally distributed.
- What acceptance sample sizes are large enough to provide sufficiently accurate measures of quality.
- The effect of variability on pavement performance.
- Interrelationships among quality characteristics.

More literature was found on some of these topics than on others. Because the literature combined several of these topics, the subject headings in this chapter differ from those above.

The emphasis on achieving high quality during construction is based on the belief that this will lead to increased pavement life and lower life-cycle costs during the pavement life. Although the ability of a paving contractor to place a quality product is an important part of initial quality, a quality design of the pavement by the responsible agency is of equal importance. It is critical for an agency to understand the impact that variation in material and construction properties (e.g., strength, air content, asphalt content, density, and smoothness) will have on future distress manifestations. To establish the optimum quality level in the construction of pavements, one of the necessary ingredients is for the responsible agency to understand the relation between variability and pavement performance. This relationship can then be used to develop and measure a desired level of quality. To address these issues a quality measurement system is a necessity.

LEVEL OF QUALITY IN TODAY'S CONSTRUCTION

Traditionally, quality in pavement construction was considered the ability of a contractor to adhere to method specifications developed by the highway agency. These method specifications provided requirements for material proportioning, mixing limits, and the procedures to follow for a job to be acceptable.(11) Although these specifications allowed for control of certain material and construction characteristics, the overall quality of the pavement and its effect on performance were not directly addressed.

More recently, such statistical measures as percent within limits (PWL), its complement percent defective (PD), and the average absolute deviation from the mean have been used as measures of quality level. Associated with measures of quality level is another item typically included in specifications — pay factors. For quality levels below a standard, a reduced pay factor is computed; for quality levels at the standard, the pay factor should be 1.00 (i.e., the unit bid price); and for quality levels above the standard, increased pay factors are becoming more prevalent.(6)

State Specifications

The following section details a sampling of the current state of pavement specifications used by SHA's across the country. In particular, emphasis has been placed on those agencies that are using statistical quality assurance specifications in an effort to improve the quality of their pavements. Quality assurance (QA) has two primary ingredients — quality control and acceptance.(3) Quality control (QC), often referred to as process control, is the control of the processes as the material is being mixed, placed, compacted or consolidated, finished, cured, etc.

Asphalt Concrete

In asphalt concrete specifications used by many States, including New Hampshire, California, and South Dakota (and perhaps as many as 30 others), quality control is in the hands of the contractor.(12,13,14) Some State specifications require that a contractor provide a QC plan to the State detailing the procedures that the contractor will follow during construction.(12,13,14) Other States simply require the contractor to have an approved plan.

Acceptance, sometimes referred to as quality acceptance, is the testing that is performed to determine whether the contractor's QC program has kept the processes within acceptable control limits, usually around the job mix formula or other target value. Acceptance testing is usually done by the State. For instance, Texas requires contractors to test properties that are not used to determine pay and State personnel to test all properties that are used to determine pay. But with the downsizing that is taking place within many SHA's, the acceptance testing is sometimes being assigned to the contractor, under prescribed conditions. Acceptance for HMAC has traditionally included asphalt content, percent passing critical sieves such as the 2.36-mm and the 0.075-mm sieves, and percent compaction of the in-place mat on the basis of the maximum theoretical specific gravity (MTSG).(15) A review of several State specifications illustrated that these items were not always included and in some cases additional items were included. New Hampshire's quality acceptance plan includes testing the following on a random basis: gradation, asphalt cement content, percent of MTSG, air voids in total mix, viscosity, thickness, and ride smoothness.(16) California tests compaction, stability, abrasion smoothness, gradation, and asphalt content for acceptance.(17) Many SHA's and the FHWA are introducing volumetric properties into acceptance plans in lieu of gradation because the volumetric properties are viewed as being more related to performance than gradation.

AASHTO's Quality Assurance Guide Specification includes gradation and asphalt content, and allows for acceptance criteria based on Hveem or Marshall Design test results.(1)

Portland Cement Concrete

New Jersey The State of New Jersey, which has been one of the leading developers of performance-related specifications, uses the AASHTO Design Procedures as the basis for its acceptance procedure. Conceptually, the method consists of using as-built measurements obtained in the field and working backwards through the design equation to determine the pavement's actual load-carrying capacity. By comparing this to the loads computed from the design values, it is possible to estimate the degree to which deficient quality will shorten the pavement's service life. Engineering economics procedures are then used to determine the change in present-worth value, which is the basis for the pay adjustment. In essence, this is the equivalent of a liquidated-damages clause.

If the as-built measurements taken in the field exceed the design values, this results in a surplus load-carrying capacity that extends the pavement's service life. Accordingly, the New Jersey Department of Transportation (NJDOT) pay schedule has been designed to pay slightly greater than 100 percent (currently a maximum of 102 percent) for truly superior quality.

A vital step in the development of the specification is the construction of operating characteristic (OC) curves to confirm that the acceptance procedure and pay schedule will perform as intended. This ensures that contractors who deliver the acceptable quality level (AQL) will receive 100 percent payment on average. This is made possible in part by the inclusion of a bonus provision that allows errors in quality estimates (and the corresponding errors in pay factors) to balance out in a natural way without biasing the average pay factor downward at the AQL. At lower levels of quality, this analysis allows agency engineers to confirm that sufficient payment is being withheld to cover the anticipated cost of future repairs.

The NJDOT uses PD as the statistical measure of quality, which provides an effective way to account for both mean level and variability. The AQL is defined as a percent defective level of PD = 10 below the class design strength for most classes of concrete. A rejectable quality level (RQL) is also defined but is based on the percentage of material falling below a lower limit that corresponds approximately to the structural design strength (f_c').

The PD is obtained from an appropriate table after first computing the quality index (Q) as follows:

$$Q = \frac{\bar{x} - L}{S} \quad \textbf{(3)}$$

where:

Q = quality index

$\bar{x}$ = sample mean

S = sample standard deviation

L = lower limit

The NJDOT has also used a composite pay equation that accepts PCC pavement on the basis of three independent quality measures: strength, thickness, and smoothness. Pavement smoothness was made the dominant factor in the equation because excess strength and thickness would be perceived as being of little value if the pavement did not provide a smooth ride. Another unique feature of this specification is that, within reasonable limits, it allows surpluses and deficiencies in strength and thickness to offset one another, consistent with the AASHTO Design Procedure.

This specification produced excellent results on the one trial project for which it was used. However, NJDOT engineers noted that two changes would be contemplated for future applications. Because the specification defined a lot as a day's production, and the contractor elected to construct the project in a piecemeal fashion, the resulting sampling and testing effort was felt to be excessive. Accordingly, a somewhat different lot definition will be considered. Also, for the initial application, the pay equation was based on the mean values of the three quality measures. For any future applications, the acceptance procedure will be structured around PD as the quality parameter.

<u>Oklahoma</u> The Oklahoma DOT PCC pavement specification uses five measures of pavement quality: gradation, air content, strength, thickness, and smoothness. Testing of PCC pavement materials is done on a lot-by-lot basis, with a lot defined as 8,360 m^2. Samples are taken and

tested for gradation, air content, strength, and thickness on the basis of sublots. Smoothness is measured on the basis of 160-m increments throughout the paving section.

Adjustments to the pay factor are based on the difference between the test results and the specified design values. For air content and gradation, the absolute difference is considered, so that actual values above and below both decrease the pay factor. For strength, thickness, and smoothness, the sign of the difference is considered, though no incentives are provided for high strengths or thicknesses. Roughness values less than the maximum allowable are used to increase the pay factor and provide an incentive for the contractor to place the smoothest pavement possible.

South Dakota As with many other SHA's, the South Dakota DOT bases the pay factor for concrete paving operations on the smoothness of the PCC pavement once it has been placed. A special provision to the South Dakota DOT PCC pavement specifications provides for payment incentives for profile indices less than the specified value.[18] There is no reduced pay factor for an excessive profile index; rather, the contractor must grind the existing pavement or remove and replace the pavement until an acceptable profile is achieved.

Illinois The Illinois DOT QA specification contains five quality control characteristics to be monitored: strength (compressive or flexural), air content, slump, smoothness, and thickness. Of these characteristics, only thickness and smoothness are used for adjusting the pay factor. No incentives are given for pavement thickness greater than specified, though pay factors range from 80 percent of bid price for thicknesses 2 to 3 percent less than designed to 50 percent of bid prices for thickness 7.5 to 10 percent less than designed. Pay adjustments for smoothness range from 103 percent of bid price for a profile index less than 35.5 mm/km to 90 percent of bid price for a profile index between 221 and 237 mm/km.

Michigan The Michigan DOT PCC specification calls for testing air content, slump, concrete temperature, and compressive strength.[19] For determining the pay factors, the PD is calculated using the mean and standard deviation of the strength. Design strengths and sampling rates are determined by the class of concrete being placed.

West Virginia The West Virginia DOT PCC specification includes compressive strength, cement factor, water content, coarse aggregate size, and entrained air.[20] Pay adjustment factors are calculated for compressive strength, slab thickness, and pavement smoothness. For each of these quality characteristics, the pay factors are defined as reductions in the bid price for failing to meet the specified quality levels. For smoothness and thickness, the mean values are used to calculate the pay factors. For strength, the mean value and standard deviation of the strength measurements, along with the design stress, are used to calculate the percent reduction in bid price.

Washington The Washington DOT PCC specification includes requirements for temperature, air content, slump, compressive strength, smoothness, and thickness.[21] The mean and standard deviation of the compressive strength are used to calculate pay adjustment factor on a lot-by-lot basis. Air content and slump also are used to reduce the pay factor for excessive deviations from the design values.

Indiana The Indiana DOT PCC specification includes requirements for cement content, water-cement ratio, cement-fly ash ratio, air content, flexural strength, coarse aggregate size, and fine aggregate amount.[22] Pay adjustment factors are determined for the slab thickness, with reductions determined as a ratio of the in-place thickness squared to the designed thickness squared. This adjustment is only for in-place thicknesses less than designed; no pay increases can result from thicker pavements. Air content and flexural strength also are used to determine reductions in pay factors for deviations from the designed parameters. Pavement smoothness does provide for pay increases for very smooth pavements, though no reductions are specified for rough pavements.

Minnesota The Minnesota DOT PCC specification provides for payment incentives based on the mean flexural strength, standard deviation of the flexural strength, and the number of tests used to calculate those two values. As more tests are performed, the reduction in calculated standard deviation is offset by increases in the k factor in the agency's quality index (QI) calculation:

$$QI = \bar{x} - (S \times k) \tag{4}$$

where:

- QI = quality index, MPa
- $\bar{x}$ = mean compressive strength value, MPa
- S = standard deviation of compressive strength, MPa
- k = factor determined by number of tests used to calculate mean and standard deviation

AASHTO As mentioned previously, the AASHTO Subcommittee on Construction prepared a specification for SHA's to use for QA of PCC pavements.[1] This specification considers compressive strength, thickness, and surface smoothness as the key quality parameters. In this specification, the PWL is estimated based on the mean, standard deviation, specification limit(s), and number of tests. It also is specified that the average of any five consecutive compressive strength tests shall be greater than the design strength. Pay factors are calculated by a pay factor equation based on the estimated PWL; the equation allows for an incentive as well as a disincentive.

This limited survey of the current status of the implementation of quality assurance specifications indicates that most of these specifications at least attempt to control variability, however, much improvement is still possible.

Construction Quality

Good quality construction does not "just happen," as evidenced by a "worse-case scenario" study reported by the Waterways Experiment Station on the examination of an airport pavement in Egypt.[23] An asphalt concrete overlay was constructed on an airfield parking apron and taxiway. The overlay exhibited significant deformation and depressions under normal aircraft traffic. Cores taken from the apron showed that the aggregates in the mix were gap-graded and the mix had an extremely high natural sand content. In general, the mixture was inconsistent and did not meet specifications. The asphalt cement used in the mix also was found to be unstable. An examination of construction operations showed that the stockpiles at the quarry contained uncrushed particles. Stockpiles at the plant were contaminated with fines and the cold feed bins were filled to overflowing. The absence of adequate quality control was a major factor in the poor performance.

Asphalt Concrete Pavement

Significant variations in either the transverse, longitudinal, or vertical direction can be the cause of problems. The problems will show up in the form of differential pavement distress along the length of a construction project. The causes of significant variation can be found all along the construction process. Several of these potential causes are discussed below.

The asphalt plant is one source of variation. "The key to quality mix starts at the cold-feed bin stage."[24] Many batch plant operators believe that carryover from one cold feed bin to another will be corrected at the gradation screens. This is not the case, especially if the plant's gradation screens are not working properly. It is necessary that the proper proportions are taken from the cold feed bins to ensure that a steady supply of material is fed to the hot bins. Some batch plants have operated with fewer than four screens but use all four hot bins. The four screens provide for a mechanism to reject oversized material and, if less than four are used, that mechanism is taken away. Clogged and damaged screens can cause carryover of one size of material to the next bin. This carryover will cause improper proportioning of material into the mix and could cause an overflow of one bin into another.

Drum mix plants have their own set of problems. Again, the key to a quality mix starts at the cold feed bin stage.[24] Proper stockpiling will ensure that uniform material is fed into the plant.[25] Overflow of the cold feed bins in a drum mix plant will cause a problem. These bins determine the aggregate proportioning, and if overflow occurs, the resulting mix will have an improper gradation.

The most likely area for generating segregation is the surge or storage bins.[25] One way to minimize the amount of segregation in loading the surge bin is to use a bin-loading batcher. The material can be dropped from the batcher into the chute. However, it may cause the material to pile up in the middle of the bin. Once the material has created a pile, new material will roll down the sides with the larger material traveling farther than the small material and causing segregation. If a rotating chute is used to drop the material, it must rotate and allow the material to turn directly downward.[25] If the material drops from a rotating chute that does not rotate properly, the material will pile up at the bottom of the bin and, as material is dropped, the larger material will have a tendency to roll down the pile and cause segregation. When the material is loaded into a truck, it should be allowed to drop in one mass rather than dribbling from the bin into the truck. The truck should be loaded in three increments: the first drop at the front of the truck, the second adjacent to the tailgate, and the third in the center. This will minimize, to some extent, the large aggregate rolling, and the material that has rolled will be covered by other material.

The paver can be another source of segregation if it is not operated correctly. The hopper should not be emptied between each truck load; leaving material in the hopper forces the coarse material to be fed more uniformly into the drag flights.[25] Since coarse material has a tendency to collect on the wings of the hopper, the wings should never be dumped. Paver operations should be conducted as continuously as possible. However, if a truck is late, the paver should be stopped before it runs out of material.

PCC Pavements

One problem with quantifying construction quality is the difficulty in generating consistent, repeatable measurements of quality characteristics. For example, when a testing laboratory is not able to achieve within-test COV's below 5 percent for cylinder compressive strength, it is generally indicative of a quality control problem at the laboratory.[26] In most of the SHA specifications, the mean values are used to determine whether a quality material is being placed, with no consideration of the variability inherent in the mean value. The risk to the agency of erroneously accepting a mean compressive strength of 30 MPa is much higher when the associated standard deviation is 5 MPa than when the standard deviation is 2 MPa.

Another problem that has been the subject of many reports is the use of profile measurements to determine pay factors on the basis of the final smoothness of the pavement without considering all of the factors that influence the final values. In South Dakota, differences in the way that a computerized, automated profilograph measured smoothness led to discrepancies in the amount of positive pay factor due to a contractor.(18) In this instance, the filtering process that was intended to remove spikes from the profile also attenuated longer features that impact the ride quality. The profile values generated by the contractor's device indicated a pay incentive roughly double the amount that would be due based on the agency's profile value (generated with a manual profile device).

In Wyoming, a study was conducted to determine the impact of operator influence and environmental changes on the measurement of smoothness using a South Dakota road profiler.(27) Findings of the study indicate that: (1) the variability of rut depth measurements was much higher than for the longitudinal smoothness, (2) the majority of statistically significant differences in smoothness occurred on smooth pavements, and (3) changes in environmental conditions resulted in significant changes in smoothness measurements. These findings underscore the need for having a good understanding of the factors affecting each particular performance characteristic being measured to ensure that pay incentives and disincentives are applied rationally.

A similar study in Arizona, using mechanical and automated California profilographs, indicated that operator interpretation of the profile trace data can impact the final results by as much as 16 mm/km.(28) Comparisons of the different pieces of equipment showed significantly different smoothness readings at a 1-percent significance level. Comparisons of the different operators showed significantly different smoothness readings at a 7-percent significance level, indicating that operator interpretation has a larger impact on the variability of profile values than the type of device used. Another interesting aspect of this study was the use of different filter settings on the automated devices, which produced significantly different results at a 1-percent significance level, the same level as using different machines.

One of the other problems with the monitoring of quality in PCC pavement construction is the lack of correlation among initial material characteristics (initial quality) and future performance of the pavement (cost due to differences in performances). The PAVESPEC computer program developed under a previous FHWA project runs simulations of future pavement performance on the basis of as-constructed material properties, calculates anticipated quantities of distress, and then calculates a present worth cost of the pavement (including maintenance and rehabilitation) over the analysis period.[29] The comparison of the as-constructed present worth to the as-design present worth is then used to calculate the pay factors.

QUALITY CHARACTERISTIC DISTRIBUTIONS

Assumption of Normal Distribution

The use of the normal or Gaussian distribution is often made when applying QA specifications. The use of this distribution simplifies what could otherwise be an arduous task of trying to define populations. Defining a normal distribution requires only an estimate of the average and standard deviation. Two of the important properties of the normal distributions are that it is unimodal, i.e., has one peak, and it is symmetrical. In practice, few populations are truly normal but many are approximately normal. This raises the question of how far a population can be from truly normal and not create large errors in estimates of the population.

This question of the assumption of whether material and construction populations follow a normal distribution has been in engineers' minds since at least the early 1960's when the analysis of the data was obtained from construction of the AASHO Road Test.[30] As one example of this concern, Shook concluded that the density of the asphalt concrete pavement from the Road Test "...were distributed approximately normal."[31]

Asphalt Concrete Pavements

The information from the AASHO Road Test data analysis led to an appreciable research effort in the 1960's by the Bureau of Public Roads (BPR) and several SHA's.[6] Several of the States, including Louisiana, Pennsylvania, Virginia, and West Virginia began to gather data that triggered the evolution to what are now QA specifications. Some of these States did look at the applicability of normality in highway construction. One Louisiana report indicated that bitumen content, percent passing 13-mm, 9.5-mm, 2-mm, 0.180-mm, and 0.075-mm sieves, tended to be normal.[32] Virginia did a comparison between the number of predicted samples for several highway properties with an actual count of samples within one, two, and three standard deviations.[33] The results shown in table 2 indicate approximate normality by the agreement between actual and predicted number of samples for both construction and nonconstruction properties.[34] More recently, Schmitt, et al., in a study of asphalt density tests in Wisconsin, were able to verify "... that pavement density is normally distributed."[34]

In addition to the studies mentioned above, a Pennsylvania DOT report found that most of the data elements are skewed to the right.[35] However, if the actual distribution of these elements is skewed, it may not be necessary to throw out all of the traditional methods shown in QA specs. As noted by Shapiro, "the concern in testing for a distributional assumption should be whether or not it is reasonable to approximate the data with the model, not whether the data came from the hypothesized distribution."[36] Furthermore, if the distribution is found to be non-normal, the data often can be transformed using the logarithm or some other mathematical function to fit a normal distribution. The data also can be combined into average results and analyzed, a step that generally improves normality. Therefore, in most cases, if care is taken to examine the distribution of data before making a decision, it will not cause significant errors if the data are assumed to be normally distributed. However, quantification of these errors, or bias, was the impetus for the computer program SKEWBIAS, developed in this study.

Table 2. Predicted vs. Actual Samples Within 1, 2, 3 σ Limits[33]

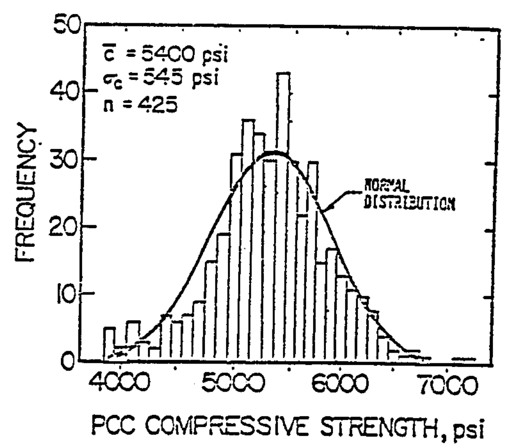
Data	X	σ	Limits	Predicted No. of Samples	Actual No. of Samples
Core Strength, psi	3,772	565	$\bar{x} \pm 1\sigma$	124	134
			$\bar{x} \pm 2\sigma$	176	179
			$\bar{x} \pm 3\sigma$	186	185
Core Depth, in	9.73	0.32	$\bar{x} \pm 1\sigma$	124	135
			$\bar{x} \pm 2\sigma$	176	179
			$\bar{x} \pm 3\sigma$	186	183
Air Content, PCC, %	4.1	1.09	$\bar{x} \pm 1\sigma$	23	23
			$\bar{x} \pm 2\sigma$	32	33
			$\bar{x} \pm 3\sigma$	34	33
Asphalt Content, %	4.68	0.17	$\bar{x} \pm 1\sigma$	40	39
			$\bar{x} \pm 2\sigma$	57	58
			$\bar{x} \pm 3\sigma$	60	60
Vehicle Speed, mi/h	56.5	5.7	$\bar{x} \pm 1\sigma$	439	427
			$\bar{x} \pm 2\sigma$	626	634
			$\bar{x} \pm 3\sigma$	658	659

1psi = 6.894 kPa   1 mi/h = 1.6 km/h
1 in = 25.4 mm

PCC Pavements

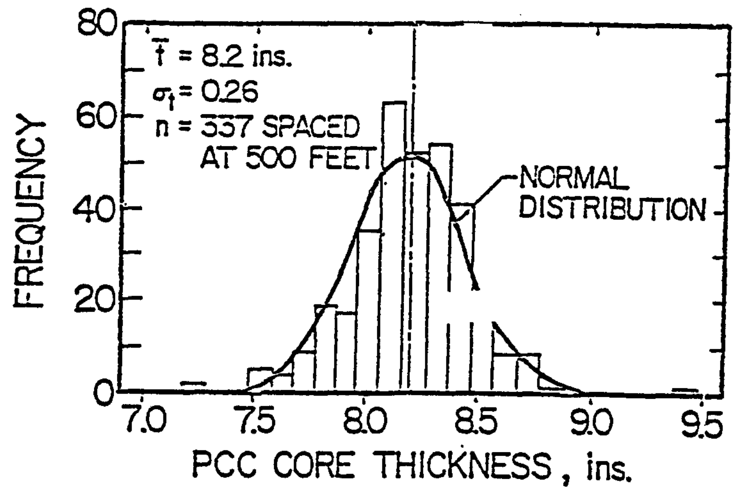
A 1994 FHWA process review of several PCC pavements in Delaware indicated that two out of six projects had thickness measurements that were skewed to the right, supporting the Pennsylvania study mentioned above. But four were only very slightly skewed. It is surmised that, on the projects with the skewed data, the contractor was constantly monitoring the thickness, to stay as close to the design as possible, and changing his operation often to avoid placing an excessive amount of PCC or falling into a price adjustment range. This may indicate good QC, but it also can be construed as not having a good idea of the process capability and thus overreacting. The constant change produces a multimodal distribution, *unless the data are* *subdivided into lots of constant operation and analyzed in that manner.* On the other hand, the four projects that had only slightly skewed data were likely obtained from operations that were allowed to continue with minor or few changes.

Over the past 40 years, there have been several other studies that presented data showing that the thickness, flexural strength, and compressive strength in most cases are approximately normally distributed.[37-41] A long-term study conducted by the Portland Cement Association reported normalized distributions for compressive strength, although the shapes of the distributions were dependent on the type of cement, type of curing, and the age at which strength measurements were made.[42] Distributions were provided for concrete samples ranging in age from 3 days to over 20 years. Figures 2 through 8 show typical distributions of various concrete parameters, with most showing close to normal distributions.



1 psi = 6.894 kPa

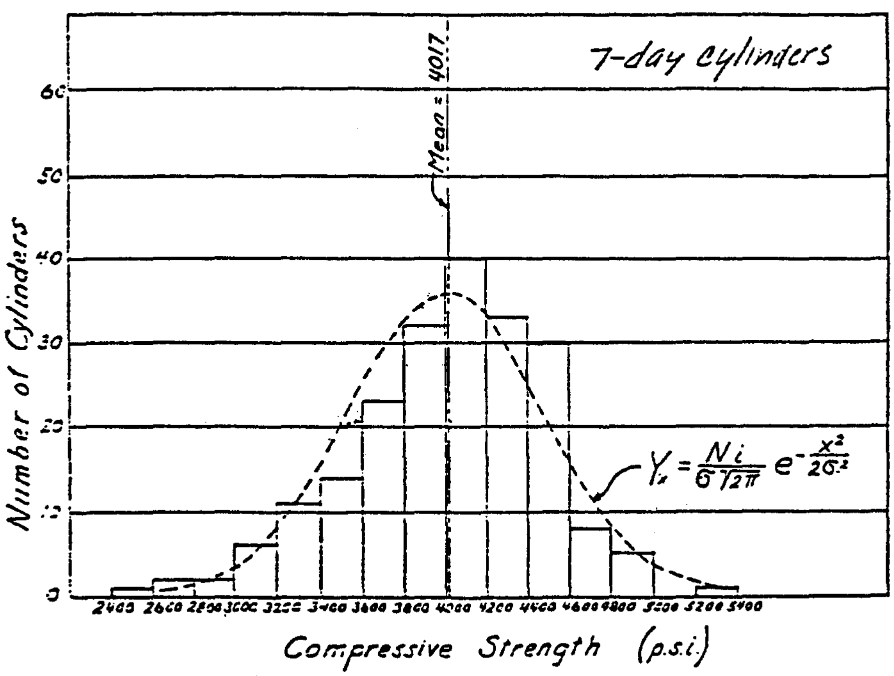
Figure 2. Distribution of PCC Compressive Strength[43]



1 in = 25.4 mm

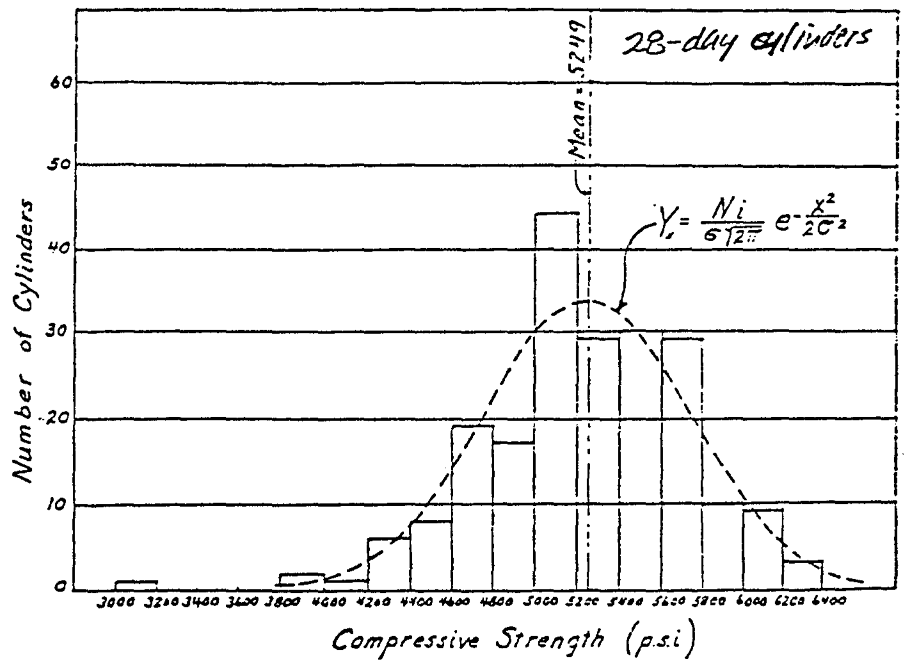
1 ft = 0.305 m

Figure 3. Distribution of PCC Core Thickness[43]



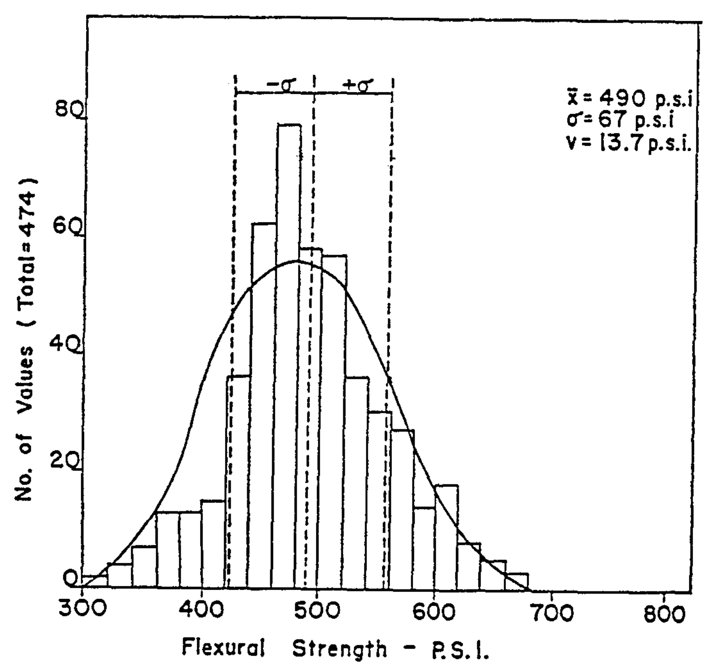
1 psi = 6.894 kPa

Figure 4. Histogram of 7-day Concrete Compressive Strength[38]



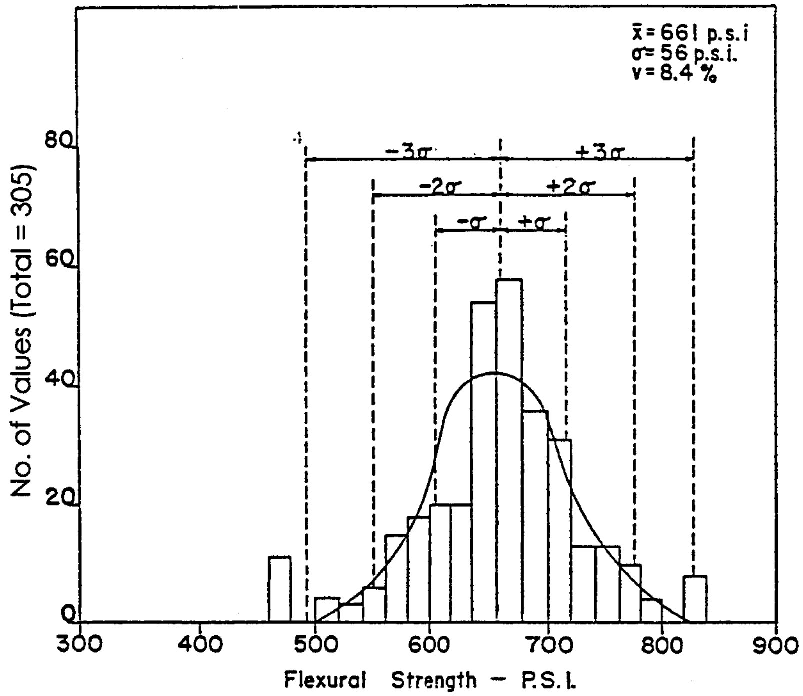
1 psi = 6.894 kPa

Figure 5. Histogram of 28-day Compressive Strength[38]



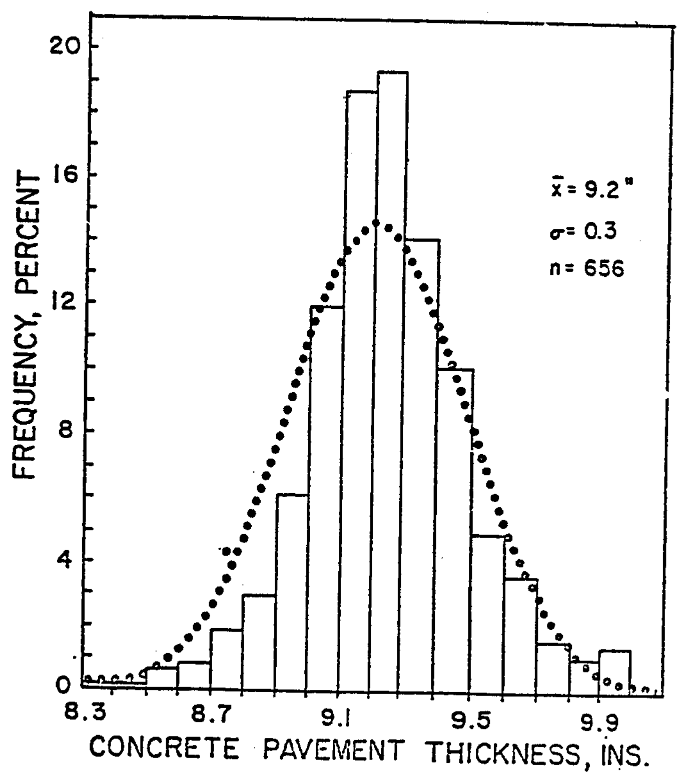
1 psi = 6.894 kPa

Figure 6. Histogram of Concrete Flexural Strength[40]



1 psi = 6.894 kPa

Figure 7. Histogram of Concrete Flexural Strength[40]



1 in = 25.4 mm
1 ft = 0.305 m

Figure 8. Histogram of Concrete Slab Thickness[41]

Another recent study collected data from 100 different concrete projects around the country and compared each set of data against normal and log-normal distributions.[44] The study compared the estimated number of low tests based on the normal and log-normal distributions [using the mean, standard deviation, and coefficient of variation (COV)] with the actual low test results for the particular data set. The level of acceptance also was varied for each comparison, with from 3 out of every 10 tests being low (t = -0.52) to 1 test in 741 being low (t = -3.00). The t-values were used to determine the limit of low values, as shown in the following equation:

$$L = \bar{x} + (t \times S) \quad \textbf{(5)}$$

where:

- L = limit of acceptable strength values, MPa
- $\bar{x}$ = mean value of the data set, MPa
- t = Student's t-value
- S = standard deviation of the data set, MPa

The findings of this study indicated that the log-normal distribution was better than the normal at predicting low test results in the range of 25 to 0.5 percent low tests (1 in 4 to 1 in 200). The normal distribution led to higher estimates of the number of low tests in the range of 2.5 to 0.5 percent low tests (1 in 40 to 1 in 200). Another aspect of the study was the use of a chi-square test to improve the prediction of low strengths. Essentially, no improvement was observed for prediction of low strengths, which is where improvements would be needed.

The importance of the distribution shape for concrete parameters is in estimating the probability of failure when comparing the strength of the material with the anticipated stress levels from traffic. Conceptually, if the loaded-axle flexural stress developed in a PCC slab also can be assumed to follow a normal distribution, a portion of the area of overlap between the stress distribution and PCC flexural strength distribution represents the probability of failure. If the true distribution of the PCC strength was skewed to the low-strength side, then the probability of failure would increase. An increase in the standard deviation for the PCC strength distribution also would increase the probability of failure. Of course, this scenario is only for discussion purposes, since cracking is ultimately caused by fatigue damage from repeated loads. On the other hand, the concept appears reasonable.

However, considering the limited amount of data available to verify normality in light of the above-mentioned explanation of the consequences of non-normality, concern often is expressed whether many highway properties are normally distributed or, stated differently, how close to normal they are. One study suggests the use of the absolute mean deviation from the target as a way to address contractor process changes that typically may occur within a lot.(45) This procedure produces a skewed distribution that may fit the data from a changing process better than the assumption of a normal distribution.

QUALITY ASSURANCE SPECIFICATIONS

Many States are moving toward QA specifications. These types of specifications give more control and responsibility to the contractor to meet a set of criteria provided by the SHA. Some of these criteria have been listed in the sections above. However, in an attempt to make the specifications more performance-related, criteria better related to performance are constantly being sought.

In the past, several States have based acceptance criteria for HMAC on asphalt content, gradation, and percent compaction in the mat on the basis of maximum specific gravity.(14,15,16) However, more recent specifications use volumetric properties in lieu of gradation in the belief that the former are better quality measures than gradation. "The move toward statistically-based quality control/quality assurance construction specifications is motivated by the desire to control the quality of the finished product," according to Parker, et al.(45) Another quality characteristic that is gaining favor in both HMAC and PCC specifications is smoothness of the final surface.(6)

Specification development related to performance has progressed further for PCC than for HMAC. An example is an FHWA project that developed a prototype performance-related specification.(29) The prototype specification was based on the best of the existing SHA specifications available at the time and incorporated the PAVESPEC program to perform life-cycle cost comparisons of as-designed and as-constructed pavement.

The PAVESPEC program takes into consideration four quality characteristics: air content, thickness of slab, strength of concrete, and surface smoothness. Target mean and standard deviation values for as-designed quality characteristics are among the input parameters, as are pavement design conditions (such as joint spacing, base type, average annual daily traffic, freezing index, etc.) and maintenance strategies and costs. The program uses test results reflecting as-constructed quality characteristics to predict expected distresses and serviceability and to estimate an overall present worth (PW) per mile of the pavement section. By comparing the as-designed to as-constructed PW costs, the pay factor is determined, with more than 100 percent paid for lower as-constructed PW costs and less than 100 percent paid for higher as-constructed costs. Both the population mean and standard deviation affect the pay factor for a lot.

Quality Assurance Risks

Both the contractor and the State have a certain amount of risk involved in the purchase of the contractor's services. This is true no matter what type of specifications are used; however, it is easier to quantify these risks when a QA specification is used. The risks involved can be viewed conceptually. The seller, or α risk, is that the contractor will build a product that actually meets the specifications but due to an inadequate estimate of the population parameters, the SHA will reject the product. The buyer, or β risk, is that the work that should have been rejected is accepted at full payment.

These risks can be minimized in several ways. The consequences to the contractor of risk associated with making wrong acceptance decisions can be minimized to some extent by using pay factors. Pay factors allow the contractor to get paid for work that does not fully meet the specifications but is not fully rejectable. Pay factors also should allow for a positive pay factor for work that not only meets specifications but also comes very close to being "perfect." Direct incentive pay or a crediting provision in the specification can balance the risk and offset unwarranted penalties.[46] As noted by Parker, "unless the adjusted pay schedule is designed to allow bonuses and reductions to balance out in a natural way, the average pay factor will be biased downward at the acceptable quality level and acceptable work may be unfairly penalized."[46] Some believe that pay adjustments should be applied to manageable size portions, or lots, rather than the entire project.[45] One reason for this belief is that, because processes are being changed often, a lot-by-lot acceptance or adjustment is more realistic than those based on an entire project.

To further quantify risks, an AQL and RQL often are defined. The contractor's risk is that material truly at or above the AQL will be rejected or receive a negative pay factor. An effective QA specification should contain AQL and RQL values that are realistic, given the available materials, equipment, and conditions.

To minimize the contractor's risk, the AQL should be set at a level that satisfies design requirements but is not so high that extraordinary construction methods or materials will be necessary. To minimize the State's risk, the RQL should be set at a level such that the option to require removal and replacement at the contractor's expense is truly justified.

Many States are apprehensive about moving to QA specifications. These anxieties include losing control over construction and facing higher bid prices. A pilot project in Colorado showed that the successful bidders had no apparent concern about the specification because their bid price was comparable to bid prices under a method specification.[47] "Under a well-designed statistical end-result specification that used a quality index or percent defective approach, there would be much more incentive for contractors to tighten the control of their operations and minimize variability," according to Benson.[48] In other words, a QA specification gives the contractor the incentive to find better ways of constructing the pavement. Method specifications have a tendency to limit the contractor's ingenuity in construction. With this incentive comes a better final product and a reduced risk to the State of having work that does not meet specifications. With the introduction of the NQI, the fears of many SHA's concerning the use of QA specifications appear to have abated.

Another way of decreasing risks is to increase the number of acceptance samples. Because of the inherent variability in testing, sampling, and construction, it cannot be expected that one sample will be representative of an entire lot, let alone an entire project. If several samples are taken and analyzed to find their average and standard deviation, the population can be more accurately estimated and the risks to both parties reduced. This is potentially the greatest advantage of large sample sizes.

In most cases, the acceptance of undesirable material for pavements will not result in an increased safety risk, as might happen if low-strength material was accepted for structural use. However, the costs associated with accepting undesirable material for pavements — added maintenance, increased lane closure, user delay, shorter time until rehabilitation — do add up and can be a substantial amount compared with the overall budgeting of the responsible agency. For this reason, it is important to reduce the risk of accepting rejectable materials.

Selection of Lot Sizes

Lot size is a very important factor in acceptance testing, and many decisions are necessary to establish the proper lot size. A lot is a subset of a construction project on which acceptance decisions are made. According to Brakey, "the material contained in the lot needs to come from a relatively continuous operation, and the quality parameters should be unimodally distributed."(47) As stated before, it is much fairer to both the contractor and the State to accept or reject material on a lot-by-lot basis.

The choice of lot size is sometimes an attempt to balance costs: testing cost on the one hand versus the cost of accepting RQL or rejecting AQL work on the other. Conceptually, small lot sizes prevent a large quantity of RQL material from being incorporated into a project and protect a contractor from having a large quantity of construction rejected or pay adjusted, but they also may result in an increase in testing costs. On the other hand, large lot sizes allow a larger sample size to be used, which provides a more accurate estimate of the population. One caution that should be recognized is that it is logical that the definition given above "...that the lot needs to come from a relatively continuous operation, and the quality parameters should be unimodally distributed..." is harder to meet as the lot size becomes larger. As previously mentioned, materials and plant and construction operations change often in highway construction. Periodic changes to meet a design criteria or JMF tend to change the population being produced. Combining populations with changes into a single estimate can create multimodal distributions and exaggerated estimates of the population variability.

The most popular choice for lot sizes are a day's production, a quantity commensurate with a day's production, or more recently, an entire project. The FHWA Direct Federal Regions and some SHA's have successfully implemented QA specifications that use an entire project as a lot. This success is at least partially due to the use of contractor QC tests with agency verification.[48] This allows the contractor to know the population parameters as the project progresses and minimizes the probability of a large price adjustment or the construction of a large quantity of below-AQL product.

For most SHA PCC specifications, the definition of a lot size is intended to reduce the chances of significantly different materials being placed and evaluated as a single unit. For this reason, the most common specified lot size is a single day's production, with the provisions for creating smaller lot sizes if the production process is interrupted by changes in the concrete class, aggregate gradation, aggregate moisture condition, mix formula, water-cement ratio, and so on.

But a day's production is not without its own problems. The major one is the indefinite nature of production resulting from weather, plant and equipment breakdowns, etc. This can compromise the number of samples taken to define the lot parameters.

For HMAC, lot size and test frequency have been noted to be quite variable from agency to agency.[49] Most State specifications designate a lot size to be approximately one day's production.[50,51] One advantage of the one-production-day lot size is that it is likely that, in one production day, the placement of the mix will be fairly uniform. California uses a lot size of 1,360 metric tons.[17] However, recommendations have been made to increase the lot size to 5,000 metric tons. Benson states that "...increasing the lot size for determining asphalt concrete density to 1 week's production is feasible from a statistical standpoint. Initially, risks will be higher as contractors learn to control the quality of their product. In the long run, the savings realized by a more efficient testing program and the improvements in product quality expected under an ERS [end-result specification] will reduce these risks to acceptable levels."(48)

For the prototype PCC specification developed under the earlier FHWA project, the lot sizes are defined as no more than a day's production, but no shorter than 0.5 km. This is to allow for designation of a minimum that can be accommodated by the surface profile measuring devices.(29)

In the New Jersey DOT specifications, lot sizes are defined as a day's production for a given class of concrete. Subdividing a day's production into smaller lots is allowed but is left to the discretion of the engineer in charge.(11) The Michigan DOT also uses a day's production to define a lot size.(19)

In the Oklahoma DOT specification, lot sizes are defined as 8,360 m^2, which is then broken into four equal sublots of 2,090 m^2. Sampling rates for gradation, air content, strength, and thickness are defined by sublot. Sampling rates for the pavement roughness are in 0.16-km increments.(11)

The Indiana DOT defines a lot size as 5,760 m^3, with three equal sublots of 1,920 m^3 each.(22) Flexural tests are conducted at a rate of two per sublot, with one air content and one unit weight performed for each sublot. Water-cement ratio and gradation verifications are performed on every two lots or once per week, whichever is more frequent.

The Illinois DOT recently changed to different lot sizes for different quality characteristics. PCC slump, compressive strength and flexural strength are now sampled from 2,670 m^3 lots. Air content is sampled on the basis of 135-m^3 lots, and thickness sampling is based on lots 300 m long. Lot sizes of 305 m are defined for smoothness sampling.

The most recent *AASHTO Quality Assurance Guide Specification* contains guidelines for sublots for thickness and strength (152 lane-m), but leaves the determination of lot sizes up to the agency.[1] For measuring smoothness, a lot is defined as a project, with 161-m sublots. [1]

The Washington DOT does not define lot sizes. Instead, DOT personnel sample from each truck until two consecutive trucks are found that meet all applicable requirements.[21] The tests performed include slump, air content, and concrete temperature. Once two consecutive trucks with acceptable test results are found, testing continues at a rate of once every five trucks. If unacceptable results are found at any point, the original testing plan of every truck is resumed until two consecutive trucks with acceptable results are found.

INFLUENCE OF SAMPLE SIZE ON QUALITY MEASURES

Traditionally, the sample size has been chosen to be compatible with the lot size, with primary concern given to labor requirements. For asphalt concrete, for instance, four or five samples per day often have been used because it was known that one technician could run gradation and asphalt content tests on four samples per day, and not because the information provided by this number of samples was ideal.[52] The testing capability often has been the driving force to establishing the sample size and, subsequently, the lot size.

One analysis indicates that excellent approximations of PD can be made with the use of the noncentral-t distribution as long as the sample size is 10 or more.[53] Several statistical formulas provide guidance to the number of samples needed to estimate the population average within certain limits. The goal of the sampling rates in a specification is to estimate the mean and standard deviation of quality characteristics as accurately as possible with as few tests as possible. A simulation conducted during the *Performance-Related Specifications for Concrete Pavements* study proposed a chart of pay factor standard deviation versus the number of samples per sublot.[54] The simulation was intended to see where the cost of additional testing began to experience diminishing returns with respect to getting the most accurate pay factors possible. The reduction in standard deviation is minimal from 4 or 5 samples per sublot up to 10 samples per sublot. There is a distinct reduction between one and two samples per sublot, no difference between two and three, and a slight drop between three and four. On the basis of this simulation, the optimum number of samples per sublot would appear to be about four or five, depending on the cost and time involved in performing the test.

FHWA's prototype performance-related specification for PCC recommends a minimum of three sublots per material lot with sampling rates given for each sublot, as required by AASHTO R-9.[29] For concrete thickness and strength, a minimum of two cores are to be taken from each sublot. The mean values of thickness and strength determined from the cores taken for a given sublot are to be reported as the value for that sublot. Air content also is determined by taking two samples per sublot. Again, the average value is used as the value for that sublot. For roughness measurements, a minimum of two parallel lines of roughness are to be performed for each paving lane for a given sublot. The average of all parallel roughness measurements is to be used as the roughness value for that sublot. In addition to the mean values for strength, thickness, air content, and roughness, target levels of standard deviation are defined for each of these quality characteristics.

For the New Jersey DOT specification, the number of samples per lot depends on the class (criticality) of the concrete being placed and whether or not the item has been designated as a pay-adjustment item. For pay-adjustment items of Class P (prestressed) and Class A (structural), six samples per lot are taken. For Class B (pavement, structural foundations) and Class C (other foundations, slope protection, etc.) the sampling rates are five and four samples per lot, respectively. For non-pay-adjustment concrete, the sampling rates for Classes A, B, and C are three, two, and one samples per lot, respectively. If either a pay-adjustment or a non-pay-adjustment item is determined to be rejectable, a retest sampling rate of six cores per lot applies. When non-pay-adjustment concrete items are retested, they then are treated as pay-adjustment items to make the final acceptance determination.

For the Oklahoma DOT specification, the gradation and air content tests are performed on one specimen for each sublot; strength testing is performed on two cylinders per sublot (taken from different trucks), with the average result reported as the value for that sublot. For thickness, three cores are taken from random locations from each sublot. Because the Oklahoma DOT divides the lots into equal sublots, each of the sampling rates per sublot would be repeated four times for each material lot. For roughness measurements, values are taken in 0.16-km increments for 100 percent of the length of each sublot.[11]

TYPICAL PAVEMENT VARIABILITY

QC and quality acceptance are the two keys to building a quality pavement. A malfunction in either of these two testing plans will cause an inferior pavement to be built. An inferior pavement is one that has substantial variation in all directions — longitudinal, transverse, and vertical. However, a certain amount of variation is inherent in the pavement no matter how well the quality control and quality acceptance plans are written and followed. The NCHRP Synthesis 232, *Variability in Highway Pavement Construction,* discusses the "Sources of Variability" and quantifies typical variabilities of many pavement material and construction properties.[6]

Pavement smoothness, in particular, is receiving a lot of attention in contemporary specifications. Smoothness is being viewed as important for several reasons, one being that the traveling public tends to recognize smoothness.

Asphalt Concrete Pavement Variabilities

A recent study evaluated some asphalt concrete pavement characteristic variabilities.[48] One of the findings was that variability in compaction across the mat at a single station was the same as it was from station to station. Variability in the compaction operation, not the compactibility of the material, was the major contributor to the total density variability of the pavement density. Statistically significant differences were found between the pooled longitudinal and transverse data and the vertical data. Variations in fundamental engineering properties of cores taken both transversely and longitudinally were essentially the same.

The variability of density measurements also is discussed in a 1997 Association of Asphalt Paving Technologists (AAPT) paper.[34] "Pavement density ranged about 3 to 4 percent in the transverse direction... The cross-sectional variation and longitudinal fluctuations collectively yield a density range of 3 to 10 percent," according to Schmitt, et al.

A 1967 study referenced in an FHWA summary report contains the following conclusion regarding the effect of rolling pattern on density:(55)

> Normal rolling procedures used by roller operators result in wide lateral variations in compactive effort. The number of roller passes applied to the center of the lane is usually from three to six times greater than at the lane edges. The lateral pattern of density is similar to the lateral pattern of compactive effort; i.e., high-in-the-middle and low-at-the-edges.

For smoothness, FHWA Central and Western Lands Divisions have reported typical variabilities of California-type profilographs, of about 0.03 m/km (1.9 in/mi).(6)

PCC Material Variabilities

Some interesting studies have been done on the variability of structural concrete materials and the variability of roughness measurements along PCC pavement sections.

One such paper describes an experiment using three 1 x 1 x 2-m concrete columns made with three different strength concrete mixtures.(56) The findings of this research effort were essentially that the higher strength concrete experienced no statistically significant changes in measured strength values when different layers of the columns were compared (top, middle, and bottom) or when different positions within a layer were compared (interior, edge, and corner). The lower strength concrete showed statistically significant differences for both compressive strength and split tensile strength values, though the split tensile had only one significant difference and the compressive strength values had five.

In addition to the variability in PCC materials placed in the field, the laboratory tests used to determine the quality values have some variability associated with them. An article published in the December 1994 *Concrete International* details an investigation of a project that was showing concrete compressive strengths as much as 30 percent below the required strengths.[26] The article discusses ACI 214-89 "Recommended Practice for Evaluation of Strength Test Results of Concrete," which provides the following guidelines for determining the quality of a laboratory operation based on the within-test coefficient of variation:

Within-test COV*	Laboratory Control
Below 3.0	Excellent
3.0 to 4.0	Very Good
4.0 to 5.0	Good
5.0 to 6.0	Fair
Above 6.0	Poor

*Variation in compressive strength between replicate cylinders tested by the same operator.

The laboratory in question was found to be operating at a 13.0 percent within-test COV. Another problem discovered during the follow-up investigation was that approximately 10 percent of the 7-day strength values were above the corresponding 28-day strength values.

Additional data reported COV values for PCC pavement thickness generally less than 3 percent (though sometimes as high as 8 percent), COV values for elastic modulus ranging from 21 to 49 percent, COV values for Poisson's ratio ranging from 9.3 to 20.2 percent, and modulus of rupture COV values ranging from 2.8 to 17.6 percent for 7-day flexural strength and from 3.5 to 9.6 percent for 28-day flexural strength.[26] Guidelines for rating construction control also are given for the total COV values of compressive strength as follows:

Total COV Value*	Construction Control
Below 10	Excellent
10 to 15	Good
15 to 20	Fair
Above 20	Poor

*Includes inherent material, sampling, and testing variation.

A study conducted in Western Canada provides an alternative for evaluating compressive strength test results for PCC materials.[57] The findings indicate that the 95-percent repeatability limit for within-laboratory variability is approximately 10 percent and the between-laboratory 95-percent reproducibility limit variability is approximately 15 percent. These values imply that two test results, based on test error alone, can vary by at least 10 or 15 percent, 5 percent of the time (1 time in 20), without there being a significant difference between the values.

RELATIONSHIPS BETWEEN QUALITY CHARACTERISTICS AND PERFORMANCE

Knowing the relationships between quality characteristics can help to limit the amount of testing that is necessary for acceptance. By knowing that one parameter is directly related to another, it will not be necessary to test both parameters for acceptance.

For instance, the gradation of adjacent sieves often has been shown to correlate; as one sieve increases in percent passing, the adjacent sieve can be expected to increase proportionally in percent passing. Within a hot bin in asphalt batch plants, one sieve often is viewed as the critical sieve. As the percent passing of that sieve changes, the other sieves within the gradation of that bin change similarly.

QC and acceptance plans have different purposes. QC tests are for the contractor to use to control the process so as to minimize and, hopefully, eliminate non-specification product. Acceptance tests determine whether or not the product does meet specifications. From this viewpoint, the QC and acceptance tests do not have to be the same, although, realistically, they usually are.

In determining which parameters should be monitored for acceptance, it is important to know what relationships exist between the quality parameters and pavement performance. Knowing these relationships will aid in determining which parameters should be used for acceptance testing. "The objective of monitoring the engineering properties is to verify that the plant-mixed materials have the engineering properties that are the same as those of the ...design," according to Hughes.[58]

Performance Relationships

Asphalt Concrete

The influence of low in-place air voids of well-designed asphalt mixes on pavement performance has been well documented.(58) Strength and durability have been shown to be adversely affected when high air voids are found after construction compaction.(59)

Other work has demonstrated that fatigue life is generally higher for higher moduli, lower applied stress levels, higher voids in mineral aggregate (VMA) and, at times, lower asphalt contents.(60) A step-wise linear regression demonstrated that rutting was a function of voids filled with asphalt (VFA), the hump in the gradation curve, air voids between the wheelpaths (i.e., a measure of construction compaction, not traffic consolidation), and Marshall stability.(61)

One laboratory study focused on the development and verification of relationships between materials variables, construction variables, and fundamental response variables.(62) Compaction was found to be a function of VMA, air voids, percent of HMAC aggregate passing the 0.6-mm sieve, percent of HMAC aggregate passing the 0.075-mm sieve, and asphalt content. Resilient modulus (tested at 25 °C) was related to compaction, asphalt type, VMA, percent deviation from optimum asphalt content, and percent HMAC aggregate passing the 0.075-mm sieve. The retained modulus, after moisture conditioning, is a function of the presence of lime, asphalt type, VMA, and compaction. The retained strength, after moisture conditioning, is a function of percent HMAC aggregate passing the 0.075-mm sieve, presence of lime, asphalt type, and VMA.

Another study compared the effects of varying gradations.(63) Gradation variations have the greatest effect when they change the slope of the JMF gradation curve. Creep stiffness was the lowest for the gradations that had the greatest change in slope. Gradation variations that produced a coarser mix produced the lowest indirect diametral tensile strength. However, Elliott, et al. states that "within the range of variations normally encountered, tensile strength is more sensitive to air void content than it is to gradation variation."(63)

Yet another study concentrated on density.[64] This study found that, because of the relatively rapid cooling rate of thin layers, it was more difficult to achieve density in thin layers than in thick ones. The projects that had segregation or tenderness problems did not achieve densities as high as those without any problems. No meaningful relationship was found between the average density and dust-to-asphalt cement (AC) ratio. Another study found that "...mixtures made with uncrushed aggregates compact into a denser arrangement than do crushed aggregates which have more macro-texture and micro-texture."[65]

A Strategic Highway Research Program (SHRP) study examined the sensitivity of various distresses and roughness to pavement attributes.[66] Tables 3, 4, and 5 list the results of these analyses for HMAC pavements. The distresses studied include rutting, roughness, and transverse crack spacing. The variables on the left-hand side of each table were those found to be significant in the prediction of that distress. The numbers within the table list the relative significance of the independent variable to the prediction of the distress. For instance, the cumulative 18-kip equivalent single axle load (KESALs) was the most significant in the prediction of rutting. Furthermore, rutting was most sensitive to changes in KESALs in all environmental zones and pavement types except full-depth pavements in the dry zone and granular-base pavements in the dry/freeze zone.

Another study recently completed compared the attributes of pavements that performed well and those that performed poorly, as shown in table 6.[67] In this table, the "I" means that an increase in the variable on the left will cause an increase in the distress. The "D" indicates that an increase in the independent variable will cause a decrease in the expected distress. The question mark indicates an uncertain or variable effect. Finally, the air voids used in this study were measured after consolidation by traffic; hence, any sign indicated by the analyses may not be the same sign as for air voids measured during construction.

Tables 3, 4, 5, and 6 provide a list of the most important factors affecting pavement performance. A review of these tables provides a list of variables that can be controlled by the contractor and are important to the development of pavement distress. Among these characteristics are air

Table 3. Orders of Significance for Independent Variables, All Models for Rutting of HMAC Pavements[66]

Independent Variables	HMAC on Granular Base					Full-Depth HMAC					HMAC on Portland Cement-Treated Base	No. of Models Found Significant	Average Rankings
	All Zones	WNF Zones	WF Zones	DNF Zones	DF Zones	All Zones	W Zones	D Zones	NF Zones	F Zones			
KESALs	1	1	1	1	5	1	1	3	1	1	1	All	1.5
HMAC Air Voids	2	3	4	2	4	3	4	-	-	2	2	9	4.2
HMAC Thickness	3	2	5	7	3	6	5	5	7	6	6	All	5.0
HMAC Aggr.< #4	4	5	3	5	-	-	7	-	-	-	4	5	7.1
Asphalt Viscosity	-	-	8	-	-	-	6	-	3	-	5	4	8.4
Asphalt Content	-	-	-	-	-	-	-	-	-	4	-	1	9.5
Base Thickness	5	6	7	8	2	N/A	N/A	N/A	N/A	N/A	3	All	5.2
Base Compaction	-	-	-	-	1	N/A	N/A	N/A	N/A	N/A	-	1	8.5
Subgrade <# 200	7	4	-	6	7	2	3	4	4	5	-	9	5.6
Days > 32°C	-	8	-	3	-	-	2	1	5	3	-	6	6.5
Annual Precipitation	-	-	6	4	-	5	-	-	2	-	-	3	7.9
Freeze Index	6	-	2	-	6	-	-	-	-	-	-	3	8.5
Annual Freeze-Thaw Cycles	-	7	-	-	-	-	-	-	-	-	-	1	9.7
Daily Temp. Range	-	-	-	-	-	4	-	-	6	-	-	2	9.1
Avg. Annual Min. Temp.	-	-	-	-	-	-	-	2	-	-	7	2	9.0

Note: N/A = Variable not applicable for the data set.
Empty cells are considered as 10 for averaging.
WNF: Wet-No Freeze WF: Wet-Freeze DNF: Dry-No Freeze DF: Dry-Freeze
W: Wet D: Dry NF: No Freeze F: Freeze

Table 4. Orders of Significance for Independent Variables, All Models for Change in Roughness of HMAC Pavement[66]

Independent Variables	HMAC on Granular Base					Full-Depth HMAC	HMAC on Portland Cement-Treated Base	No. of Models Found Significant	Average Ranking
	All Zones	WNF Zones	WF Zones	DNF Zones	DF Zones				
KESALs	1	3	1	1	5	1	2	All	2.0
Air Voids in HMAC	9	8	2	-	7	-	6	5	7.7
HMAC Thickness	5	5	4	3	6	6	5	All	4.9
HMAC Aggr. < #4	-	-	-	-	-	-	-	None	11.0
Asphalt Viscosity	6	7	3	5	2	-	-	5	6.4
Asphalt Content	10	-	-	-	-	4	-	2	9.9
Base Thickness	7	2	7	7	3	N/A	7	All	5.5
Base Compaction	4	10	-	-	-	N/A	-	3	9.7
Subgrade < #200	8	4	-	6	-	3	1	4	6.3
Days > 32°C	2	9	6	2	4	5	3	All	4.4
Annual Precipitation	11	4	-	4	-	-	4	4	8.0
Freeze Index	3	-	5	-	1	2	8	5	5.9
Annual Freeze-Thaw Cycles	-	1	-	-	-	-	-	1	9.6
Daily Temp. Range	-	6	-	-	-	4	-	2	9.3

Note: N/A = Variable not applicable for data set
Empty cells are considered as 11 for averaging.

Table 5. Orders of Significance for Independent Variables, All Models for Transverse Cracking in HMAC Pavements(66)

Independent Variables	HMAC on Granular Base and Full-Depth HMAC					No. of Models Found Significant	Average Rankings
	All	WNF	WF	DNF	DF		
Annual KESALs	-	-	-	-	2	1	8.4
Air Voids in HMAC	8	-	-	-	-	1	9.6
HMAC Thickness	5	5	7	2	6	All	5.0
HMAC Aggr. < #4	-	-	5	-	-	1	9.0
Asphalt Viscosity	3	9	-	-	1	3	6.6
Asphalt Content	-	7	-	-	-	1	9.4
Base Thickness	7	4	6	4	4	All	5.0
Base Compaction	1	3	-	-	-	2	6.8
Subgrade < #200	-	-	4	-	7	2	8.2
Days > 32°C	-	-	1	-	-	1	8.2
Annual Precipitation	2	6	3	1	-	4	4.4
Freeze Index	-	2	-	-	5	2	7.4
Annual Freeze-Thaw Cycles	4	-	-	-	-	1	8.8
Daily Temp. Range	-	8	-	-	-	1	9.6
Age	6	1	2	3	3	All	3.0

Note: Empty cells are considered as 10 for averaging.

Table 6. Effects of Variables on HMAC Performance[67]

Characteristic	Distress Type			
	Rutting	**Fatigue Cracking**	**Transverse Cracking**	**Roughness**
AC Thickness	D	D	D	D
Base Thickness	D	?	?	?
Air Voids in AC	*	*	*	*
Asphalt Viscosity	I	I	D	I
Base Compaction	?	?	?	I
Structural Number	D	D	?	D
Expected ESALs	I	I	I	I
Annual No. of Days > 32°C	I	D	D	?
Freeze Index	?	?	I	I
Annual No. of Freeze-Thaw Cycles	?	?	I	I
Annual Precipitation	I	I	I	?
Subgrade < 0.075-mm Sieve	?	?	?	I
Age	?	?	I	?

* Only initial air voids are controllable, and data available are for air voids after consolidation by traffic.

D = indicates a decrease

I = indicates an increase

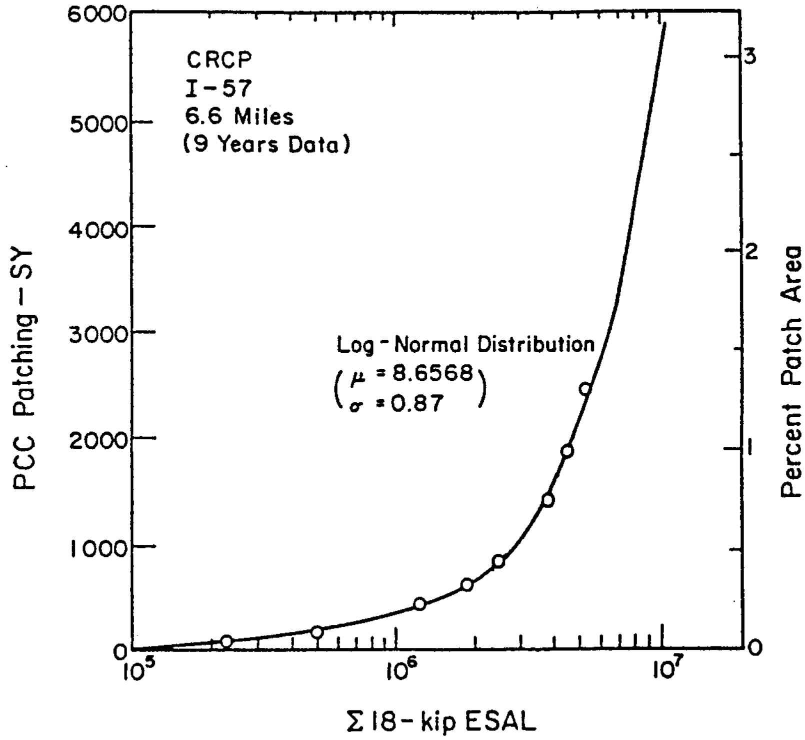
? = indicates uncertain or variable effects voids, HMAC thickness, asphalt cement viscosity, base thickness, and base compaction. These factors should be given serious consideration when developing a quality control specification.

The only conclusion that can be drawn from this compilation of references is that, for HMAC construction, many different lot and sample sizes exist in various States, each with their own justification. Also, variability of most quality characteristics changes from project to project but can be typified.

PCC

The point-to-point variation in materials and construction variables along a pavement has a significant effect on performance. As the variation of such items as strength, entrained air content, smoothness, slab thickness, steel depth, density of concrete near doweled joints, slab support, and others increase along a given lot, the variation in distress and rideability over time should increase. This would result in increased maintenance and rehabilitation costs and corresponding lane closures causing increased delays and congestion.

An example that shows the rate of early failures for a continuously reinforced concrete pavement (CRCP) over time and traffic in Illinois is shown in figure 9.[68] Patching quantities done each year over an 8-year time period are plotted versus the cumulative 80-kN ESAL on the pavement. A log-normal distribution curve has been fitted through the data points as shown. The patched areas are those areas that failed structurally early in the life of the pavement, probably from deficiencies in quality characteristics like concrete strength, slab thickness, reinforcement placement, and slab support. Had there been less variation during construction, this curve may have been much flatter than the one shown in figure 9. This pavement failed after about 10 years and required a major rehabilitation.



1 mi = 1.61 km

1 sy = 0.836 m^2

Figure 9. PCC Patching Amounts Versus Loading[68]

Relatively little research has been conducted in this important area. The PAVESPEC software for PCC pavements discussed earlier makes it possible to change the point-to-point variation in four quality characteristics: strength, thickness, air content, and smoothness along a project for the as-constructed lot.[29,54] As point-to-point variation increases, cracking, spalling, and other distresses develop at an earlier age. The result is increased maintenance costs and eventually major rehabilitation. The impact of this variation on the ultimate contractor pay factor as computed from life-cycle costs is shown in figures 10 and 11 for slab thickness and strength. As the COV's of the quality characteristics increase, the contractor's expected pay decreases. For example, if the slab thickness COV increases from 2 percent (the specification target) to 8 percent, the expected pay decreases from 100 to 92 percent. This represents a significant life-cycle cost increase for the pavement lot and a large economic loss to the contractor. Additional research is greatly needed to relate variation within a lot to the performance of the lot.

A performance-related quality specification requires a relationship between the quality characteristic and the development of distress and roughness. Several studies have developed relationships between quality characteristics and distress development. Table 7 shows a summary of the material properties and pavement distresses that were found to be significant by the various studies.

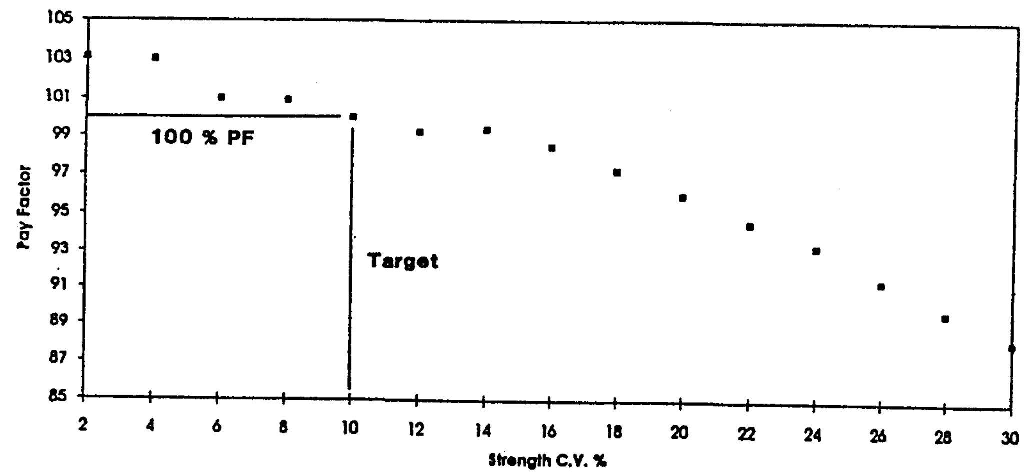


Figure 10. Pay Factor Versus COV of Compressive Strength(29)

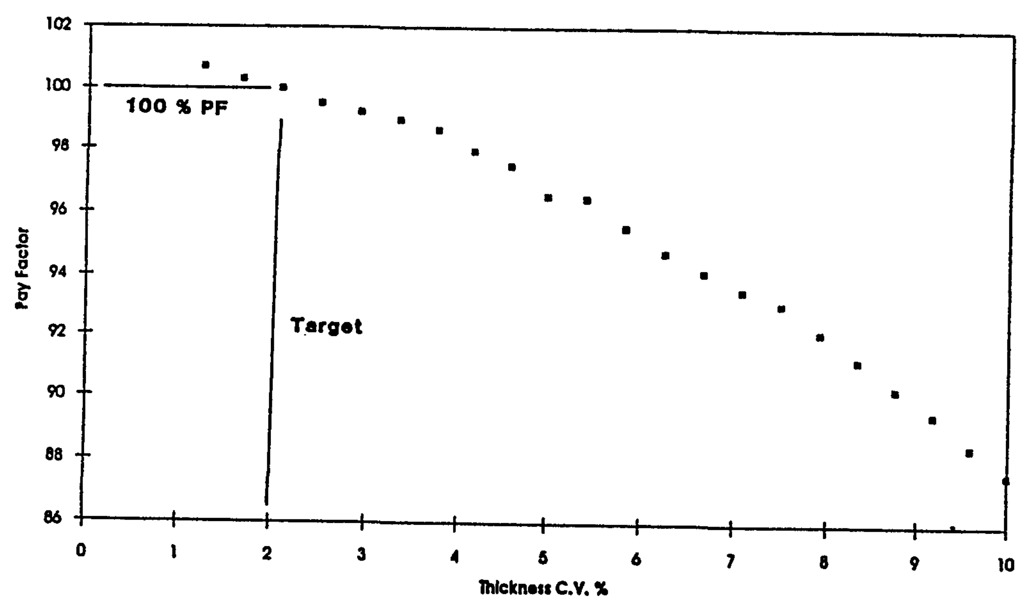


Figure 11. Expected Pay Factor Versus COV of Slab Thickness(29)

Although the development of relationships between quality characteristics and expected performance is important, one must remember that a myriad of other variables influence PCC pavement performance, and cannot be taken into account when establishing QA requirements to ensure good future performance of pavement sections. A study of PCC pavements in Illinois showed tremendous variability in the faulting, cracking, and number of deteriorated joints that developed for 12 seemingly identical projects.(71) The ability to predict highly variable field performance on the basis of a limited number of quality characteristics is a very difficult, though necessary, part of QA specifications.

Table 7. Relationships Between Material Properties and Distress Development for PCC Pavements

Pavement Distress or Smoothness Quality	**Material Properties and Construction Variables**			
	Slab Thickness	**Elastic Modulus**	**Concrete Flexural Strength**	**Initial Smoothness**
Joint Faulting	COPES[c]			
Pumping	COPES			
Transverse Cracking of JPCP	SHRP[a] RPPR[b]	SHRP	SHRP COPES	
Transverse Cracking of JRCP	RPPR COPES		COPES	
Longitudinal Cracking	RPPR			SHRP
IRI of Doweled JPCP	SHRP			SHRP
IRI of JRCP	SHRP			SHRP
IRI of CRCP	SHRP			SHRP

Notes: [a] SHRP P-020 Study,(66) [b] FHWA Project RD-89-136,(69) [c] NCHRP Study 1-19 (70)
JPCP: Jointed Plain Concrete Pavement
JRCP: Jointed Reinforced Concrete Pavement
RPPR: Rigid Pavement Performance and Rehabilitation
IRI: International Roughness Index
COPES: Concrete Pavement Evaluation System
CRCP: Continuously Reinforced Concrete Pavement

A laboratory study performed as part of the *Performance-Related Specifications for Concrete Pavements* study investigated the effects of various material properties on the development of two distress types: transverse cracking, caused by repeated loading and thermal curling; and joint spalling, caused by an inadequate air-void system.[11] The variables included in the laboratory study were as follows:

- Coarse aggregate hardness
- Coarse aggregate geometry
- *Coarse aggregate maximum size*
- *Fine aggregate fineness modulus*
- Air content
- Coarse aggregate volume percentage
- Cement volume percentage
- *Water volume percentage*
- *Fine aggregate type*
- *Consolidation level*
- *Mineral admixture*
- *Cement type*
- *High-range water reducer*

Those variables shown in italics only were included at a single level, so their significance to the two distress types could not be determined.

Several tests were performed on the plastic concrete, including the following:

- Slump
- Unit weight
- Initial concrete temperature
- Air content

Several tests also were performed on the hardened concrete, including the following:

- Compressive strength (152- × 305-mm cylinders)
- Splitting tensile strength (152- × 305-mm cylinders)
- Third-point flexural strength (152- × 533-mm long beams)
- Modulus of elasticity (152- × 305-mm cylinders)

Some of the more interesting findings of the study indicated that the higher consolidation levels improved the resistance to spalling after freeze-thaw cycling in the presence of salt, but even lower consolidation levels did not develop spalling without the presence of salt. The laboratory study also found that there was a significant decrease in flexural strength as the consolidation level decreased from 100 to 94 percent. The flexural strength decrease ranged from 8 to 25 percent for various aggregate types, whereas the corresponding decrease in compressive strength only ranged from 1 to 12 percent for the same aggregate types. Another portion of the study indicated that there were no statistically significant differences between the strength measurements obtained from cores and cylinders cured under identical maturity conditions. These results are based on testing eight different mixes at 7, 14, and 28 days.

Results of the above-mentioned study of the effects of consolidation levels showed a relationship between inadequate consolidation and lower expected pavement life.(11) By determining the ratio of flexural strength at 100-percent consolidation to flexural strength at lesser degrees of consolidation and by using the reduced flexural strengths in the AASHTO design equation, the expected service life dropped from 20 years at 100-percent consolidation to 5.4 years at 90-percent consolidation.

A more recent study looked at the relationship between concrete density and 28-day compressive strength, in particular when the 28-day strengths were less than specified.(72) The emphasis of the study was that, although the 7-day strengths do provide a good indicator of 28-day strength, the relationship is not very strong for 28-day strengths less than specified. The ability of concrete density to predict 28-day strengths where the 28-day strength is less than specified provided an R^2 value of 0.766 versus an R^2 of 0.213 using the 7-day strength to predict 28-day strength for the same less-than-specified data set.

The most frequently used relationships between different quality characteristics of PCC pavements are those used to convert concrete compressive strength to flexural strength and elastic modulus values since they are values needed for predicting the pavement performance given the available PCC pavement performance models. Laboratory data obtained during the FHWA "Performance-Related Specifications for Concrete Pavements" project determined that the flexural strength was a function of the square root of compressive strength.(11) The same study also determined that the splitting tensile strength was a function of the $\log_{10}$ of the compressive strength. The elastic modulus also was found to be a function of the square root of the compressive strength, though the error mean square (EMS) ratio indicated a serious lack of fit that was not indicated for the other two relationships.

To reduce the variability inherent in the relationships between the various strengths, individual relationships can be generated for each mix produced. Transformations for the dependent and independent variables can be accomplished using square root, logarithmic, and inverse transformations, with the model having the highest R^2 value considered to be the most representative.(11)

Another recent study proposed a method for converting concrete core strength to an equivalent in-place strength through the use of several correction factors.(73) These correction factors include a length-to-diameter ratio correction, core diameter correction, reinforcing bar correction, moisture condition correction, and a damage-during-coring correction. For example, a 100-mm-diameter (Fdia = 1.0), 200-mm-long core (F1/d = 1.0) with no reinforcing bars (Fr = 1.0) that is soaked before testing (Fmc = 1.09) and shows normal coring damage (Fd = 1.06) would have an in place equivalent strength factor of $1.0 \times 1.0 \times 1.0 \times 1.09 \times 1.06 = 1.16$. This means that the concrete strength values in place are approximately 16 percent greater than those determined in the laboratory when testing core samples.

One other property of PCC jointed pavements that has been shown to have a significant effect on pavement performance is the correct placement of dowel bars. Dowel bar misalignment has been shown to affect spalling, cracking, and load transfer.(74) Another study conducted in Wisconsin estimated that the misalignment of dowel bars could be within 3.5 percent horizontally and 2.0 percent vertically without significantly affecting the movement of the joint.(75)

SUMMARY

The current trend in pavement construction is toward performance-related specifications that attempt to correlate future performance to the initial measured values of certain quality characteristics. Although some present-day specifications simply compare the mean quality values to the targeted design values, the impact of variability associated with the mean value is being recognized as substantial. Distributions with high mean values and high variability can lead to more substandard materials in place than distributions with slightly lower means and with much smaller variability. The increase in the number of agencies using the PWL (or PD) concept is an indication that variability is being viewed as more important than in the past.

The desirability of decreasing risks for both the contractor and the agency is a strong incentive to go to larger lot sizes and to include more test results per lot than in the past. It may be possible to use a procedure similar to sequential analysis in which the distribution is analyzed periodically for normality; if the distribution is normal, one can continue to add to the data set, essentially increasing the lot size.

It appears that many construction properties are reasonably normally distributed. Some are not. However, some of the properties that appear skewed may represent multimodal distributions that are products of constant process change.

CHAPTER 4. DESCRIPTION OF FIELD PROJECTS

A total of six projects were selected for field evaluation. Of the six projects studied, three were flexible and three were rigid construction. To assess differences in the acceptance specifications, no more than two projects were selected from any one State.

HMAC PROJECTS

A total of three locations (i.e., projects) totaling 9.1 km of HMAC paving were monitored as part of this study. The three locations were:

- Two locations on State Highway 7 (SH-7) outside of Duncan, Oklahoma (Projects OK1 and OK2).
- IH-10 outside of Baton Rouge, Louisiana (Project LA).

The following sections summarize the conditions at each test site layout and the particulars of construction monitoring.

Duncan, Oklahoma Site No. 1

Figure 12 shows the schematic of this location. The test site consisted of a four-lane road, running primarily east-west between Duncan and Lawton, Oklahoma, in Stephens County. Two different sections of the paving job were monitored from Station 486+30 to 436+30 and from Station 539+30 to 489+30.

The test site was new construction consisting of 200 mm of HMAC base, 51 mm of HMAC surface, and a 19-mm open-graded friction course placed on prepared subgrade. Deflection testing was conducted on the subgrade on August 14-15, 1995. The road was opened to traffic in early October 1995, while FWD testing on the finished surface was completed on February 23, 1996. Bulk samples of the HMAC surface mix were obtained in September 1995 and tested in December 1995 in the laboratory. GPR testing was performed on this site on October 14, 1995.

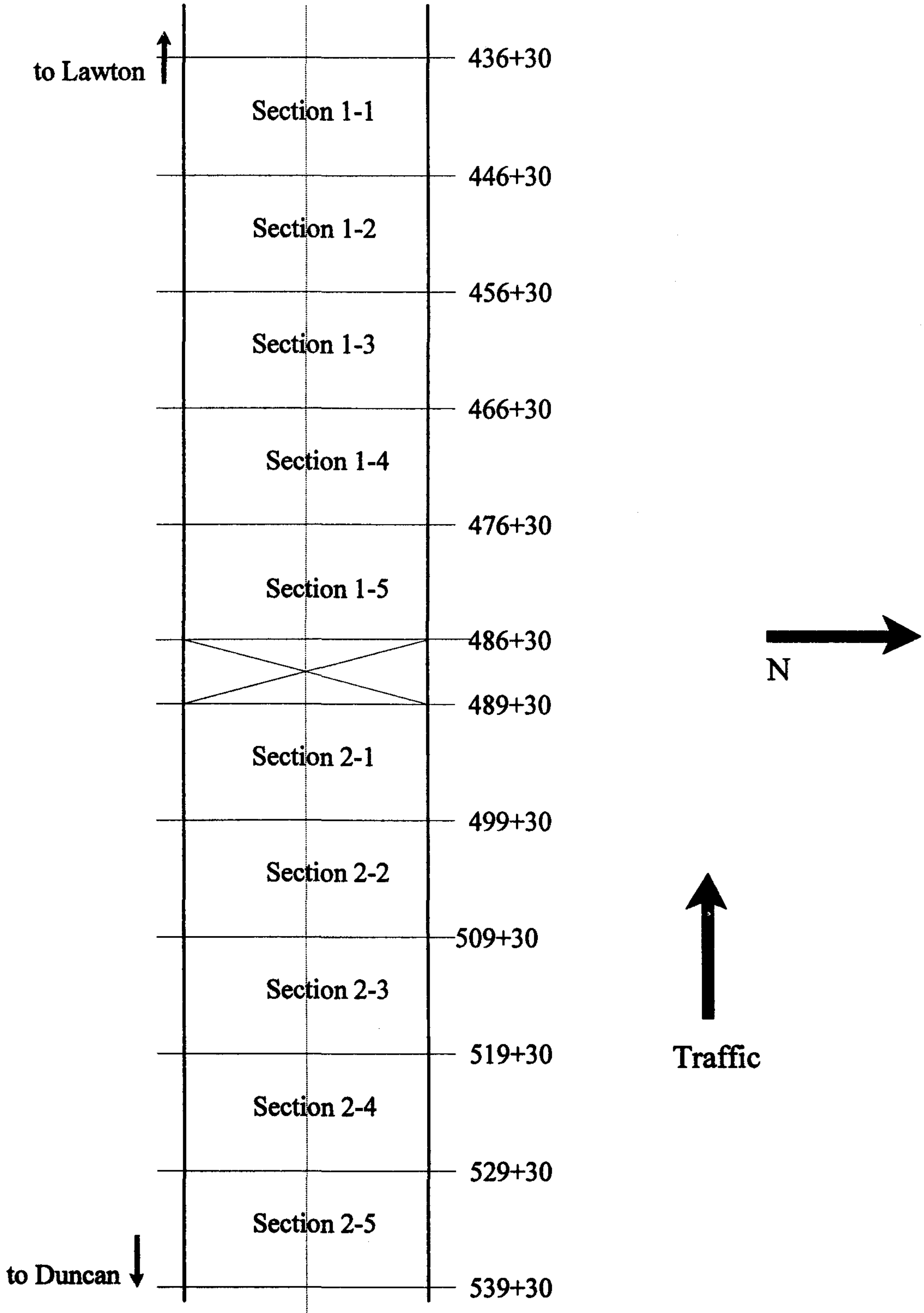


Figure 12. Location of HMAC Test Sections, Duncan, OK Test Site No. 1

Duncan, Oklahoma Site No. 2

Figure 13 shows a schematic of this test site. This test site is located directly west of the site previously mentioned on SH-7 and was built by a different contractor 6 months later. Two different sections of the construction were monitored — from Station 315+50 to 265+50 and from Station 262+50 to 212+50.

Like site 1, this site was new construction consisting of 200 mm of HMAC base, 51 mm of HMAC surface, and a 19-mm open-graded friction course. Deflection testing was conducted on the prepared subgrade on February 22-23, 1996. Monitoring of the placement of the asphalt layers was conducted on April 24-25, 1996. At that time, bulk samples of the mix were obtained and sent to the laboratory for further testing. Laboratory testing was completed on these samples in July 1996. Deflection testing was conducted on the finished surface of the test site on May 23, 1996, and GPR testing was performed on May 24, 1996.

Baton Rouge, Louisiana

Figure 14 shows a schematic of this test site. The test site consisted of a four-lane road running primarily east-west through Baton Rouge. Two different sections of the construction were monitored — from Station 3243+25 to 3193+25 and from 3071+25 to 3021+25.

This construction involved rehabilitation of the existing pavement. The original construction was 254 mm of jointed concrete pavement. In 1989 the joints of the pavement were repaired and the pavement was overlaid with 75 mm of HMAC. The rehabilitation being monitored involved milling 50 mm and replacing with 50 mm of SUPERPAVE™ mix. Deflection testing on the milled surface was completed on June 30, 1996. Construction monitoring and bulk sampling activities were conducted on July 1, 1996. Laboratory testing of these samples was completed in August 1996. Deflection testing was performed on the finished surface on July 3, 1996, and GPR testing was conducted on July 12, 1996.

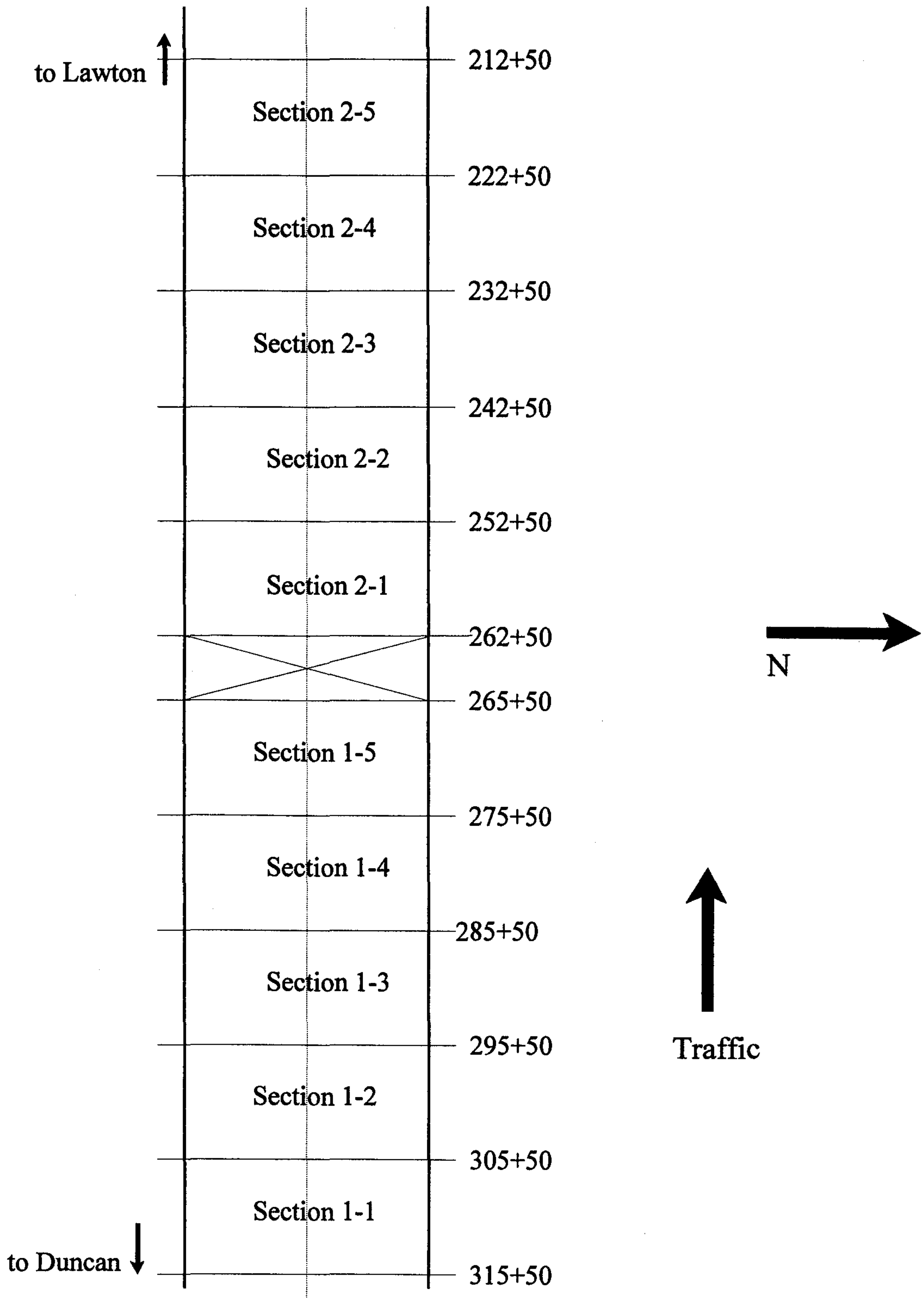


Figure 13. Location of HMAC Test Sections, Duncan, OK Test Site No. 2

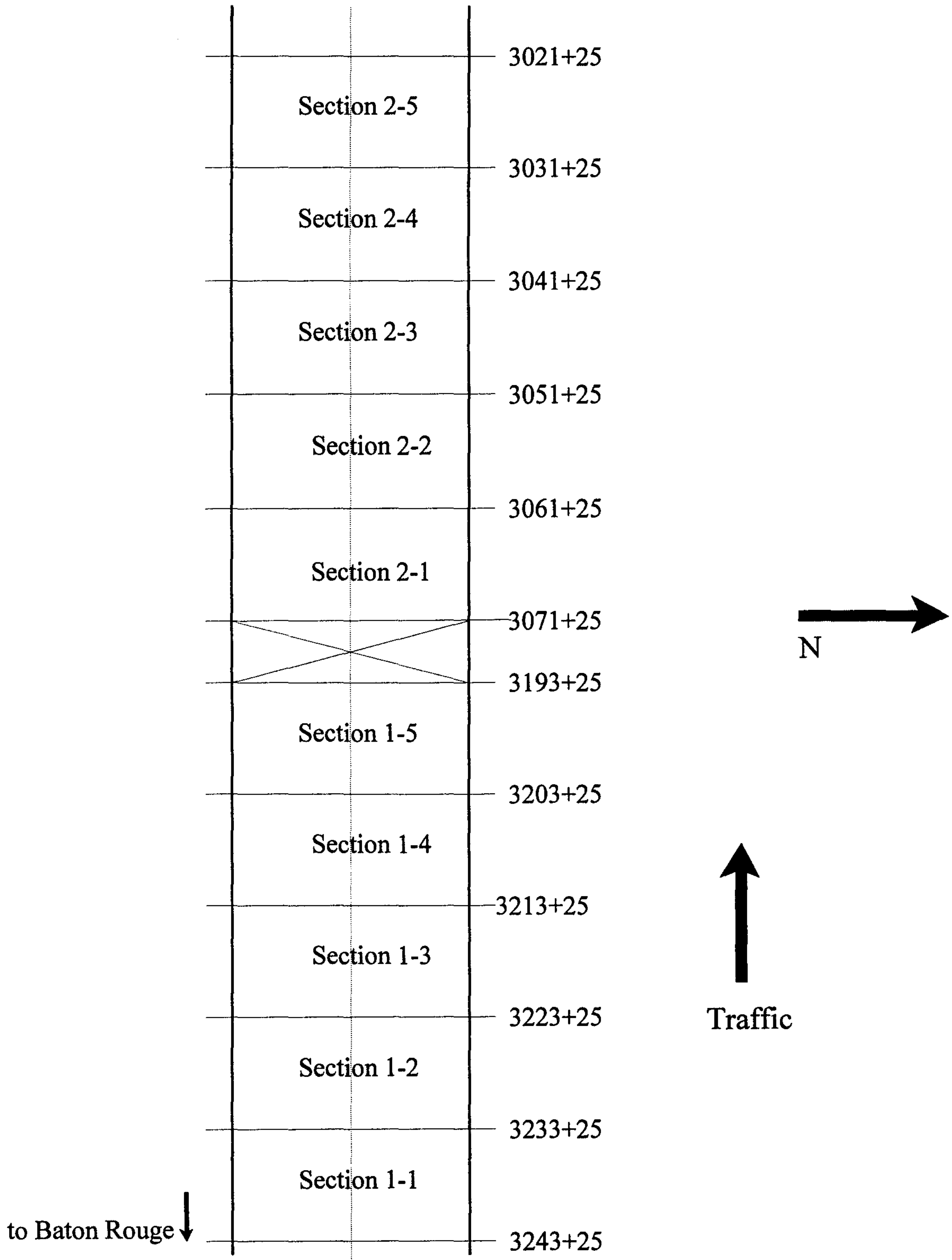


Figure 14. Location of HMAC Test Sections, Baton Rouge, LA

PCC PROJECTS

The construction of three locations (i.e., projects) totaling 9.1 km of PCC paving was monitored as part of this project. The three locations were:

- TH 169 south of Mankato, Minnesota (Project MN).
- Rt. 38 east of Rochelle, Illinois (Project IL).
- US 33 east of Bellefontaine, Ohio (Project OH).

The following sections summarize the conditions at each test site layout and the dates on which the paving operations were monitored.

Mankato, Minnesota

Figure 15 shows the location of this test site. The test site at Mankato, Minnesota, consisted of a two-lane road, running primarily north-south between Mankato and Garden City, Minnesota. Two different sections of the paving job were monitored, from Station 2202+00 to 2252+00 and from Station 2300+00 to 2350+00.

The Mankato test site was a 200-mm nominal thickness, plain jointed pavement, with dowel bars at the joints. The width of the pavement was 7 m for both lanes. Deflection testing on the granular base course layer was performed on May 22, 23, and 26, 1995. Concrete paving was monitored on May 24 and 25, 1995, for section 1 and on May 30 and 31, 1995, for section 2. Deflection testing for the finished PCC surface was performed on both test sections on June 27, 1995. Coring of the pavement in accordance with the sampling and testing plan was performed on July 8, 1995. Cores were turned over to the Minnesota DOT for testing in conjunction with cores taken by the QC contractor on the project. These cores were tested for strength and thickness at 60 days as per Minnesota DOT standards.

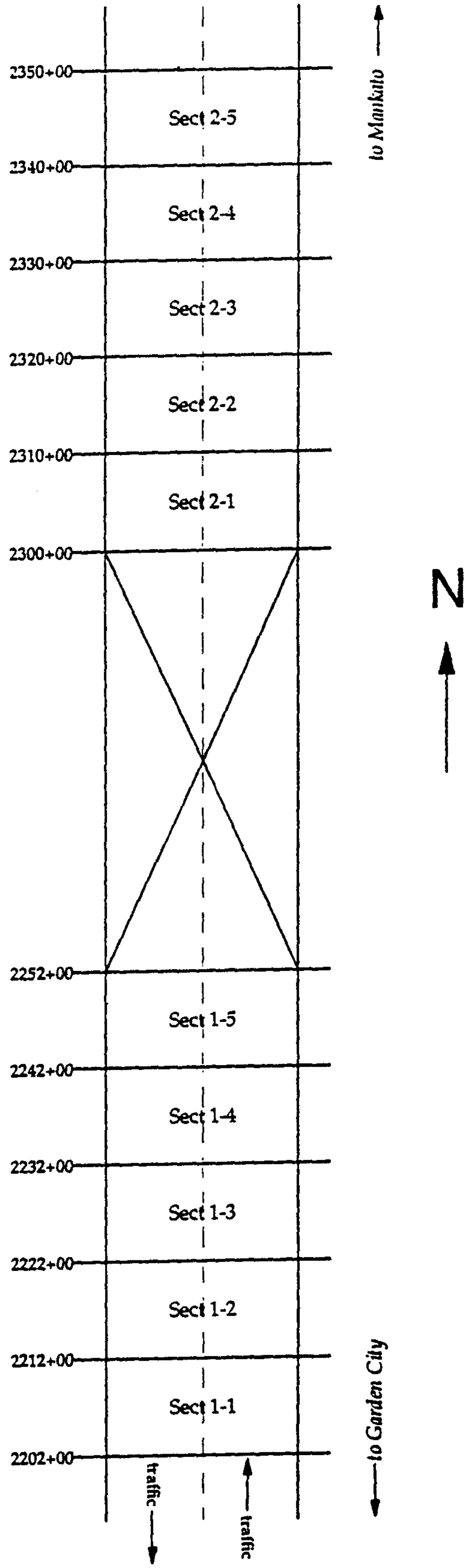


Figure 15. Location of PCC Test Sections, Mankato, MN

Rochelle, Illinois

Figure 16 shows a schematic of the Rochelle, Illinois, test site. The test site at Rochelle consisted of a four-lane divided pavement on Rt. 38 running from the interchange at Interstate 39 west into the town of Rochelle. Two different sections of the paving operation were monitored, both of which extended from Station 132+00 to 190+00, one in the eastbound and one in the westbound lanes. A 2600-m stretch of pavement was not monitored as part of this project from Station 172+00 to 180+00 because of the presence of an intersection that interrupted the paving process.

Because of the layout of the site and delays in construction, final grading of the prepared subgrade was performed in the eastbound lanes between Station 180+00 and 190+00 several months before the subsections located between Station 132+00 and 172+00. Also, there were (and will continue to be) significant differences in traffic between the eastbound and westbound subsection 5 and all other subsections; these are due to the presence of a major truck stop that causes truck traffic to get off Interstate 39, go to the truck stop, and head back to Interstate 39 without traveling over subsections 1 through 4 in either the eastbound or westbound lanes.

The Rochelle test site consisted of a 225-mm nominal thickness, plain jointed pavement, with dowel bars at the joints. The width of the pavement was 7 m for both sections monitored. Deflection testing on the granular base course layer was performed on August 29, 1995, for section 1 and on April 26, 1996, for section 2. Concrete paving was monitored between October 10-12, 1995, for section 1 and on May 30 and 31, 1996, for section 2. Deflection testing for the finished PCC surface was performed on November 3, 1995, for section 1 and on July 2, 1996, for section 2. Coring of section 1 pavement was performed on November 2, 1995, and section 2 pavement was done on June 24, 1996. Cores were tested for compressive strength at 28 days.

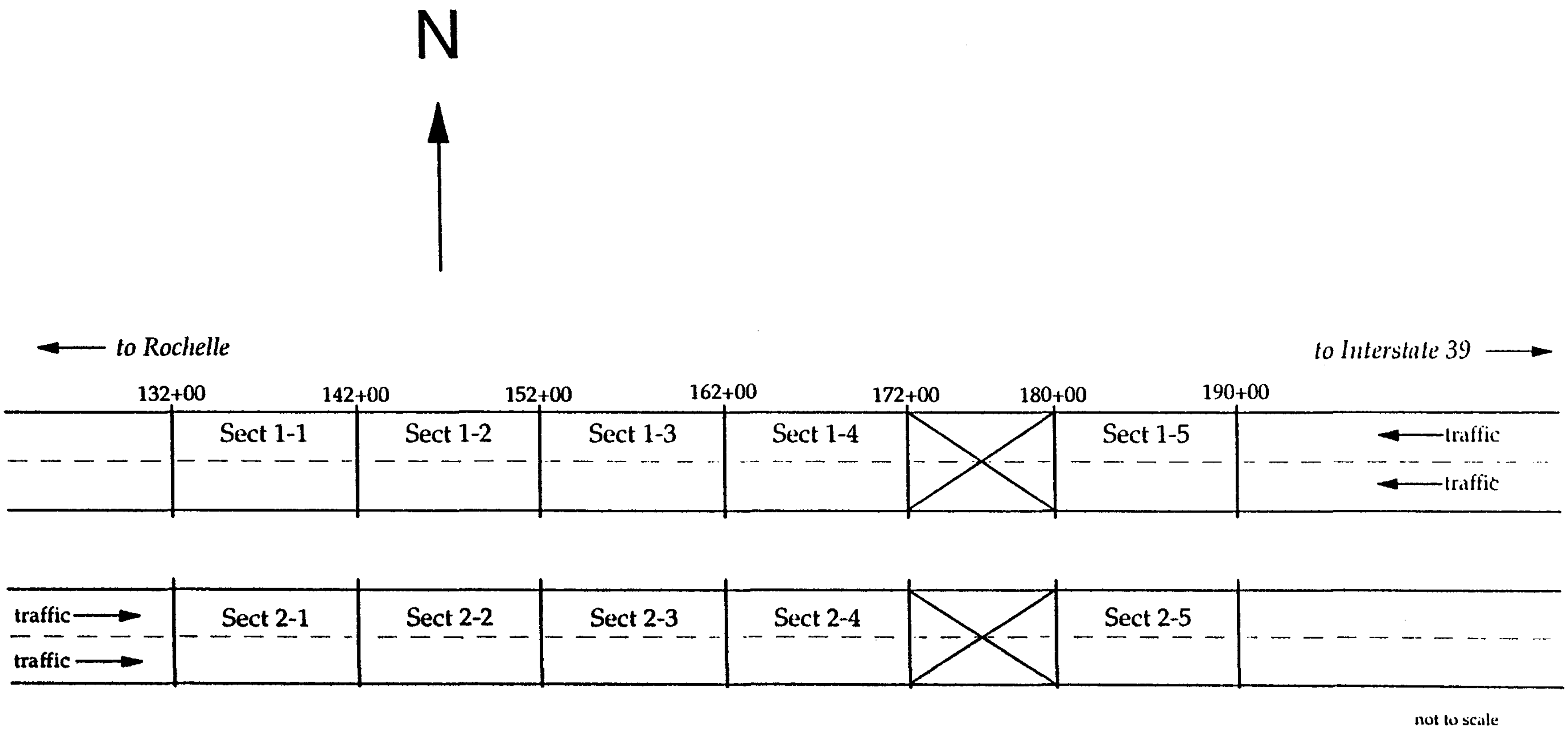


Figure 16. Location of PCC Test Sections, Rochelle, IL

Bellefontaine, Ohio

Figure 17 shows the location of the Bellefontaine, Ohio test site. The test site at Bellefontaine consisted of the two eastbound lanes of a four-lane divided pavement running from approximately CR 10 to CR 153 along US 33, east of Bellefontaine and just south of Zanesfield, Ohio. Two different sections of the paving operation were monitored, from Station 342+00 to 392+00 and from Station 394+00 to 444+00. As with the Rochelle test site, an intersection in the paving job between Station 392+00 and 394+00 caused that portion of the project to be omitted from the monitored sections.

The Bellefontaine test site consisted of a 275-mm nominal thickness, plain jointed pavement, with dowel bars at the joints. The width of the pavement was 7 m for both sections monitored. Deflection testing on the granular base course layer was performed on September 13, 1995, for both sections. Concrete paving was monitored for both sections between October 11-19, 1995.

Deflection testing for the finished PCC surface was performed on both sections on May 1 and 2, 1996. Coring of both sections was performed on May 2 and 3, 1996, with all cores tested within the next week (approximately 6 months after placement).

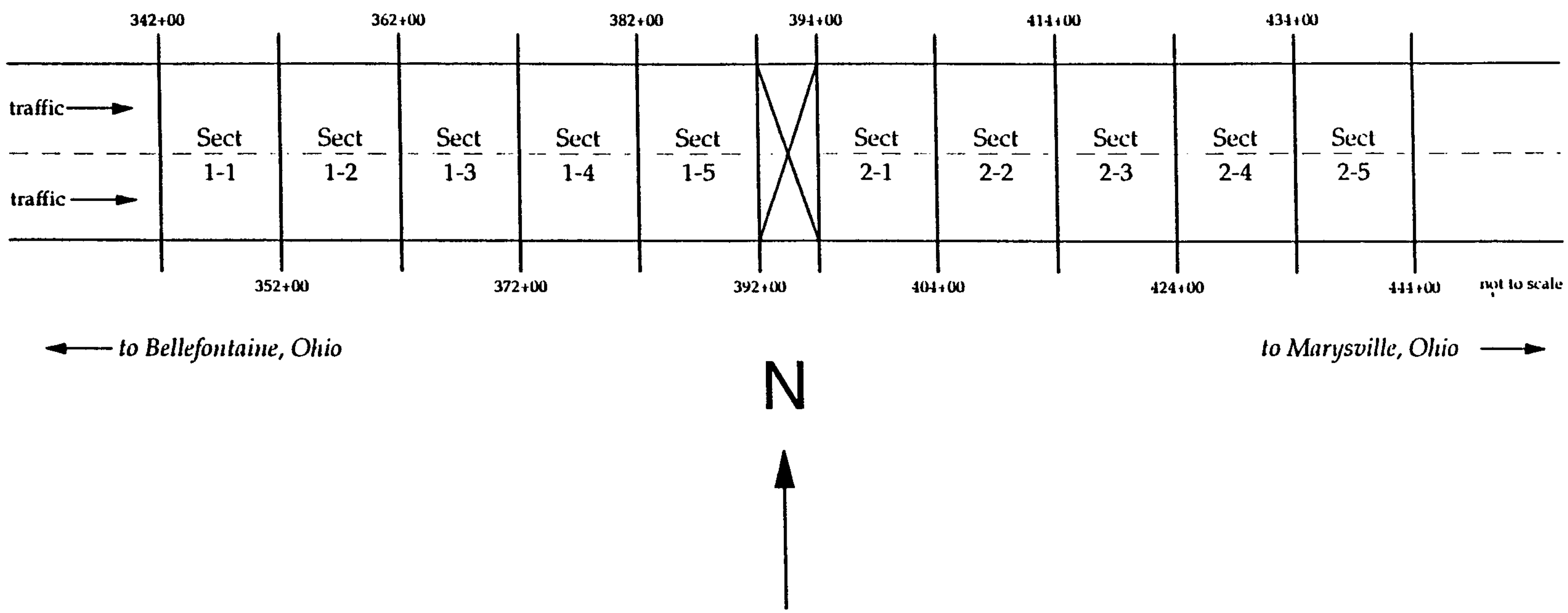


Figure 17. Location of PCC Test Sections, Bellefontaine, OH

CHAPTER 5. SAMPLING AND TESTING

As indicated under subtask B.3 of chapter 2, a detailed experimental plan was submitted to the FHWA. The experimental plan contained the testing and sampling frequencies to be conducted on each project. A summary of the approved sampling and testing plan that was followed appears below.

It was anticipated that, in addition to the testing performed by the researchers, contractor QC and SHA acceptance tests would be available for analysis. The State and study team laboratory test results for the PCC projects were combined for the analyses of the PCC data. The analyses were performed separately for the HMAC projects.

SAMPLING AND TESTING FOR HMAC PAVEMENTS

The following is a description of the sampling and testing plan for the flexible pavement test sites. Asphalt content, bulk specific gravity, MTSG, gradation, and air voids were identified as the mixture quality characteristics of interest while density, thickness, and deflection (strength) were the in-place quality characteristics of interest.

Test Sections (Lot Description)

Two replicate test sections 1600 m in length defined the lot size. These sections consisted of continuous paving areas with no bridges or other breaks in the paving operations. Each of the replicate test sections were further subdivided into five sublots, each 305 m in length. Within each sublot, four sample areas, each 76 m in length, were tested. Each of the four 76-m segments were further divided transversely into two sampling areas, representing the two wheelpaths.

Bulk HMAC Samples

Forty bulk HMAC samples were obtained at the plant from each of the three HMAC projects for testing of mixture quality characteristics including air voids, gradation, and asphalt content. It was anticipated that results from both the contractor QC and agency acceptance samples would be available. However, only agency data were available; these data were compared with those obtained by the contract laboratory in the data analysis. Samples were obtained from the hauling units at the plant. The hauling units were coordinated with the placement of material at the core locations within each sublot at the project site.

The 40 bulk samples were sent to Atlanta, Georgia, for testing. Each sample was heated in a convection oven set at 154°C for 45 to 75 min so the sample could then be separated into appropriate quantities for the various ASTM tests.[76,77,78] Nearly 24 kg of the approximate 55- to 66-kg quantity received for each sample was needed for the testing.

Bulk specific gravity testing followed ASTM D2726-90 procedure. A loose sample of approximately 1,220 g was split from each material source and placed in an oven at 163°C for approximately 2 h. The mix then was placed in a Marshall mold (nominal 102-mm diameter), spaded, and allowed to cool to approximately 135°C. The specimen then was subjected to 75 blows from a Marshall hammer on each of its two faces. The specimen was allowed to cool to room temperature prior to its extrusion from the mold. Bulk specific gravity tests were performed on each air-cooled specimen in a manner prescribed by the ASTM test procedure.[76]

MTSG was determined using ASTM D2041-91 procedure. An appropriate 2,000- to 2,500-g sample was used for this analysis. A 4,000-ml flask (pycnometer) was used for the test, and the water was maintained to within the specified temperature tolerances of the test procedure.[77] Once testing was completed, air voids were calculated.

Extractions were performed on the bulk samples to determine the asphalt content and aggregate gradations using ASTM D2172-92. An approximate 1,200-g sample was used for this test. Method E of this specification (vacuum extractor) was used as the device, and methylene chloride was used as the solvent.[78] Typically three to four washes of the solvent was all that was necessary to dissolve the asphalt from the aggregate. Asphalt contents are based on total mixture weight as opposed to dry aggregate weight.

Field Sampling and Testing

Field Sample Code Designation

For ease in interpreting the sample code designation, the following description is provided. Samples or tests performed on the samples from the list of quality characteristics were identified by a code number consisting of a letter set (cs for core sample, nu for nuclear density tests, etc.), sublot number (1-5), sample area (1-4), and transverse location (1-2). For example, a core sample taken from the first 1.6-km test section, second sublot, third sample area, and outside wheelpath would be identified as #12CS31. Similarly, a nuclear density test performed in the second 1.6-km test section, third sublot, second sample area, and inside wheelpath would be identified as #23NU22.

Density

Nuclear density testing was performed on the LA and OK2 projects. Core density tests were performed on the OK1 project. MTSG values were obtained from the bulk HMAC samples and were used for determination of percent compaction.

Thickness

On the OK1 project, the cores were used for thickness measurements. On all three projects, GPR was performed for thickness determination.

SAMPLING AND TESTING FOR PCC PAVEMENTS

The following sections describe the sampling and testing plan used for the PCC pavements. Air content, thickness, density, smoothness, and strength were identified as the quality characteristics of interest.

Two 1.6-km test sections were defined as lots. These sections consisted of continuous paving areas, with no bridges or other breaks in the paving operation. Each of the two replicate test sections was further subdivided into five sublots, each 305 m in length. Each slab was subdivided into six transverse units, with three units in each traffic lane.

Samples and test results for each of the quality characteristic tests were identified by a code number consisting of the replicate number (1 or 2), sublot number (1-5), slab number (01-66), and transverse position (1-6). For example, a core taken from the second replicate, third sublot, tenth slab, outside wheelpath would be identified by the number 23101.

Air Content

Air content testing for this project included both plastic concrete and hardened cores. For testing of plastic concrete, one slab was chosen at random for each 305-m sublot, and six air content samples were taken from across the pavement width. Additionally, three evenly spaced locations were chosen for air content testing within each sublot.

ASTM C231-91b was used for determining the air content of the plastic PCC. Samples of plastic concrete were taken from the paver at the point of placement so the samples could be correlated to a lot, sublot, and slab of the roadway.[79]

ASTM C642-90 was used to measure the air content of cores obtained from the hardened PCC. Cores to measure hardened air content and density were taken from a minimum of five random locations (one per each 305-m sublot).[80]

Air content variability was examined in the longitudinal and transverse directions using the plastic air contents determined by the contractor's QC testing, the SHA's QA testing, and the additional testing conducted according to the research test plan. Air content variability through the depth of the pavement was examined using the hardened air content results for the five cores chosen in each test section.

Thickness

All cores, including those taken by the contractor and by the SHA, were measured for thickness. These measurements were made prior to strength testing of the cores or sawing of the cores for hardened air content and density measurements. All thickness measurements were identified by the test section replicate (1 or 2), sublot (1-5), slab number, and transverse position (1-6). On all three projects, thickness also was determined by GPR.

Strength

Both compressive and flexural strength tests were performed as directed by ASTM C39-86 and ASTM C42-90, respectively.[81,82] Compressive strength was performed on both cast cylinders and cores taken from the pavement. Pairs of cylinders were cast for three evenly spaced locations within each sublot, and the cylinders were tested after 28 days. Testing of cores was performed at 60 days. Testing of compressive strength also was conducted by the contractor and the SHA.

Smoothness

For each test section, four profile lines were established, one in each of the wheelpaths.

Density

Density testing consisted of specific gravity values for the cores collected, unit weight of the material as it was tested for plastic air content, and unit weights of the disks created to evaluate density variability with depth. The density of the hardened concrete was performed as specified by ASTM C642-90.[80]

FIELD TESTING

Each of the six test sections was tested to determine the variability of the quality characteristics. Testing included laboratory testing of sampled materials, FWD testing, and GPR testing to determine thickness. Details of these tests and their results are given in volume II of this report.

FWD Testing

Deflection testing was conducted twice on each site. For the two sites in Oklahoma, this testing was completed on the prepared subgrade and then on the finished surface. Testing on the

Louisiana site was completed on the milled surface and then on the finished surface. Testing on the PCC sites was completed on the base course and then on the finished surface.

Testing was conducted in the center of the lane at 36-m intervals. Three passes were made along each lane, with the passes staggered longitudinally every 12 m. Two drops at 40.0 MN were used to seat the plate. Then the testing consisted of two drops of 26.7 MN and four drops at 40.0 MN. The sensors were spaced at 0, 0.2, 0.3, 0.5, 0.6, 0.9, and 1.5 m according to the SHRP protocol on sensor spacing.[83]

For the base testing of the PCC, three passes were made with tests conducted at 37-m intervals, staggered transversely from one side to the next. FWD testing was performed on the finished PCC pavement surface only in the center of the slabs, with testing of joint-load transfer. Two passes were made along each lane, with the passes staggered longitudinally so that mid-panel tests were performed throughout the length of the site. Joint-load transfer tests were performed every 305 m as well.

GPR Testing

The objective of this work was to use GPR as a means for characterizing as-built pavement layer thickness variations. Unlike coring, which can only obtain local thickness data, GPR can obtain continuous thickness data at any desired longitudinal spacing. The availability of this quantity of data can lead to statistical thickness representations that are truly indicative of the as-built construction.

GPR testing was used to obtain thickness data on each of the test sections. Data were collected at four distinct transverse locations across the pavement. These locations were 0.46 m, 1.37 m, 2.29 m, and 3.20 m from the shoulder of the pavement. At each of these transverse locations, data were taken approximately every 7.6 m longitudinally down the pavement. This method of data collection allows for a comparison of transverse variability. A total of approximately 800 measurements were taken on each lot of each project or approximately 200 measurements were taken on each lot at each of the four transverse locations, with a total of 1600 measurements for the entire project.

Ordinarily, GPR can detect surface layer thickness within 5 to 10 percent of actual values. With the availability of a small number of cores for calibration, the accuracy can be enhanced. The ability to accurately determine a large number of thickness data points has been applied to this project and is described in chapter 6 of this report.

CHAPTER 6. DATA ANALYSIS

INTRODUCTION

As noted in chapter 3, a number of States have published information on population distributions and variability related to the quality characteristics of interest to this project. Most of the SHA's view quality as conformance to design. Conformance to design means meeting the SHA specifications and allowable tolerances. The assumption is that meeting these specification tolerances implies that the pavement will perform as designed. The data were available from this study to evaluate conformance to the normal distribution theory and perform other statistical analyses that might provide insights into better measurements and specifications of quality characteristics of paving products.

It is the objective of QA specifications to govern those characteristics that are believed to be strongly related to the ultimate performance of the final product. In most cases, the *qualitative* relationships between commonly measured construction characteristics and performance have been documented even though the *quantitative* nature of these relationships may be somewhat vague.

Although various statistical measures of quality are available, most SHA's have exhibited a strong preference for the concept of lot PD, the estimated percentage of a lot falling outside specification limits (or its counterpart, the PWL).[84] This measure is particularly appealing for at least three reasons:

1. It can be applied to virtually any construction quality characteristic.
2. It encourages uniformity in that it controls both the average level and variability of the product in a statistically efficient way.
3. It is believed to be associated with good ultimate performance.

Acceptance Plans

PD can be controlled by either of two types of quality acceptance procedures: attributes plans or variables plans. Attributes plans typically involve the counting of some type of defect or the number of failing tests, and lead to the inspected lot being classified as either satisfactory or unsatisfactory. Variables plans apply to quality characteristics that are measured on a continuous scale and involve the computation of statistical parameters such as the mean and standard deviation. Either type of plan may be used for pass/fail decisions, although variables plans are somewhat more convenient as a basis for adjusted pay schedules.

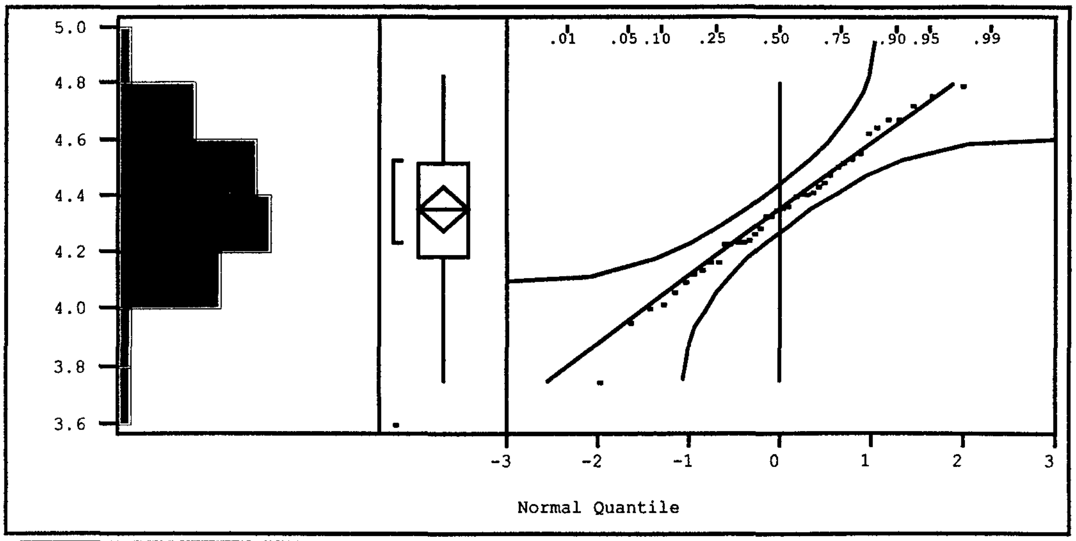
As a general rule, variables plans are more efficient than attributes plans. This means that for a given sample size, greater protection against risk is provided, or for a given level of protection, a smaller sampling effort is required. Either way, substantial economic benefits can be realized with the use of variables plans.

A basic assumption of variables acceptance theory in highway construction is that the population (lot) being sampled is normally distributed. As previously discussed in chapter 3, many construction characteristics have been found to closely approximate the normal distribution, thereby justifying the widespread use of variables procedures. When a situation occurs in which the construction characteristic is distinctly non-normal, there are two possible remedies. Either the individual tests can be replaced with the averages of two or more random tests, a step that greatly improves normality, or an attributes procedure that requires no distributional assumptions can be used.

TESTS FOR NORMALITY

The data collected from the six projects by the study team were used to study many of the statistical characteristics of the data, including the normality of quality characteristics. Several methods were used to make these determinations. A typical statistical characterization used in this study is shown in figure 18 for one property on one project. This property is asphalt content from the Louisiana project. The entire collection of data for all properties from all projects is

AC Content



Quantiles

maximum	100.0%	4.8000
	99.5%	4.8000
	97.5%	4.7990
	90.0%	4.6800
quartile	75.0%	4.5175
median	50.0%	4.3550
quartile	25.0%	4.1850
	10.0%	4.0240
	2.5%	3.7553
	0.5%	3.7500
minimum	0.0%	3.7500

Moments

Mean	4.35175
Std Dev	0.23752
Std Error Mean	0.03755
Upper 95% Mean	4.42771
Lower 95% Mean	4.27579
N	40.00000
Sum Weights	40.00000
Sum	174.07000
Variance	0.05641
Skewness	-0.15279
Kurtosis	-0.14422
CV	5.45799

Test for Normality

Shapiro-Wilk W Test

W	Prob<W
0.985393	0.9240

Figure 18. Typical Statistical Characterization of One Quality Property from One Project

contained in volume II of this report. The following data are included in each statistical characterization:

- A histogram.
- A normal probability plot.
- A table of quantile statistics that includes minimum, 25-percent quartile, median, 75-percent quartile, and maximum.
- A table of statistical moments that includes the mean, standard deviation, variance, standard error of the mean, upper and lower 95-percent confidence limits of the mean, the sample size, skewness, kurtosis, and the COV.
- The Shapiro-Wilk test for normality.

Histograms

Visual observations of the histograms show the difficulty of subjectively judging normality. Several of the histograms appear to be skewed, e.g., for HMAC, air voids from the first Oklahoma project and for PCC, GPR thickness from the Ohio project. Some appear bimodal, e.g., for HMAC, the 10-mm sieve from the Louisiana project and for PCC, to a lesser extent, core compressive strength from Illinois. Other histograms contain apparent outliers, e.g., for HMAC, percent of MTSG from the second Oklahoma project and for PCC, profilograph measurement from the Ohio project. Thus, it is apparent from visual, subjective judgment that more scientific tests of normality are warranted.

Observations of the normality of a distribution from normal probability plots is one attempt to remove the subjective judgment. On this plot, the data values appear on the ordinate axis in a linear (equally spaced) scale and the abscissa axis contains the cumulative probabilities. If the data are from a normal distribution, this plot will produce a straight line. Viewing the plots in volume II indicates that many of the properties appear to be from normal distributions, but again evidence of non-normal distributions are apparent, especially for the properties in the examples mentioned above.

Shapiro-Wilk Test, Skewness, and Kurtosis

More scientific evaluations of normality are found in the Shapiro-Wilk statistic, and skewness and kurtosis values. The Shapiro-Wilk statistic, W, is the ratio of the best estimator of the variance to the usual corrected sum of squares estimator. W must be between 0 and 1, with large values indicating that the data are normally distributed.[85] This test is very sensitive and has been known to indicate that data are not normally distributed when they are from a known normal distribution. As shown in tables 8 and 9 for HMAC and PCC and the discussion of the values of the statistical tests, all Shapiro-Wilk values are above 0.68, with a large majority being 0.90 or greater.

Skewness is another measure of non-normality. The standard deviation is a measure of the overall magnitude of the deviations from the mean. But the standard deviation does not indicate if the data have larger positive deviations or larger negative deviations. Skewness is a measure of the tendency of the deviations to be larger in one direction than in the other. Skewness values that have a large absolute value are likely to be from a non-normal distribution.[85] This statistical characteristic also is an indication that many of the properties analyzed are approximately normally distributed since most values of skewness are less than ±1.0. The equation for skewness is shown below:

$$\text{Skewness} = \Sigma \frac{(x_i - \bar{x})^3}{s^3} \times \frac{n}{(n-1)(n-2)} \tag{6}$$

where:

x_i = ith observation of distribution

$\bar{x}$ = sample mean

s = sample standard deviation

n = number of samples.

Table 8. Shapiro-Wilk (W), Skewness, and Kurtosis Results for HMAC Projects

Project	Property	n	W	Skewness	Kurtosis
LA	BSG	40	0.981	0.101	-0.627
OK1		40	0.931	-0.740	0.133
OK2		40	0.968	-0.405	-0.231
LA	TSG	40	0.977	0.184	0.037
OK1		40	0.978	-0.129	0.698
OK2		40	**0.766**	**2.865**	**12.474**
LA	AV	40	0.955	-0.160	-0.721
OK1		40	0.930	**0.979**	1.331
OK2		40	0.979	0.297	-0.220
LA	AC	40	0.985	-0.153	-0.144
OK1		40	0.952	-0.750	0.927
OK2		40	0.982	0.022	0.130
LA	19-mm	40	**0.794**	**-0.942**	0.818
LA	13-mm	40	0.967	0.085	-0.366
OK1		40	0.808	-0.515	0.557
OK2		40	0.878	-0.578	-0.326
LA	10-mm	40	0.950	-0.013	-0.759
OK1		40	0.823	**-1.708**	**3.902**
OK2		40	0.917	-0.819	0.373
LA	5-mm	40	0.963	0.432	0.788
OK1		40	0.928	**-1.061**	**3.074**
OK2		40	0.973	-0.377	-0.033
LA	2.36-mm	40	0.877	**1.169**	**4.103**
OK1		40	0.944	-0.762	1.377
OK2		40	0.975	-0.233	0.024
LA	1.18-mm	40	0.915	0.611	0.966
OK1		40	0.922	-0.643	0.158
OK2		40	0.950	0.127	0.063
LA	0.30-mm	40	0.838	0.842	1.546
OK1		40	**0.757**	-0.612	-0.722
OK2		40	0.902	0.210	-0.758
LA	0.15-mm	40	**0.716**	**1.711**	**5.219**
OK1		40	**0.796**	**1.432**	**7.127**
OK2		40	0.916	0.262	-0.707
LA	0.075-mm	40	**0.685**	**3.630**	**18.235**
OK1		40	0.939	-0.647	-0.338
OK2		40	0.949	0.054	-0.552
LA	GPR	1600	0.985	-0.208	0.279
OK1	Thickness	1512	0.879	**1.690**	**5.016**
OK2		1592	0.967	-0.555	**2.971**
LA	Density,	40	0.893	-0.072	-1.533
OK1	% Max, LWP	18	0.955	-0.152	-1.139
OK2		19	0.957	-0.166	0.021
LA	Density,	40	0.893	0.029	-1.537
OK1	% Max, RWP	18	0.946	0.258	-0.916
OK2		19	0.881	**-1.279**	**3.722**
LA	FWD,	167	0.914	**1.088**	1.5057
OK1	Deflection	252	0.929	**1.145**	**1.945**
OK2		254	0.967	0.530	0.123
LA	FWD,	167	0.946	0.619	0.028
OK1	Modulus	252	0.949	**0.959**	1.833
OK2		254	0.917	0.274	-0.921

Note: Shapiro-Wilk values less than 0.8, skewness values greater than ±1.0, or kurtosis values greater than ±2.0 are bold to indicate higher levels of skewness.

Table 9. Shapiro-Wilk (W), Skewness, and Kurtosis Results for PCC Projects

Project	Property	n	W	Skewness	Kurtosis
IL	Plastic Air	43	0.978	0.221	-0.235
MN		29	0.977	0.053	-0.384
OH		115	0.970	-0.255	-0.201
IL	Unit Weight	30	0.965	-0.261	-0.860
MN		10	0.958	0.488	-0.605
IL	Cylinder	106	0.967	-0.200	-0.486
MN	Compressive	56	0.957	0.377	-0.549
OH	Strength	100	0.971	-0.194	-0.687
IL	Core	30	0.971	-0.267	-0.727
MN	Compressive	30	0.909	**-1.208**	1.796
OH	Strength	31	0.950	-0.379	-0.827
IL	Core	30	0.927	-0.574	-0.628
MN	Thickness	41	0.954	0.020	-0.593
OH		30	0.942	0.512	0.916
IL	GPR	1450	0.814	**1.787**	**9.368**
MN	Thickness	1187	0.901	0.014	**3.593**
OH		1188	0.928	**-1.149**	**4.658**
IL	FWD, mm	302	0.915	**1.010**	**0.839**
MN		369	0.841	**1.835**	**4.456**
OH		472	0.949	**0.464**	0.306
OH	Profile	73	0.813	**2.108**	**7.057**

Likewise, kurtosis measures the "heaviness" of the tails of a distribution.[85] A large value of kurtosis indicates a heavy-tailed distribution. Once again, the kurtosis values are usually less than ±1.0. The population kurtosis equation is shown below:

$$\text{Kurtosis} = \Sigma \frac{(x_i - \bar{x})^4}{s^4} \times \frac{n(n+1)}{(n-1)(n-2)(n-3)} - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (7)$$

where:

x_i = ith observation of distribution

$\bar{x}$ = population mean

s = population standard deviation

n = number of samples.

Somewhat arbitrarily, a Shapiro-Wilk test value less than 0.8 was chosen as an indication of non-normality. But according to statistical textbooks, for a sample size of 40 for the HMAC data, skewness values greater than ±0.9 and kurtosis values greater than +1.9 are indications of non-normality at an α value of 1.0 percent.[86] For the PCC data that had sample sizes from 10 to 1450, the critical skewness values vary from ±1.3 to ±0.2, and critical kurtosis values vary from greater than 2.0 to 0.3 for an α value of 1.0 percent. These values are in bold in tables 8 and 9.

HMAC Projects

For the HMAC projects, of the 52 properties measured on the three projects, six (11.5 percent) Shapiro-Wilk values were less than 0.8. Five of the six were from gradation results. Two were from the Louisiana project; one for the top size sieve (19-mm) and one from the finest sieve (0.075-mm). In the case of the 19-mm sieve, this indication of non-normality is typical when the mean value is close to a physical barrier. Specifically in this case, the mean value is 99.2 percent and the maximum percent passing the 19-mm sieve is 100 percent, so values can only be below 100 percent. Therefore, any appreciable deviation must be below the mean creating the skewness. However, the other properties that have W values less than 0.8 are not as easily explained. The presence of outliers is another source of skewness. For the 0.15-mm sieve on the first Oklahoma and Louisiana projects, the relatively low W values appear to be the result of apparent outliers. Outliers also appear to be the cause of the low W value for the theoretical specific gravity from the Louisiana project.

Again for the HMAC projects, of 52 properties measured, skewness values greater than ±1.0 occurred for 44 (27.0 percent) properties. Five of these were the same properties that had relatively low W values. Seven were from gradation measurements of either the Louisiana or first Oklahoma project. Skewness can be relatively easily visualized from the histograms but kurtosis cannot be. For these projects, kurtosis values exceeding 1.9 occurred for 11 (21.2 percent) properties.

PCC Projects

For the PCC projects, of the 21 properties measured on the three projects, none of the Shapiro-Wilk values were less than 0.8. Skewness values exceeded the critical value for seven (33.3 percent) properties, two were from GPR thickness measurements, three from falling weight deflectometer measurements, one from profile, and one from core compressive strength results. Kurtosis values exceeded the critical value for six properties (28.6 percent), five were the same properties that exceeded the critical skewness value.

SKEWBIAS Program

Although the properties measured in this study tend to be normally distributed, it is of interest to determine to what extent the normality assumption can be violated without significantly impairing the effectiveness of the PD estimation procedure. To do this a computer program was developed that enables the user to simulate variables sampling under a wide variety of conditions. The development and use of the program is discussed first, followed by the applicability to the test data.

Program SKEWBIAS is written in Microsoft QuickBASIC for a DOS operating system on an IBM compatible PC. For best results, it should be run on a machine using an Intel compatible 80386 (or higher) processor. A monochrome monitor is sufficient, but considerably greater clarity is achieved on a color monitor.

Program SKEWBIAS is menu driven, requires very little input for each run, and is virtually self-explanatory. It may be run from a disk drive or can be loaded directly onto a hard drive. In either case, support program TABLEPD.FIL must be present on the drive from which the program is run. TABLEPD.FIL is a standard table of PD estimates associated with variables acceptance sampling and is part of the program provided.

SKEWBIAS uses computer simulation to determine the bias of PD estimates associated with sampling from a skewed normal population with skew coefficients ranging from -10 to +10 and sample sizes from n = 3 to n = 30. This is accomplished by generating a perfect normal distribution, transforming it into a population with a known skew coefficient and level of PD, repeatedly sampling the skewed population using the desired sample size, computing the estimated PD, and noting the degree of bias of each estimate. After many such samplings, a reasonably precise estimate of the bias can be obtained. *In this program, the bias is expressed as the estimated PD value minus the true population PD in each case.*

For example, suppose it were desired to determine the bias for the case in which n = 5 samples are taken from an approximately normal population having a positive skew coefficient of 1.0 and a true percent defective level of 10.0 in the lower tail. To accomplish this by computer simulation, first it is necessary to create a normal population. Program SKEWBIAS uses a table of standard normal variates to create a population of 1000 normal numbers with a mean of μ = 0.0 and a standard deviation of σ = 1.0. For this example, a population percent defective of PD = 10.0 in the lower tail is desired and the standard normal value that cuts off exactly this area is -1.282.

One method for transforming the normal population into a skewed population is to raise each value to an appropriate power. To use this method, first it is necessary that each value in the population be greater than or equal to zero. This is done by adding the same constant value to each population value which, in effect, just slides the population to the right along the axis of real numbers without changing either its shape or its PD level.

In program SKEWBIAS, the constant that is added to each value is 10.0. Therefore, the new mean value is μ = 0.0 + 10.0 = 10.0, the standard deviation remains unchanged at σ = 1.0, and the limit that cuts off 10 PD in the lower tail is now -1.282 + 10.0 = 8.718.

Next, it is found by trial and error that raising each population value to the power of 4.38 produces a skew coefficient of exactly 1.0. The new limit that cuts off PD = 10.0 in the lower tail is $8.718^{4.38} = 1.315 \times 10^{4}$. The mean and standard deviation of this transformed distribution are no longer of concern because the objective simply was to produce a distribution of known skewness and percent defective. As an aid to visualization, SKEWBIAS displays a histogram of the transformed population before completing the analysis.

It should be noted that this is not a unique solution and that other similar transformations can be used to produce the desired skewed-normal population. During the development of SKEWBIAS, other transformations were tested to check if there were any appreciable effects on the bias determination. The bias values produced by the program under these different conditions were all found to be in close agreement.

Once the desired skewed population has been created, the next step is to sample it, compute the sample mean ($\bar{x}$) and standard deviation (s), calculate the quality index ($Q = (\bar{x} - L) / S$), and obtain the PD estimate from the appropriate table (accessed as a subrouting by SKEWBIAS). No single estimate such as this is sufficiently reliable to judge the degree of bias, so this process must be repeated many times, usually 1000, or more. The average of many such estimates is then compared with the true population PD to provide a reliable estimate of the bias.

For example, if the average of 1000 PD estimates was 11.62 with a standard deviation of 3.16, and the true population PD is 10.0, then the bias would be estimated as 11.62 - 10.0 = +1.62 PD units with a standard error of $3.16/\sqrt{1000} = 0.10$. With each run of SKEWBIAS, the average bias and its standard error are displayed for PD levels of 1, 5, 10, 20, ..., 80, 90, 95, 99.

The opening screen provides basic operating information, including the use of the <ESC> key to repeat certain steps and the <END> key to exit the program. The input screen appears next, and the user is prompted to enter a transformation exponent (to obtain the desired degree of skewness of the population to be sampled), the sample size, and the number of replications to be used at each quality level.

After the number of replications is entered, the program proceeds to create the skewed distribution. Because it may be useful to have a visual impression of the degree of skewness that has been selected, the program next displays a histogram of the distribution to be sampled. A typical display is shown in figure 19.

At this point, the user can either press the <ESC> key to return to the input menu, the <END> key to exit the program, or any other key to proceed with the bias computations. The final display screen appears next and, depending on the speed of the computer and the number of replications selected, it may require anywhere from a few seconds to a few minutes to compute the bias values. A typical output display is presented in table 10.

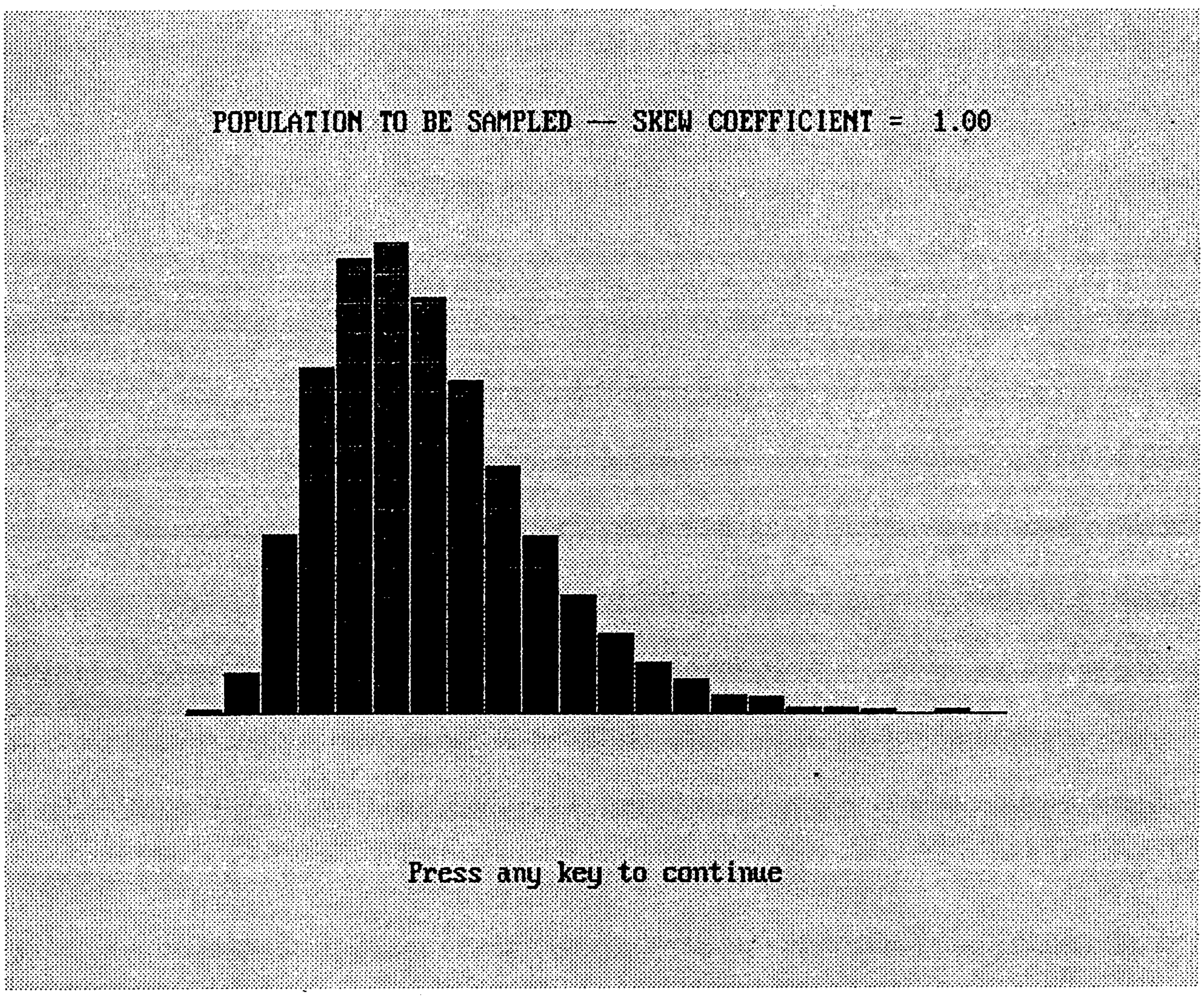


Figure 19. Typical Display of Skewed Distribution by Program SKEWBIAS

Table 10. Typical Output of Program SKEWBIAS

True PD	Bias of PD Estimate (Standard Error)	
	Positive Skew	Negative Skew
1	2.27 (0.05)	-0.41 (0.03)
5	2.34 (0.08)	-0.72 (0.08)
10	1.62 (0.10)	-0.19 (0.13)
20	-0.41 (0.13)	0.93 (0.18)
30	-1.98 (0.16)	1.88 (0.20)
40	-2.80 (0.17)	2.86 (0.20)
50	-3.24 (0.19)	3.24 (0.19)
60	-2.86 (0.20)	2.80 (0.17)
70	-1.88 (0.20)	1.98 (0.16)
80	-0.93 (0.18)	0.41 (0.13)
90	0.19 (0.13)	-1.62 (0.10)
95	0.72 (0.08)	-2.34 (0.08)
99	0.41 (0.03)	-2.27 (0.05)

Notes: Population Size = 1,000 Number of Replications = 10,000
Sample Size = 5 Skew Coefficient = ±1.00

Table 10 provides the expected bias when the sample size is N = 5 and the population being sampled has a skew coefficient of ±1.00. The column for a positive skew coefficient indicates that the estimated defective portion is in the shortened tail and, when the skew coefficient is negative, it is in the elongated tail.

There are three interesting observations to be made from the data in this table. First, the degree of bias cycles back and forth between positive and negative values as the PD increases, with the greatest value occurring at PD = 50. This was found to be generally true and is a consequence of the difference in the shape of the tails of the assumed normal population and the actual skewed population. Second, the standard errors (in parentheses) are all fairly small, indicating that the bias values are determined reasonably precisely when 10,000 replications are used. Third, the bias values for a specific value of PD and positive skewness is the same for the complementary value (100 - PD) and negative skewness, but with the opposite algebraic sign, as would be expected. (If it were desired to have a table giving bias values for PWL estimates, the first column heading would be changed to "True PWL" and the headings "Positive Skew" and "Negative Skew" would be reversed).

To see how this table is used, if a sample of size N = 5 were to be taken from a population having a skew coefficient of approximately 1.00 and a true PD in the shortened tail (positive skew in table 10) of PD = 10, there would be a tendency to overestimate the true PD value by about 1.6 percent, making the estimated PD = 11.6. Furthermore, the standard error of 0.10 indicates that the estimated bias value (11.6 PD) is accurate to within ±0.2 PD, assuming two standard deviation limits. Similarly, if defective material were in the elongated tail (negative skew in table 10), there would be a tendency to underestimate the true PD value by about 0.2 percent, making the estimated PD = 9.8. In this case, the standard error of 0.13 indicates that the bias value in the table (-0.19) would be between positive 0.07 and negative 0.45, assuming two standard deviation limits. Since this range includes zero, it means there may be no bias in this particular case.

In these two examples at least, bias does not appear to be a serious problem. For quality levels in the range that is generally regarded as acceptable, such as PD = 10, there is a tendency for the quality estimates to be in error by between about 0.2 to 1.6 PD, depending in which tail of the skewed population the defective material lies. For considerably lower levels of quality, such as PD = 50, the error will not be much larger than about 3.2 PD. In most characteristics measured in this study, the populations to which variables sampling procedures are applied will have skew coefficients less than the value of 1.00 assumed for these examples, in which case the amount of bias will be less.

Results of SKEWBIAS Program

To create a concise table that provides the values of the expected bias over a wide range of conditions that might be encountered, program SKEWBIAS was run many times, and the results are presented in table 11. For most typical highway construction sampling applications, the expected amount of bias can either be read or interpolated from this table. For conditions outside the range of the table, or for more precise determinations within the range of the table, the computer program itself can be used.

In table 11, it is interesting to note that the bias associated with sampling from any given skewed distribution increases as N increases. This should not be too surprising. "The more samples one takes, the better the estimate of quality (i.e., PD)" is a well-known principle; but, for it to be true, the normality assumption must be valid.

To judge the effect in an actual specification application, both the correct PD (or PWL) value and the expected biased estimate of PD (or PWL) must be entered into the acceptance procedure, and the difference in the acceptance decision (or pay factor) noted at different levels of PD (or PWL). In this way, it will be possible to judge at what degree of skewness the operation of the acceptance procedure might be adversely affected.

LOT SIZE AND BIAS RESULTS

Precision, Accuracy, and Bias

The terms precision and accuracy are often confused. Precision is the degree to which tests or measurements on identical samples tend to produce the same result.[3] Accuracy is the degree to which a measurement, or the mean of a distribution of measurements, tends to coincide with the true population mean.[3] Bias is an error, constant in one direction, that causes a measurement, or the mean of a distribution of measurements, to be offset from the true population mean.[3] Good accuracy, then, can be considered the lack of bias.

Table 11. Bias of Single-Tailed Percent Defective (PD) Estimates for Selected Values of Sample Size and Skew Coefficient

True PD	Bias for Selected Skew Values and Sample Size N = 5								
	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
1	-0.4	-0.4	-0.4	-0.4	0.0	0.8	2.3	4.3	6.4
5	-0.6	-0.8	-0.7	-0.5	0.0	0.8	2.3	3.8	5.4
10	-0.1	-0.2	-0.2	-0.4	0.0	0.6	1.6	2.5	3.7
20	2.0	1.6	0.9	0.4	0.0	-0.2	-0.4	-0.1	0.2
30	4.1	3.5	1.9	1.1	0.0	-1.0	-2.0	-2.4	-3.0
40	5.3	4.4	2.9	1.5	0.0	-1.5	-2.8	-3.8	-4.8
50	5.4	4.4	3.2	1.6	0.0	-1.6	-3.2	-4.4	-5.4
60	4.8	3.8	2.8	1.5	0.0	-1.5	-2.9	-4.4	-5.3
70	3.0	2.4	2.0	1.0	0.0	-1.1	-1.9	-3.5	-4.1
80	-0.2	0.1	0.4	0.2	0.0	-0.4	-0.9	-1.6	-2.0
90	-3.7	-2.5	-1.6	-0.6	0.0	0.4	0.2	0.2	0.1
95	-5.4	-3.8	-2.3	-0.8	0.0	0.5	0.7	0.8	0.6
99	-6.4	-4.3	-2.3	-0.8	0.0	0.4	0.4	0.4	0.4
True PD	Bias for Selected Skew Values and Sample Size N= 10								
	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
1	-0.6	-0.6	-0.6	-0.5	0.0	1.2	3.2	5.8	8.2
5	-1.2	-1.2	-1.1	-0.8	0.0	1.4	3.2	5.3	7.2
10	0.1	-0.4	-0.6	-0.5	0.0	0.9	2.0	3.6	5.1
20	4.1	2.4	1.7	0.6	0.0	-0.4	-0.3	0.0	0.4
30	6.8	5.1	3.4	1.6	0.0	-1.4	-2.4	-3.3	-3.6
40	8.2	6.4	4.6	2.2	0.0	-2.0	-4.0	-5.4	-6.4
50	8.2	6.5	4.8	2.3	0.0	-2.3	-4.8	-6.5	-8.2
60	6.4	5.4	4.0	2.0	0.0	-2.2	-4.6	-6.4	-8.2
70	3.6	3.3	2.4	1.4	0.0	-1.6	-3.4	-5.1	-6.8
80	-0.4	0.0	0.3	0.4	0.0	-0.6	-1.7	-2.4	-4.1
90	-5.1	-3.6	-2.0	-0.9	0.0	0.5	0.6	0.4	-0.1
95	-7.2	-5.3	-3.2	-1.4	0.0	0.8	1.1	1.2	1.2
99	-8.2	-5.8	-3.2	-1.2	0.0	0.5	0.6	0.6	0.6
True PD	Bias for Selected Skew Values and Sample Size N = 15								
	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
1	-0.8	-0.8	-0.7	-0.5	0.0	1.3	3.6	6.4	9.1
5	-1.4	-1.5	-1.5	-0.9	0.0	1.5	3.6	5.8	7.9
10	0.1	-0.4	-0.7	-0.6	0.0	1.0	2.4	4.0	5.6
20	4.5	3.3	1.8	0.8	0.0	-0.5	-0.2	0.1	0.8
30	8.2	6.0	4.0	1.9	0.0	-1.6	-2.6	-3.2	-3.5
40	9.4	7.5	5.1	2.6	0.0	-2.4	-4.5	-5.9	-6.9
50	9.1	7.4	5.3	2.7	0.0	-2.7	-5.3	-7.4	-9.1
60	6.9	5.9	4.5	2.4	0.0	-2.6	-5.1	-7.5	-9.4
70	3.5	3.2	2.6	1.6	0.0	-1.9	-4.0	-6.0	-8.2
80	-0.8	-0.1	0.2	0.5	0.0	-0.8	-1.8	-3.3	-4.5
90	-5.6	-4.0	-2.4	-1.0	0.0	0.6	0.7	0.4	-0.1
95	-7.9	-5.8	-3.6	-1.5	0.0	0.9	1.5	1.5	1.4
99	-9.1	-6.4	-3.6	-1.3	0.0	0.5	0.7	0.8	0.8

Note: Bias values are in units of PD and were obtained by random sampling from transformed standard normal populations using 10,000 replications for each table value. For positive skew, the defective portion is in the shortened tail, and, for negative skew, it is in the elongated tail.

Selection of Lot Size

The selection of lot size is dictated primarily by practicality and convenience. Some agencies prefer a lot size based on a day's production with the belief that this time period best defines production homogeneity. Other agencies prefer a constant tonnage or area with the belief that either provides a more constant sample size in case of production disruptions. In either case, for variables acceptance procedures, care must be exercised in combining work produced at different times or under different conditions, since this might violate the assumption of normality, as discussed in chapter 3 concerning multimodal distributions. From a statistical viewpoint of risks, larger lot sizes better accommodate larger sample sizes, as discussed below. A procedure has been proposed in this study to add to lot size when a sublot or lot is estimated to be normal or nearly normal. Typically, either time or quantity limits are used to define lots, such as a day's production, 2,268 metric tons or 4,200 square meters.[84]

Selection of Sample Size

The sample size (i.e., number of samples) to be used for acceptance purposes is a more important consideration than lot size because it has a direct effect on the risks that are involved and the resulting operating characteristic (OC) curve. Except for attributes sampling from discrete lots (items that are counted), the lot size plays no role in the development of the OC curve. Usually, but not always, larger sample sizes reduce the risks to both the contractor and the SHA, but to be sure the plan will perform as desired, the OC curves should be constructed for all sample sizes under consideration. Typical sample sizes of one to five per lot have been reported recently.[84] A sample of size one would not be recommended because no measure of variability is possible.

To detect smaller levels of bias, it is necessary to take larger numbers of samples. The first step in determining the precision and bias values for the collected data in this study was to determine the mean, standard deviation, and 95-percent confidence interval for each of the individual sublots within a project. The determination of the required sample size depends on several factors, as explained below.

- In a project, if the variances for the sublots are unequal, then the precision and bias values are developed on the basis of the sublot with the largest variance, and the sample size required must be based on sublots. If a minimum level of precision is selected, the number of samples required per sublot to determine that level of precision can be determined.

- If the sublot variances are equal, then these variances are pooled to obtain the variance estimate for precision calculations for that project and lot, and the lot means are tested.

  - If the sublot means are unequal, then precision estimates should be determined for each sublot, and the recommended sample size applies by sublot.

  - If the sublot means are equal, but the lot means are unequal, then precision estimates should be determined for each lot, and the recommended sample size applies by lot.

  - If lot means are equal, but project means are unequal, the precision estimates are determined for each project, pooling lot means.

  - If lot means are equal and project means are equal, then a single precision estimate is developed by pooling all projects and the sample size is on a project basis.

However, in a study such as this in which material and construction properties are compared for different States and each State has different requirements, it is unlikely that project means will be equal for any property. So it can be anticipated that, for this study, the sample size will always be project-specific from a statistical viewpoint. It is possible that, because the variance estimates are so small, from an engineering viewpoint a single sample size will accommodate most any level of bias that is selected.

From the viewpoint of an agency determining the proper number of samples to take from a lot to determine compliance with a specification, the application of sample size most likely will be to determine a specific sample size for a material or construction property and designate it in the specification to avoid having to determine a different sample size for each contractor or project. In this case a conservative sample size will mean that more sampling than necessary will be required for a contractor or project with a smaller standard deviation than that assumed when establishing the sample size for the specification.

The sample size necessary to discern a specific bias value is calculated using the following formula:

$$n = \frac{\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{(\mu_2 - \mu_1)^2} \qquad \textbf{(8)}$$

where:

- n = sample size
- σ^2= variance of sublot or lot as appropriate
- z = normal distribution statistic associated with α and β
- μ = mean of lot or project as appropriate
- α = level of significance, 5 percent
- β = probability of false acceptance, 20 percent

In this study, the range of the required sample size was tabulated using an α level of significance of 5 percent and an 80-percent power of the test for each project. The power of the test is defined as the probability of rejecting the hypothesis when it is false, $1 - \beta$, with β being the probability of a false acceptance, in this case 20 percent.[87]

The formula used in this analysis for determining precision is:

$$l_{0.95} = \pm z_{0.05}\sigma \tag{9}$$

where:

- $l_{0.95}$ = precision limits for 95-percent confidence
- $z_{0.05}$ = standard deviate at 5 percent
- σ = standard deviation of lot or sublot as appropriate

It should be noted that the above equation assumes a normal distribution and its associated statistics. If the variances of sublots are found to be unequal, the "z" value is replaced by a "t" value with the appropriate degrees of freedom, and σ is replaced by s, the sample standard deviation.[87]

HMAC RESULTS

The mixture and construction quality characteristics of the three HMAC projects are discussed in this section.

HMAC Mixture Result Analysis

The mixture characteristics measured were bulk specific gravity, MTSG, air voids (75-blow Marshall compactive effort), asphalt content, and gradation from the 19-mm sieve to the 0.075-mm sieve. Each project was divided into two lots, each containing five sublots, and a sample size of four within each sublot. The test results of these 40 samples per project comprise the primary analysis in this section. As will be noted where appropriate under each mixture characteristic, the number of State samples are mentioned. The data from the State samples are compared with those of the study team where possible. It should be noted that the number of State samples was generally small and in several cases nonexistent.

Each characteristic is discussed separately. The summary and test statistics used for determining the equality of variance tests among sample areas are tabulated first. Numbers in parenthesis under project are the target values established from the JMF.

The variabilities as measured by the standard deviations in the tables represent an overall variability caused by several individual sources of variability including those resulting from material variation, plant production process, sampling, and testing. In an attempt to find a measure of comparison of the standard deviations found in this study and other referenced standard deviations, three reference sources were used. One is the standard deviation from the appropriate ASTM method, where available, and another from the Asphalt Material Reference Laboratory (AMRL) data.[88] Since neither of these references contain the material or plant production process variabilities, the reference values should be less than those found in this study. Nonetheless, they can be used to put into perspective the values found in this study. The third source is NCHRP Synthesis 232 "Variability in Highway Pavement Construction."[6]

Analysis of variance was another statistical test used. A necessary assumption in order to perform the analysis of variance tests among means is that the sample areas represent the same population with homogeneous variables. Thus, the first step is to conduct a test of equality of variances among samples within projects, lots, and sublots. The Scheffe test, which compares the ratio of the maximum to minimum variances, was used to analyze variances among sublots within a lot with an F-distribution having (n_l) and (n_s) degrees of freedom where n_l is the sample size for the larger variance and n_s is the sample size for the smaller variance. The ratio of the variances is presented in the table for each property and an * indicates that the variances were not equal at the 5-percent level of significance. Upper and lower 95-percent confidence limits are labeled UL and LL, respectively, and are based on the lot mean ±2 standard deviations of the mean. The target values for each project are in parentheses. Note that if a lot has failed the equality of variance test among sublots, these confidence intervals are invalid, from an analysis viewpoint, and separate confidence limits for each sublot should be determined using the individual sublot variances. From a practical implementation viewpoint this may not be practical. Thus, a confidence interval based on a conservative estimate of the sublot variance may be necessary. When the sublot variances were equal, the sublot means were equal for each of the HMAC quality characteristics studies.

The second analysis is the number of samples necessary to attain a specific bias value. As discussed previously, the range of the required sample size is tabulated using an α level of significance of 5 percent and an 80-percent power of the test for each project.

Bulk Specific Gravity

Table 12 shows the sublot variances for bulk specific gravity were not equal at the 5-percent level of significance at the Oklahoma site 1 (OK1) but were equal at the Louisiana site (LA) and Oklahoma site 2 (OK2). Therefore, the sampling plan and bias and precision calculations should be project-specific for OK1.

Table 12. Descriptive Statistics for Bulk Specific Gravity

Project	Lot	Means	Std. Dev.	UL	LL	Var. Test
LA	1	2.375	0.0113	2.380	2.370	4.605
(2.412)	2	2.387	0.0126	2.392	2.381	6.900
OK1	1	2.399	0.0262	2.410	2.387	5.816
(2.375)	2	2.408	0.0215	2.417	2.399	13.661*
OK2	1	2.347	0.0218	2.357	2.338	6.823
(2.375)	2	2.351	0.0148	2.357	2.344	2.693

The standard deviations found on the LA project and on lot 2 for the OK2 project are less than that contained in ASTM D2726-91 "Standard Test Method for Bulk Specific and Density of Compacted Bituminous Mixtures Using Saturated Surface-Dry Specimens," which has a single-operator standard deviation of 0.0124.[76] The standard deviation measured from the AMRL proficiency sample data base is 0.0177, which also is larger than those found on the lots mentioned above.[88] The magnitude of standard deviations found on these projects could be used as a basis of specification development, if desired.

Table 13 lists ranges of required sample size for a range of bias values for each project. *Bias in this section, compared with the previous section on skewness evaluation, is defined as the difference between the lot or project average and the target.* The target is not necessarily the centerline of the specification limits. However, the target is the value provided by the State for the JMF. In other words, this value is the one the contractor is trying to produce. For OK1, the maximum sublot standard deviation was used. For LA and OK2, lot means were equal with no significant interaction; this indicates, from a statistical analysis viewpoint, that the project would not have to be divided into sublots and that a single random sample of size two or three on the entire project would be sufficient to estimate bulk specific gravity with biases of -0.031 for LA and -0.026 for OK2. If the detection of smaller biases is desired, a larger sample size would be required. Also, from an engineering viewpoint, relatively frequent periodic testing of bulk specific gravity would be considered "good engineering practice" as a QC tool to ensure any significant changes such as in aggregate supply would not go unnoticed. Precision values are 0.026 for LA and 0.036 for OK2.

Table 13. Required Sample Size to Attain Specific Bias Values for Bulk Specific Gravity

Bias	LA	OK1	OK2
0.01	15	117	29
0.02	4	29	7
0.03	2	13	3
0.04	2	8	2
0.05	2	5	2
0.06	2	3	2

An example of the way the data in table 13 would be used is if an engineer on the LA Project wanted to determine if the lot average of the bulk specific gravity is within 0.01 of the target value, 15 tests would be required.

On only the LA project were bulk specific gravity tests performed by the State. One sample was obtained from lot 1 and two samples were obtained from lot 2. These few samples preclude a rigorous statistical analysis but a cursory comparison indicates the State values were lower than those obtained by the study team.

Maximum Theoretical Specific Gravity

Table 14 lists the summary and test statistics for the equality of variance test among the sublots for the MTSG.

Table 14. Descriptive Statistics for Maximum Theoretical Specific Gravity

Project	Lot	Mean	Std. Dev.	UL	LL	Var. Test
LA	1	2.516	0.0071	2.519	2.512	5.220
(2.512)	2	2.511	0.0060	2.513	2.508	2.200
OK1	1	2.505	0.0113	2.510	2.500	3.450
(2.476)	2	2.507	0.0079	2.510	2.503	7.314
OK2	1	2.505	0.0070	2.508	2.502	36.885*
(2.476)	2	2.507	0.0128	2.513	2.502	20.026*

The standard deviations found on all the projects were greater than that contained in ASTM D2041-91 "Standard Test Method for Theoretical Maximum Specific Gravity of Bituminous Paving Mixtures," which has a single-operator standard deviation of 0.004.(77) The standard deviation measured from the AMRL proficiency sample data base is 0.0147, which is larger than those found on any of the projects.(88) Thus, because of the additional sources of variability that exist in the study samples, the variabilities found in these projects should not be considered excessively large.

The sublot variances were not equal for OK2 but were for OK1 and LA. The sampling plan and bias and precision calculations should be project-specific for OK2. Table 15 lists ranges of required sample size for a range of bias values. The biases found are 0.001 for LA and 0.030 for OK1. Precision values are 0.014 for LA and 0.019 for OK1. For OK2, the maximum sublot standard deviation of 0.02506 was used.

Table 15. Required Sample Size to Attain Specific Bias Values for Maximum Theoretical Specific Gravity

Bias	LA	OK1	OK2
0.005	16	32	216
0.006	12	23	150
0.007	9	17	110
0.008	7	13	84
0.009	5	10	66
0.010	4	8	54
0.020	2	2	13
0.040	2	2	4

For this quality characteristic, on only the LA project were tests performed by the State. As for the bulk specific gravity, one sample was obtained from lot 1 and two samples were obtained from lot 2. A comparison indicates the State values were the same as for lot 1 and slightly higher than for lot 2 obtained by the study team.

Air Voids

Table 16 lists the summary and test statistics for the equality of variance tests among the sublots for air voids produced under a 75-blow Marshall compactive effort. The standard deviations found in this study compare favorably with those that were reported in NCHRP Synthesis 232, which ranged from 0.5 to 0.9.[6] The standard deviation measured from the AMRL proficiency sample data base is 0.8305, which is comparable with those found on these projects.[88] Thus, once again, the variabilities found in these projects should be considered typical.

The sublot variances were not equal for OK1 but were equal for LA and OK2. The sampling plan and bias and precision calculations should be project-specific for OK1. The standard deviations found on OK1 are larger than those on LA and OK2. The bias value for LA was 1.2 and OK2 was 2.2. These bias values indicate that the sublot means were appreciably higher than the target values for both LA and OK2.

Table 16. Descriptive Statistics for Air Voids, % (75-blow Marshall Compaction)

Project	Lot	Mean	Std. Dev.	UL	LL	Var. Test
LA	1	5.57	0.633	5.84	5.29	7.90
(4.0)	2	4.93	0.667	5.23	4.64	5.716
OK1	1	4.20	1.094	4.68	3.72	42.75*
(4.1)	2	3.94	0.952	4.36	3.52	10.38*
OK2	1	6.30	0.938	6.71	5.89	5.32
(4.1)	2	6.25	0.667	6.54	5.95	6.34

Table 17 lists ranges of required sample size for a range of bias values. This table indicates that the sample size required for LA, with precision of 1.41, is less than that required for OK1 or OK2. Saying this another way, an equal sample size for LA, OK1, and OK2 will result in detecting only a higher bias for OK1 or OK2. However, if detecting a bias of 0.50 percent air voids is desirable, a large sample size would be required. For the OK1 project, for this level bias, the sample size is impractical.

Since the States did not perform 75-blow laboratory compaction on any project, comparable State air void data were not available.

Asphalt Content

Table 18 lists the descriptive statistics for asphalt content. Most of the project standard deviations compare favorably with typically found asphalt content standard deviations using extraction procedures (0.20 to 0.25 percent).[6] The standard deviation measured from the AMRL proficiency sample data base is 0.24 percent, which is comparable with those found on these projects.[88] Thus, once again, the variabilities found in these projects should be considered typical.

Table 17. Required Sample Size to Attain Specific Bias Values
for Air Voids, %

Bias	LA	OK1	OK2
0.30	49	308	62
0.40	28	173	35
0.50	18	111	22
0.60	12	77	15
0.70	9	57	11

Table 18. Descriptive Statistics for Asphalt Content, %

Project	Lot	Mean	Std. Dev.	UL	LL	Var. Test
LA	1	4.29	0.253	4.40	4.17	2.60
(4.2)	2	4.42	0.207	4.51	4.33	11.06*
OK1	1	4.53	0.169	4.61	4.46	42.25*
(4.8)	2	4.46	0.226	4.56	4.36	17.67*
OK2	1	4.24	0.250	4.35	4.13	5.35
(4.8)	2	4.16	0.268	4.28	4.05	2.20

The sublot variances were not equal for LA or OK1. The bias for OK2 was -0.60. The precision value for OK2 was 0.51.

Table 19 lists the ranges of required sample size for a range of bias values for each project. Since the sublot variances were unequal for OK1, the estimate for the standard deviation used in determining sample size was the maximum standard deviation among the sublots, namely 0.260. For OK2, the sublot variances were equal and the lot means were equal. As for OK1, the estimate for the standard deviation used in determining sample size was the maximum standard deviation among the sublots, 0.267.

Table 19. Required Sample Size to Attain Specific Bias Values for Asphalt Content, %

Bias	LA	OK1	OK2
0.10	91	58	57
0.20	23	15	14
0.30	10	6	6
0.40	6	4	4
0.50	4	2	2

More testing was done by each State on this quality measure, allowing a statistical analysis between the State results and those of the study team. The number of State asphalt extraction tests performed is as follows: on the LA project, four tests were performed on lot 1, three on lot 2; on the OK1 project, 22 tests were performed on lot 1 and 26 on lot 2; and on the OK2 project, two tests were performed on both lot 1 and lot 2. The statistical data from the State and study team are shown in table 20. Statistical F and t tests were used to compare the data on a lot-by-lot basis between the State and the study team data. The statistical tests indicate no statistical differences on any lot variances at the 95-percent significance level, with the exception of OK1 lot 1, in which the State standard deviation was 0.031 percent, an unusually low value compared with the standard deviation measured by the study team on this lot of 0.169 percent. Three of the sublots (with n = 4 or 5) within lot 1 had a zero standard deviation, a very unusual occurrence. The State standard deviations also were unusually small for both lots of the OK2 project, but because there were only two test results, the variances were not statistically different from those of the study team. The State average values were significantly different from those obtained by the study team on every lot except LA lot 1. They were lower on the LA project and higher on both OK projects.

Table 20. Descriptive Statistics for Asphalt Content From State and Study Tests

Project	Lot	State Mean	Study Mean	State Std. Dev.	Study Std. Dev.
LA	1	4.18	4.29	0.283	0.253
(4.2)	2	4.00	4.42	0.218	0.207
OK1	1	4.66	4.53	0.021	0.169
(4.8)	2	4.85	4.46	0.021	0.226
OK2	1	4.64	4.24	0.032	0.250
(4.8)	2	4.64	4.16	0.212	0.268

Aggregate Gradation

The aggregate gradations were analyzed by examining each sieve individually.

19-mm Sieve Table 21 provides the descriptive statistics for the percent of aggregate passing the 19-mm sieve. On only the Louisiana project was any variability measured and, as would be expected on the top-sized sieve, it was small. The variability on this sieve was not compared with other reference values because, if the 19-mm sieve were not the top size sieve of the material used in the reference value, the variabilities measured would be quite different. The bias for the LA project was -0.4 percent.

Table 21. Descriptive Statistics for Percent Passing 19-mm Sieve

Project	Lot	Mean	Std. Dev.	UL	LL	Var. Test
LA	1	99.0	0.92	99.4	98.6	2.52
(99.6)	2	99.4	0.67	99.6	99.1	1.92
OK1	1	100.0	0.00	100.0	100.0	1.0
(100.0)	2	100.0	0.00	100.0	100.0	1.0
OK2	1	100.0	0.00	100.0	100.0	1.0
(100.0)	2	100.0	0.00	100.0	100.0	1.0

The sublot variances were equal for all of the projects. In addition, the lot means were equal for each project. However, as expected, the means between projects were not equal. This is understandable, but the lot means for LA are essentially the same as for both OK projects from an engineering viewpoint. So, although the sampling plan and bias and precision calculations must be project-specific from the statistical analysis, from an engineering analysis they do not have to be. The intent of controlling the top-size sieve is primarily to prevent an excess of oversize material and, while this sieve is not considered as critical as some of the smaller sieve sizes for acceptance purposes, for QC, it is important as a detection of changes in aggregate supply or handling.

Table 22 lists the required sample size for a range of bias values. Because the standard deviation was small to nonexistent, only one sample is required to achieve even a minimal level of bias irrespective of the project from a statistical viewpoint. For quality control purposes, more frequent sampling is considered "good engineering practice."

Table 22. Required Sample Size to Attain Specific Bias Values
for Percent Passing 19-mm Sieve

Bias	LA	OK1	OK2
0.10	1	1	1
0.20	1	1	1
0.30	1	1	1
0.40	1	1	1
0.50	1	1	1

The States measured one or two gradations for each lot. For the Oklahoma projects only the 19-mm, 13-mm, 10-mm, 5-mm, and 0.075-mm sieves were the same as those used by the study team. Because of the few State results on the 19-mm sieve, a statistical analysis was not done; however, the State results were in general agreement with those of the study team.

13-mm Sieve Table 23 provides the descriptive statistics for the percent of aggregate passing the 13-mm sieve. All but one of these standard deviations (i.e., LA Lot 2) compare favorably with standard deviations typically found in surface course mixes (1.4 to 2.3 percent).[6] The standard deviation measured from the AMRL proficiency sample data base is 1.1 percent, which also indicated that the variability of the LA project, particularly lot 2, was a little higher than typical.[88]

Table 23. Descriptive Statistics for Percent Passing 13-mm Sieve

Project	Lot	Mean	Std. Dev.	UL	LL	Var. Test
LA	1	86.9	1.77	87.7	86.1	2.99
(90.4)	2	88.0	2.86	89.2	86.7	1.58
OK1	1	98.8	0.83	99.1	98.4	1.57
(98.0)	2	98.7	0.57	98.6	98.4	1.92
OK2	1	98.7	0.75	99.0	98.4	2.00
(98.0)	2	97.9	1.33	98.5	97.3	3.00

Bias values were -3.0, 0.8, and 0.3 percent, for the LA, OK1, and OK2 projects, respectively. This indicates that the LA project was appreciably off-target. Precision values were 4.7, 1.4, and 2.2 for the LA, OK1, and OK2 projects, respectively.

The sublot variances were equal for all of the projects. However, the lot means were not equal for the LA and OK2 projects. Hence, the sampling plan and bias and precision calculations should be project-specific.

Table 24 lists the sample size for a range of bias values. Because the standard deviation was small for each of the three projects, from a statistical viewpoint, only one sample is required to determine even a minimal level of bias. However, as has been mentioned previously, for QC purposes, more frequent testing should be done.

Table 24. Required Sample Size to Attain Specific Bias Values for Percent Passing 13-mm Sieve

Bias	LA	OK1	OK2
0.10	1	1	1
0.20	1	1	1
0.30	1	1	1
0.40	1	1	1
0.50	1	1	1

The few State results that were available on this sieve agreed very well with those from the study team, generally within ±1.0 percent.

10-mm Sieve Table 25 provides the descriptive statistics for the percent of aggregate passing the 10-mm sieve. These standard deviations compare favorably with standard deviations typically found in surface course mixes (1.9 to 4.4).(6) The standard deviation measured from the AMRL proficiency sample data base is 1.2 percent, which, although lower than those found on these projects, does not include material and plant production variability.(88) Thus, once again, the variabilities found in these projects should be considered typical.

Table 25. Descriptive Statistics for Percent Passing 10-mm Sieve

Project	Lot	Mean	Std. Dev.	UL	LL	Var. Test
LA	1	66.6	3.27	68.0	65.1	3.78
(68.0)	2	67.8	3.22	69.3	66.4	4.70
OK1	1	88.5	2.59	89.6	87.4	7.58
(87.0)	2	89.2	2.45	90.3	88.2	4.20
OK2	1	88.9	1.65	89.6	88.2	2.48
(87.0)	2	88.1	2.43	89.2	87.0	2.55

Bias values were -0.8, 1.8, and 1.5 percent, for the LA, OK1, and OK2 projects, respectively. This indicates that all the projects were generally on target. Precision values were 6.4, 4.9, and 4.1 for the LA, OK1, and OK2 projects, respectively.

The sublot variances and the lot means were equal for each of the projects. However, the project means were not equal; hence, the sampling plan and bias and precision calculations should be project-specific. Again, because of the relatively small variances measured, this is theoretical.

A few more State results were available for analysis on the 10-mm sieve. The average results were in very good agreement with those from the study team, generally within less than ±1.0 percent. The standard deviations of the State tests were generally lower than those from the study, but agreed reasonably well.

Table 26 lists ranges of required sample size for a range of bias values. The standard deviation was small for each of the three projects; therefore, the largest number of samples to determine even small levels of bias is two.

Table 26. Required Sample Size to Attain Specific Bias Values for Percent Passing 10-mm Sieve

Bias	LA	OK1	OK2
0.10	2	1	1
0.20	1	1	1
0.30	1	1	1
0.40	1	1	1
0.50	1	1	1

5-mm Sieve Table 27 provides the descriptive statistics for the percent of aggregate passing the 5-mm sieve. These standard deviations compare favorably with standard deviations typically found for this sieve in surface course mixes (2.8 to 3.5).[6] The standard deviation measured from the AMRL proficiency sample data base is 1.0 percent, which only serves as a reference value of between-laboratory variability, not material and plant production process variability.[88]

Table 27. Descriptive Statistics for Percent Passing 5-mm Sieve

Project	Lot	Mean	Std. Dev.	UL	LL	Var. Test
LA	1	36.7	2.05	37.6	35.7	3.37
(37.0)	2	37.5	2.35	38.6	36.4	2.69
OK1	1	61.4	3.08	62.6	59.7	3.61
(65.0)	2	62.1	2.34	63.2	61.0	3.59
OK2	1	68.0	2.70	69.3	66.7	1.43
(65.0)	2	66.8	2.98	68.2	65.5	3.24

Bias values were 0.1, -3.2, and 2.4 percent, for the LA, OK1, and OK2 projects, respectively. This indicates that the OK1 project was appreciably off target. Precision values were 4.3, 5.3, and 5.61 for the LA, OK1, and OK2 projects, respectively.

The sublot variances and the lot means were equal for each of the projects. However, the project means were not equal; hence, the sampling plan and bias and precision calculation should be project-specific. Because the sublot variances were equal, from a statistical viewpoint, for all but the minimal level of bias, a sample of size one will suffice. This is interesting and probably unacceptable from an engineering viewpoint because the 5-mm sieve is considered one of the critical sieves for accepting and controlling gradation. As will be discussed subsequently, one sieve in a range of gradation can be used for acceptance purposes. If the sieve chosen were the 5-mm, more than one test per lot would be needed to estimate the lot population.

The same number of State results were available for analysis on the 5-mm sieve as for the 10-mm sieve. Once again, the average State results were in very good agreement with those from the study team, generally within less than ±1.0 percent. The standard deviations of the State tests were generally lower than those from the study, but agreed reasonably well.

Table 28 lists ranges of required numbers of sample areas for a range of bias values. The required sample size to achieve even a minimal level of bias is quite small.

Table 28. Required Sample Size to Attain Specific Bias Values for Percent Passing 5-mm Sieve

Bias	LA	OK1	OK2
0.10	4	2	2
0.20	1	1	1
0.30	1	1	1
0.40	1	1	1
0.50	1	1	1

2.36-mm Sieve Table 29 provides the descriptive statistics for the percent of aggregate passing the 2.36-mm sieve. These standard deviations compare favorably with standard deviations typically found in surface course mixes (1.7 to 3.2).[6] The standard deviation measured from the AMRL proficiency sample data base for this sieve is 0.9 percent, which, again, only serves as a reference value of between-laboratory variability.[88]

Bias values were -0.1, 1.3, and 7.5 percent for the LA, OK1, and OK2 projects, respectively. This indicates that the OK2 project was way off target, producing a much finer gradation than called for by the JMF. Precision values were 4.0, 4.0, and 4.5 for the LA, OK1, and OK2 projects, respectively. The OK2 project is an example of poor accuracy and reasonable precision.

The sublot variances and lot means were equal for each of the projects. However, the project means were not equal; hence, the sampling plan and bias and precision calculations should be project-specific. The same comments presented on the 5-mm sieve apply here.

Table 29. Descriptive Statistics for Percent Passing 2.36-mm Sieve

Project	Lot	Mean	Std. Dev.	UL	LL	Var. Test
LA	1	25.4	1.64	26.2	24.6	6.22
(26.0)	2	26.4	2.30	27.4	25.3	4.72
OK1	1	45.1	2.09	46.1	44.0	3.19
(44.0)	2	45.5	2.01	46.4	44.5	2.48
OK2	1	51.4	2.09	52.4	50.4	1.58
(44.0)	2	51.6	2.44	51.6	49.4	2.99

Data from State tests on this sieve were available only on the Louisiana project and only three or four samples were obtained on the two lots. The average of the State results were close to those of the study team, as were the standard deviations.

Table 30 lists ranges of required sample size for a range of bias values. The required numbers of samples to achieve even a minimal level of bias are small but given the considered critical nature of this sieve size, a small bias (0.10) and a sample size greater than one would be reasonable, if for no other reason than to provide an estimate of the variability.

Table 30. Required Sample Size to Attain Specific Bias Values
for Percent Passing 2.36-mm Sieve

Bias	LA	OK1	OK2
0.10	6	2	3
0.20	2	1	1
0.30	1	1	1
0.40	1	1	1
0.50	1	1	1

1.18-mm Sieve Table 31 provides the descriptive statistics for the percent of aggregate passing the 1.18-mm sieve. These standard deviations are low compared with standard deviations typically found on slightly finer sieves used in surface course mixes (1.3 to 2.1).[6] The standard deviation measured from the AMRL proficiency sample data base for this sieve is 0.8 percent, which, again, only serves as a reference value of between-laboratory variability.[88]

Table 31. Descriptive Statistics for Percent Passing 1.18-mm Sieve

Project	Lot	Mean	Std. Dev.	UL	LL	Var. Test
LA	1	19.2	1.09	19.7	18.6	4.44
(17.0)	2	20.0	1.12	20.5	19.5	2.63
OK1	1	30.5	1.54	31.2	29.7	3.05
(30.0)	2	30.9	1.35	31.5	30.2	3.42
OK2	1	36.4	1.39	37.1	35.7	2.17
(30.0)	2	36.1	1.52	36.8	35.4	2.08

Bias values were 2.4, 0.7, and 6.2 percent for the LA, OK1, and OK2 projects, respectively. This indicates that on this sieve, as on the 2.36-mm sieve, the OK2 project was way off target, producing a much finer gradation than required by the JMF. Precision values were 2.3, 2.8, and 2.8 for the LA, OK1, and OK2 projects, respectively.

The sublot variances were equal for each of the projects. The lot means were not equal for the LA project. Therefore, the sampling plan and bias and precision calculations should be project-specific.

Data from State tests on this sieve were available only on the LA project and only three or four samples were obtained on the two lots. The averages and standard deviations of the State results were higher than those of the study team.

Table 32 lists ranges of required sample size for a range of bias values. Since the standard deviations are quite small in comparison with the means of the data, the required sample size to achieve even a minimal level of bias is quite small.

Table 32. Required Sample Size to Attain Specific Bias Values for Percent Passing 1.18-mm Sieve

Bias	LA	OK1	OK2
0.10	7	2	2
0.20	2	1	1
0.30	1	1	1
0.40	1	1	1
0.50	1	1	1

0.30-mm Sieve Table 33 provides the descriptive statistics for the percent of aggregate passing the 0.30-mm sieve. For this sieve, the standard deviations are lower than typically found in surface course mixes (1.3 to 1.6).[6] The standard deviation measured from the AMRL proficiency sample data base for this sieve is 0.8 percent.[88]

Table 33. Descriptive Statistics for Percent Passing 0.30-mm Sieve

Project	Lot	Mean	Std. Dev.	UL	LL	Var. Test
LA	1	9.1	0.89	9.5	8.6	3.46
(8.0)	2	9.9	0.72	10.2	9.6	2.82
OK1	1	14.2	0.71	14.6	13.9	1.92
(14.0)	2	14.5	0.69	14.8	14.1	1.92
OK2	1	14.1	1.15	14.6	13.5	2.00
(14.0)	2	14.0	1.19	14.5	13.4	3.00

Bias values were 1.5, 0.3, and 0.0 percent for the LA, OK1, and OK2 projects, respectively. This indicates that on this sieve, all projects were right on target. Precision values were 1.8, 1.4, and 2.3 for the LA, OK1, and OK2 projects, respectively. These results are examples of good accuracy and good precision.

The sublot variances were equal for each of the projects. The lot means were not equal for the LA project. Therefore, the sampling plan and bias and precision calculations should be project-specific.

Data from State tests on this sieve were available only on the LA project and only three or four samples were obtained on the two lots. The averages and standard deviations of the State results were very close to those obtained by the study team.

Table 34 lists ranges of required sample size for a range of bias values. Since the standard deviations are relatively small in comparison with the means, a relatively small sample size can be used to detect almost any level of bias.

Table 34. Required Sample Size to Attain Specific Bias Values
for Percent Passing 0.30-mm Sieve

Bias	LA	OK1	OK2
0.10	12	6	3
0.20	3	2	1
0.30	2	1	1
0.40	1	1	1
0.50	1	1	1

0.15-mm Sieve Table 35 provides the descriptive statistics for the percent of aggregate passing the 0.15-mm sieve. These standard deviations are close to those typically found for this sieve size in surface mixes (1.1 to 1.2).[6] The standard deviation measured from the AMRL proficiency sample data base for this sieve is 0.5 percent.[88]

Table 35. Descriptive Statistics for Percent Passing 0.15-mm Sieve

Project	Lot	Mean	Std. Dev.	UL	LL	Var. Test
LA	1	5.6	0.75	6.0	5.2	2.52
(4.4)	2	6.2	0.77	6.6	5.8	3.00
OK1	1	7.0	0.94	7.5	6.5	2.22
(8.0)	2	7.5	1.28	8.1	6.9	4.44
OK2	1	5.3	1.26	5.9	4.7	3.00
(8.0)	2	5.3	1.13	5.8	4.8	3.00

Bias values were 1.5, -0.8, and -2.7 percent for the LA, OK1, and OK2 projects, respectively. This indicates that for this sieve, the LA and OK2 projects were slightly off target and OK1 was closer to the target. Precision values were 1.6, 2.2, and 2.3 for the LA, OK1, and OK2 projects, respectively.

The sublot variances were equal for each of the projects. The lot means for the LA project were unequal. Hence, the sampling plan and bias and precision calculations should be project-specific.

Data from State tests on this sieve were available only on the LA project and only three or four samples were obtained on the two lots. The average for lot 1 of the State results was identical to that obtained by the study team; the average of lot 2 was slightly finer than that of the study. The standard deviations of both lots were comparable.

Table 36 lists ranges of required sample size for a range of bias values.

<u>0.075-mm Sieve</u> Table 37 provides the descriptive statistics for the percent of aggregate passing the 0.075-mm sieve. These standard deviations are close to those typically found for this sieve size in surface mixes (0.5 to 1.0).(6) The standard deviation measured from the AMRL proficiency sample data base for this sieve is 0.5 percent.(88)

Table 36. Required Sample Size to Attain Specific Bias Values
for Percent Passing 0.15-mm Sieve

Bias	LA	OK1	OK2
0.10	36	18	19
0.20	9	5	5
0.30	4	2	3
0.40	3	2	2
0.50	2	1	1

Table 37. Descriptive Statistics for Percent Passing 0.075-mm Sieve

Project	Lot	Mean	Std. Dev.	UL	LL	Var. Test
LA	1	4.3	0.40	4.5	4.1	3.28
(3.2)	2	4.8	0.83	5.2	4.4	16.50*
OK1	1	4.0	0.82	4.4	3.7	4.32
(4.5)	2	4.1	0.72	4.4	3.7	1.85
OK2	1	2.5	0.57	2.7	2.2	2.83
(4.5)	2	2.8	0.92	3.2	2.3	3.62

Bias values were 1.3, -0.5, and -1.9 percent for the LA, OK1, and OK2 projects, respectively. This indicates that for this sieve, the LA and OK2 projects were slightly off target and OK1 was closer to the target. Precision values were 1.4, 1.5, and 1.5 for the LA, OK1, and OK2 projects, respectively.

The sample area variances were not equal for all of the projects. Hence, the sampling plan and bias and precision calculations should be project-specific.

State tests results on this sieve were available on all projects. The average State results were in good agreement with those from the study team, especially those from the LA project, which were within ±0.3 percent. The averages of the State results from the Oklahoma projects differed from those of the study team by 1.5 to 2.0 percent for OK1 and OK2, respectively. The standard deviations of the State tests were generally lower than those from the study, but agreed reasonably well.

Table 38 lists ranges of required sample size for a range of bias values. Similar to the discussion on the 5-mm and 2.36-mm sieves, this sieve is considered critical. It would probably be desirable to be able to detect relatively small biases of 0.1 to 0.2 percent, which would require a reasonably large sample size.

Table 38. Required Sample Size to Attain Specific Bias Values for Percent Passing 0.075-mm Sieve

Bias	LA	OK1	OK2
0.10	46	31	32
0.20	12	8	8
0.30	6	4	4
0.40	3	2	2
0.50	2	2	2

HMAC Field Result Analysis

Density

The average and standard deviations of the nuclear density tests results in terms of percent MTSG are shown in table 39. As mentioned in chapter 4, nuclear gauges were used to measure density on the LA and OK2 projects, and cores were used on OK1. Density tests were performed by the State and, thus, the number of tests was not consistent from project to project or lot to lot; the selection of transverse location also was a State decision.

Table 39. Descriptive Statistics for Density Tests, % MTSG

Project	LOT	n	Left Wheelpath		Right Wheelpath	
			Mean	Std. Dev	Mean	Std. Dev.
LA	1	20	95.6	1.54	95.2	1.83
Nuclear	2	20	87.7	1.74	85.8	1.63
OK1	1	10	94.0	0.94	94.5	1.09
Cores	2	8	94.3	1.48	95.0	1.82
OK2	1	10	94.5	1.81	96.2	1.43
Nuclear	2	9	93.5	1.57	95.6	2.53

For all the density results, at an α level of 5 percent, the variances were not significantly different between wheelpaths. In two of the four lots measured by the nuclear gauge, the means were significantly different. This indicates that a difference exists transversely in average density at least when measured by the nuclear gauge. For OK1, which used cores for density, no significant difference was found either between wheelpaths or between lots.

GPR Thickness

Table 40 provides the descriptive statistics for the GPR thickness. The thickness target values are in parentheses. Means and standard deviations presented in table 40 are obtained from all of the available measurements for the given lot.

Table 40. Descriptive Statistics for GPR Thickness, mm

Project	Lot	Mean	Std. Dev.	UL	LL	Var. Test
LA	1	56.9	4.70	57.2	56.4	4.78
(51)	2	53.4	4.92	56.2	53.1	5.13
OK1	1	50.3	7.01	50.8	49.8	8.95*
(57)	2	50.3	9.65	50.8	49.5	14.72*
OK2	1	65.5	3.28	65.8	65.5	1.67
(64)	2	63.2	3.07	65.3	63.2	1.93

Bias values were 4.1, -6.7, and 0.4 mm for the LA, OK1, and OK2 projects, respectively. This indicates that for GPR thickness, the Louisiana and OK1 projects were slightly off target, and OK2 was very close to the target.

The sublot variances were equal for LA and OK2 but not for OK1, which had a range of standard deviations from 3.4 to 13.2 mm. Hence, the sampling plan and bias and precision calculations should be project-specific. For OK1, the maximum standard deviation of 13.2 mm was used. As previously stated in chapter 5, the GPR data were collected at four transverse locations, which allowed a comparison of transverse variability. The transverse means for the LA and OK2 projects were not equal and there was no significant interaction; thus, four transverse locations per lot would be required to estimate thicknesses with biases of 9.7 to -0.8 mm for LA and 8.4 to -1.5 mm for OK2. Precision values are 4.9 mm for LA and 3.3 mm for OK2 on the basis of the maximum lot standard deviation.

Table 41 lists ranges of required sample size for a range of bias values. The sample size used in this study with four transverse locations would detect a bias of 7.5 mm for LA, 5.0 mm for OK2, and 20.0 mm for OK1. The sample sizes indicate that a 7.6-m spacing may not be necessary to obtain an accurate measurement of the thickness for the lot. Sample sizes presented are for a lot. For instance, on the LA project, to obtain a measure of bias 2.5 mm it would be necessary to take 32 measurements on each lot of the project. This would require four measurements across the pavement at every 190-m interval.

As will be seen, the standard deviations of the GPR thickness results on OK1 are greater than those from cores. But it should be kept in mind that the GPR values consist of a sample size of approximately 160 for each sublot as opposed to a sample size of 4 for the core data, so the standard error for the GPR tends to be less than that for the cores. One other advantage of GPR testing is that it is nondestructive.

Table 41. Required Sample Size to Attain Specific Bias Values for GPR Thickness, mm.

Bias	LA	OK1	OK2
2.5	32	14	230
5.0	8	4	58
7.5	4	2	26
10.0	2	2	14
12.5	2	2	9
15.0	2	2	6
17.5	2	2	5
20.0	2	2	4

A regression analysis was performed on the GPR data to discern the effect of transverse offset on pavement thickness. Table 42 provides the analysis of variance results from this analysis. Table 43 provides the results for each of the parameters included in the model. As can be seen, the linear regression is statistically significant, indicating that there is significant variation in the thickness between projects, between lots, and in the transverse direction. The F ratio (1328.6) indicates that a significant amount of the variability in all of the thickness data is explained by the differences in projects, lots, and transverse offsets. The "Prob > F" indicates the probability of getting an F ratio of this size if very little of the variation was explained by the differences in projects, lots, and transverse offsets.

As expected, the projects were significant, indicating that the same thickness was not used on all three surfaces. In addition, the transverse offset is a significant factor. The parameter estimate (0.7) for this factor is positive, which indicates that the HMAC surface layer is thicker toward the center of the pavement than it is on the pavement edge.

Table 42. Analysis of Variance Results from Linear Regression of GPR Thickness for HMAC Project

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	194055.87	38811.2	1328.609
Error	4698	137237.40	29.2	**Prob > F**
Total	4703	331293.26		0.0000

Table 43. Results for Parameters Used in Linear Regression of GPR Thickness for HMAC Projects

Term	Estimate	Std. Error	t Ratio	Prob > |t|
Intercept	52.3	0.16	323.68	0.0000
Project[LA-OK2]	-1.5	0.11	-13.22	<0.0001
Project[OK1-OK2]	-6.4	0.11	-56.98	0.0000
Project[LA-OK2]*Lot[1-2]	-0.7	0.11	-6.60	<0.0001
Project[OK1-OK2]*Lot[1-2]	1.0	0.11	8.57	<0.0001
Transverse Offset	0.7	0.02	30.84	<0.0001

Core Thickness

The cores taken on OK1 allowed an analysis, in addition to that of density, of thickness. Table 44 provides the descriptive statistics of the core thickness data for OK1.

Table 44. Descriptive Statistics for Core Thickness, mm

Project	Lot	Mean	Std Dev	LL	UL
OK1	1	52.9	4.21	50.9	54.9
(57)	2	45.8	6.12	42.9	48.8

These cores allow for comparisons between the GPR thickness results and the thicknesses obtained from the cores. As shown in table 40, the sublot variances were unequal for project

OK1; therefore, comparisons should be made between individual sublots. An F-test was used to compare variances between the measurement methods, and a t-test was used to compare the means between the measurement methods. Results are given in table 45.

Table 45. Results from Comparisons of GPR and Core Thickness Data for OK1

Lot	Sublot	t	F
1	1	2.45*	2.44
	2	0.55	17.08*
	3	6.41*	13.95*
	4	1.22	5.39
	5	1.33	2.72
2	1	0.52	1.68
	2	1.34	2.11
	3	9.48*	19.60*
	4	0.16	8.11
	5	0.14	7.02

Table 45 shows that the variances between the two measurement types are different for 3 of the 10 sublots and the means are also different for 3 of the sublots. In addition to the F- and t-tests performed on these data, a regression analysis was performed to compare the two measurement types. These results are provided in figure 20.

An r value of 0 indicates that no correlation exists, while an r value of 1.0 indicates a perfect correlation. The analysis shows that there is virtually no correlation between GPR thickness and core thickness ($r = 0.05$). However, this lack of correlation may be due to the small sample size. Further evaluation is necessary to fully determine the potential value of GPR.

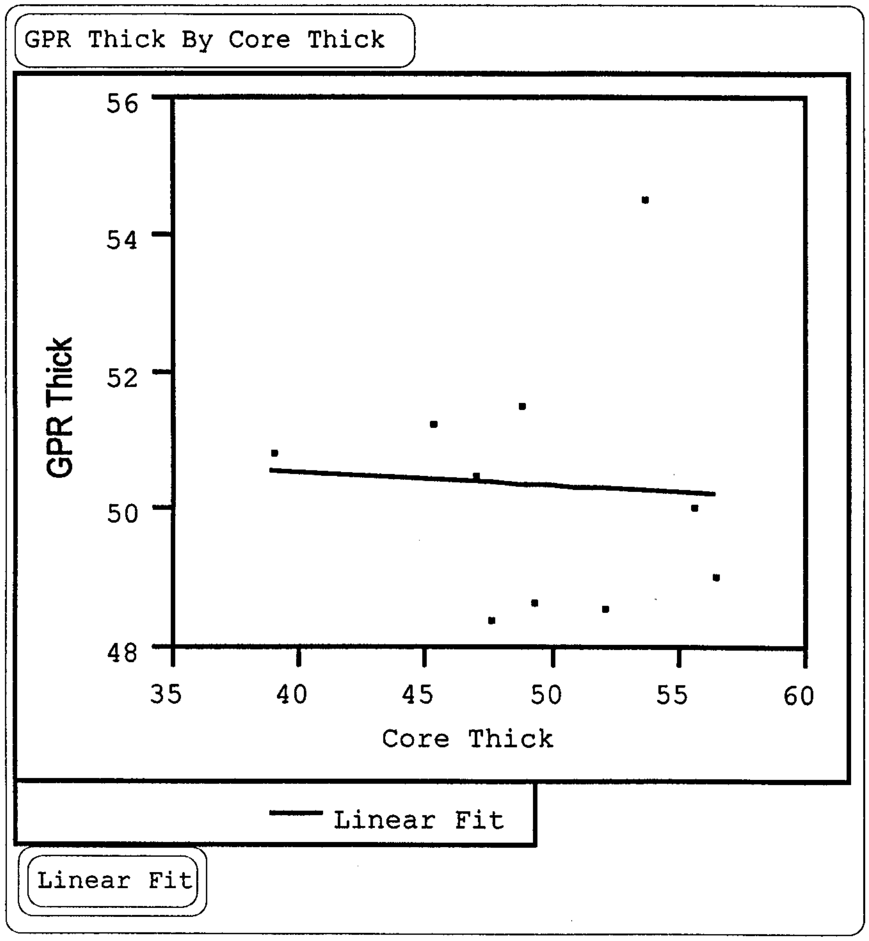


Figure 20. Linear Regression of GPR Thickness versus Core Thickness for Project OK1

CORRELATION OF INTERRELATED HMAC PROPERTIES

The statistical analysis also allowed a correlation between each of the mixture properties analyzed above. There were several properties that produced reasonably good correlation coefficient (r) values. An r value of 0.4 was chosen to indicate that the correlation is significant at 38 degrees of freedom and a conservative α value of 1.0 percent.[85] This means that there is only a 1 percent chance of getting an r value of 0.4 or higher when there is no correlation between the two variables. Those properties that have r values of equal to or greater than 0.40 for each of the projects are shown in table 46.

Table 46. Significant Correlation Coefficient Values of HMAC Mixture Variables

Project LA												
Variable	**BSG**	**MTSG**	**AV**	**AC**	**13-mm**	**10-mm**	**5-mm**	**2.36-mm**	**1.18-mm**	**0.3-mm**	**0.15-mm**	**0.075-mm**
BSG		-0.63	-0.96			0.51	0.61	0.51	0.57	0.43	0.50	0.45
MTSG	-0.63		0.83				-0.49		-0.46			
AV				-0.41		-0.47	-0.62	-0.50	-0.58	-0.44	-0.47	
13-mm						0.67	0.41					
10-mm							0.81	0.63	0.68	0.59		
5-mm								0.81	0.89	0.70	0.61	
2.36-mm									0.87	0.65	0.54	0.42
1.18-mm										0.80	0.65	0.55
0.30-mm											0.80	0.74
0.15-mm												0.86
Project OK1												
BSG			-0.94									
AV				-0.41								
10-mm							0.52	0.49	0.50			
5-mm					0.40			0.95	0.91	0.54	0.50	
2.36-mm									0.92	0.56	0.50	
1.18-mm										0.71	0.57	
0.30-mm											0.59	0.83
0.15-mm												0.50
Project OK2												
BSG			-0.88									
13-mm						0.62	0.51	0.51				
10-mm							0.91	0.88	0.78			
5-mm								0.98	0.93	0.41		
2.36-mm									0.94			
1.18-mm										0.55	0.41	
0.30-mm											0.88	0.82
0.15-mm												0.85

The strongest correlations were between bulk specific gravity versus air voids for all three projects, which is expected since one variable is used to calculate the other; and among the percent passing several adjacent sieve sizes, e.g., the 10-mm versus the 5-mm, particularly the 5-mm versus the 2.36-mm, and the 2.36-mm versus the 1.18-mm, etc. Typically, percent passing adjacent sieves are related; this is a primary reason that SHA's choose a critical sieve size to use for acceptance of gradation over a gradation range. For example, the 13-mm sieve may be chosen to quantify the gradation of the coarse aggregate; the 5-mm or 2.36-mm, since they are so strongly correlated, to quantify the combination of the coarse and fine aggregate; and the 0.075-mm to quantify the gradation of the fine aggregate. This practice appears to be supported by the correlations found here. It also is interesting to note that asphalt content was only correlated with air voids (r = -0.41 for LA and OK1) and no other variable. This is an indication that asphalt content should be a quality measure; or stated differently, if asphalt content is not measured, the only other quality measures that can act as a surrogate would be air voids and, because of the lack of a strong correlation, that variable would not serve well.

CONFORMAL INDICES FOR HMAC PROPERTIES

An alternate approach to the use of bias as an indication of being able to hit a target value is a statistic referred to as the CI.[9,10] As mentioned in chapter 2, this statistic is a direct measure of process capability and can be employed to estimate the size and incidence of deviations (variations) from the quality level target, such as the approved target JMF.

The data collected in this study allowed for the computation of the CI values for the three projects. The values were determined for each lot for each project and are presented in table 47. The most notable differences in CI on the three projects are in thickness, the 2.36-mm, and 1.18-mm sieves. The largest deviations from target for thickness were experienced on the OK1 project, which concurs with the biases seen in the GPR thickness data. The same trend is seen on the 2.36-mm and 1.18-mm sieves for OK2, which had higher bias values. Therefore, the relative difference in bias values is confirmed by the relative difference in CI values.

Table 47. Conformal Indices for HMAC Projects

Project	Lot	BSG	MTSG	AV	AC, %	19-mm	13-mm	10-mm	5-mm	2.36-mm	1.18-mm	0.30-mm	0.15-mm	0.075-mm	Thickness, mm
LA	1	0.044	0.007	1.75	0.26	1.1	3.9	3.1	2.0	1.6	2.0	1.4	1.4	1.2	8
	2	0.234	0.006	9.45	0.30	0.7	3.7	3.5	2.4	2.3	2.7	2.1	1.9	1.8	6
OK1	1	0.050	0.031	1.94	0.31	0.0	1.1	2.9	4.7	2.3	1.8	0.8	1.1	0.9	10
	2	0.052	0.032	2.00	0.41	0.0	0.9	3.3	3.7	2.4	1.9	0.9	1.3	0.8	12
OK2	1	0.047	0.030	1.87	0.61	0.0	1.0	2.5	4.0	7.5	7.0	1.1	2.7	2.1	4
	2	0.057	0.034	2.60	0.69	0.0	1.3	2.6	3.4	6.8	6.7	1.2	2.7	2.0	3

PCC DATA

The mixture and construction quality characteristics of the three PCC projects are discussed in this section. In contrast to the sampling and testing done on the HMAC projects, on the PCC projects, the study team and agency jointly collected samples that were tested jointly in the field and at either an agency or independent laboratory. Therefore, no comparison between study team and agency data can be made.

PCC Mixture Result Analysis

The mixture characteristics measured were plastic air content, compressive strength, and unit weight. The sampling plan for the PCC projects differed from that used on the HMAC projects. On each PCC project, two sections were tested, with each section containing five sublots; however, a different number of tests were conducted on each sublot. The effect of this on the statistical analysis will be noted where appropriate.

Each characteristic is discussed separately. The summary and test statistics used for determining the equality of variance tests among sublots are tabulated first. As in the HMAC data analysis, the overall variability is reported as measured by the standard deviation. The study standard deviations are compared with other referenced standard deviations where the latter were found. One reference source is the Cement and Concrete Reference Laboratory (CCRL).[89] The first step is, as in the HMAC analysis, to conduct a test of equality of variances among sublots within projects and lots. Just as with the HMAC results, the ratio of variances is presented in the table, and an * indicates that the variances were not equal at the 5-percent level of significance using the Scheffe test to perform the comparisons.

The second analysis is the number of samples necessary to attain a specific bias value. As discussed previously, the range of the required sample size is tabulated using an α level of significance of 5 percent and an 80-percent power of the test for each project.

A third table contains the bias values for each lot on each project. This bias is the level of bias that can be detected with the number of samples taken. This analysis is in lieu of determining the actual bias on each project, since many of the variables had either no requirement or a minimum requirement.

Plastic Air Content

Table 48 lists summary and test statistics for plastic air content for the equality of variance test among sublots using the same statistical techniques used in the HMAC analyses. Also, as with the HMAC data, the target values are in parentheses. For this quality characteristic, two States accept on a range of air contents and one accepts on a single target value. For Minnesota, which has a single target value, the bias was 0.3 percent. The sublot variations were not equal for Illinois (range of sublot standard deviations 0.12 to 2.27 percent) and Minnesota (0.06 to 0.49 percent). Therefore, the sampling plan and bias and precision calculations should be project-specific. The maximum variance among sublots is used for Illinois (standard deviation 2.27 percent) and for Minnesota (0.49 percent). Since sublot variances were equal for Ohio, the pooled sublot variance (standard deviation of 0.717 percent) is used. A test of equality of sublot means indicated that there are significant differences in sublot means; therefore, sampling should be done in multiple sublots. The standard deviations measured on these projects is in the range reported in the literature (0.53 to 1.11 percent).[6] The standard deviation from the CCRL for air content is 0.53 percent, but does not include material or plant production variability.[89] The standard deviations found in this study can be, for the most part, considered typical.

Table 48. Descriptive Statistics for Plastic Air Content, %

Project	Lot	Mean	Std. Dev.	n	Var. Test	df
IL (5-8)	1	6.11	0.900	15	96.70*	2,2
	2	4.77	1.17	28	11.47*	3,5
MN (5.5)	1	6.03	0.50	15	56.25*	2,2
	2	5.46	0.51	14	1.85	2,2
OH (6-10)	1	6.88	0.87	55	4.18	7,11
	2	8.11	0.83	60	3.84	11,11

Table 49 lists ranges of required sample size per sublot for a range of bias values for each project.

Table 49. Required Sample Size to Attain Specific Bias Values for Plastic Air Content, %

Bias	IL	MN	OH
0.50	78	17	25
0.75	35	8	11
1.00	20	4	6
1.25	13	3	4
1.50	9	2	3
1.75	7	2	2
2.00	5	2	2

Table 50 lists the bias values for the sample sizes used in the study. As an example, IL section 1, with a sample size of 15, yields a bias of 1.138 and section 2, with 28 sample areas, had a bias of 0.83. The question is similar to that addressed in the HMAC analysis: Is the detection of a larger level of bias acceptable or should a larger sample size be required? The question should be addressed by the specification writer when the limits for acceptable variability are established. If the specification does not contain a variability component, a permissible variability should be assumed in order to establish a reasonable sample size. However, if variability is not measured in the acceptance plan, the validity of the assumed variability will not be determined.

Table 50. Bias Values for Plastic Air Content, %

Site	Section	n	Bias
IL	1	15	1.138
	2	28	0.830
MN	1	15	0.529
	2	14	0.547
OH	1	55	0.334
	2	60	0.320

Compressive Strength

Table 51 lists summary and test statistics for compressive strength for the equality of variance test among sublots.

Table 51. Descriptive Statistics for Compressive Strength, kPa

Project	Lot	Mean	Std. Dev.	n	Var. Test	df
IL (24,100 Min.)	1	34,030	3640	60	4.680*	11,11
	2	41,740	3180	46	7.69*	5,9
MN (NR)	1	38,600	3010	30	57.14*	5,3
	2	34,280	1990	26	4.94	5,5
OH (NR)	1	29,020	2160	50	3.13	9,9
	2	29,990	2600	50	18.84*	9,9

Note: NR = No requirement

The sublot variances were not equal for IL (range of standard deviations 1048 to 3903 kPa), MN (407 to 3075 kPa), and OH (669 to 2903 kPa). The sampling plan and bias and precision calculations should be project-specific. The maximum standard deviation among sublots is used: IL (3903 kPa), MN (3075 kPa), and OH (2903 kPa).

Table 52 lists ranges of required sample size per sublot for a range of bias values for each project. These bias values are in terms of kPa.

Table 52. Required Sample Size to Attain Specific Bias Values for Compressive Strength, kPa

Bias	IL	MN	OH
55	76	60	57
62	60	48	45
69	49	39	36
76	40	32	30
83	34	27	25
90	29	23	22
97	25	20	19

Table 53 lists the bias values for the sample sizes used in the study. As an example, IL lot 1, with a sample size of 60, yields a bias of 62 kPa, and lot 2, with 46 sample areas, had a bias of 71 kPa.

Unit Weight

Table 54 lists summary and test statistics for a limited amount of unit weight data available from Illinois and Minnesota for equality of variance test among sublots. The ratio of variances is presented in the table, and an * indicates that the variances were not equal at the 5-percent level of significance.

Table 53. Bias Values for Compressive Strength, kPa

Site	Lot	n	Bias
IL	1	60	62
	2	46	71
MN	1	30	78
	2	46	83
OH	1	50	59
	2	50	59

Table 54. Descriptive Statistics for Unit Weight, kg/m^3

Project	Lot	Mean	Std. Dev.	n	Var. Test	df
IL	1	2311	25.6	15	17.36*	2,2
	2	2342	28.8	15	45.56*	2,2
MN	1	2287	-	5	-	-
	2	2246	-	5	-	-
OH	NO DATA					

The sublot variances were not equal for IL (range of standard deviations 6.4 to 43.3 kg/m^3). The maximum standard deviation among sublots is thus used for IL (43.3 kg/m^3). There was only one sample per sublot for MN, so no estimate of sublot variability was available. The maximum standard deviation (35.2 kg/m^3) is used to develop the sampling plan and bias and precision calculations for MN. No data were available for OH. The standard deviation from the CCRL data base for unit weight is 22.4 kg/m^3.[89]

Table 55 lists ranges of sample size per sublot for a range of bias values for the 5-percent level of significance.

Table 55. Required Sample Size to Attain Specific Bias Values for Unit Weight, kg/m^3

Bias	IL	MN
16	23	19
20	15	13
24	11	9
28	8	7
32	6	5

Table 56 lists the bias values for the sample sizes used in the study. As an example, IL lot 1 with a sample size of 15, yields a bias of 20 kg/m^3.

Table 56. Bias Values for Unit Weight, kg/m^3

Site	Lot	n	Bias
IL	1	15	20
	2	15	20
MN	1	5	31
	2	5	31

PCC Field Result Analysis

The quality characteristics measured are core compressive strength and core and GPR thickness.

Core Compressive Strength

Table 57 lists summary and test statistics for core compressive strength for the equality of variance test among sublots. The ratio of variances is presented in the table, and an * indicates that the variances were not equal at the 5-percent level of significance.

Table 57. Descriptive Statistics for Core Compressive Strength, kPa

Project	Lot	Mean	Std. Dev.	n	Var. Test	df
IL (NR)	1	36,280	7110	15	52.72*	2,2
	2	46,180	3140	15	117.6*	2,2
MN (26890 Min.)	1	43,250	4590	16	85.5*	2,2
	2	39,480	5490	15	70.4*	2,2
OH (NR)	1	40,340	3500	15	38.26*	2,2
	2	39,860	4100	15	8.36*	2,2

Note: NR = No requirement

The sublot variances were not equal for IL (range of standard deviations 441 to 11,315 kPa), MN (924 to 7750 kPa), and OH (889 to 5502 kPa). Therefore, the sampling plan and bias and precision calculations should be project-specific. The maximum standard deviation among sublots was used: IL (11,315 kPa), MN (7750 kPa), and OH (5502 kPa).

Table 58 lists ranges of required sample size per subsection for a range of bias values for the 5-percent level of significance.

Table 58. Required Sample Size to Attain Specific Bias Values
for Core Compressive Strength, kPa

Bias	IL	MN	OH
70	141	97	69
100	63	43	30
140	35	24	17
170	23	16	11
210	16	11	8

Table 59 lists the bias values for the sample sizes used in the study. As an example, IL lots 1 and 2, with a sample size of 15, yield a bias of 211 kPa.

Table 59. Bias Values for Core Compressive Strength, kPa

Project	Lot	n	Bias
IL	1	15	212
	2	15	212
MN	1	16	167
	2	15	174
OH	1	15	147
	2	15	147

Core Thickness

Table 60 lists statistics and test statistics for core thickness for the equality of variance test among sublots. The ratio of variances is presented in the table, and an * indicates that the variances were not equal at the 5-percent level of significance.

Table 60. Descriptive Statistics for Core Thickness, mm

Project	Lot	Mean	Std. Dev.	n	Var. Test	df
IL (240)	1	244.1	6.1	15	12.25*	2,2
	2	250.7	4.3	15	17.36*	2,2
MN (200)	1	206.2	8.1	21	15.12*	4,3
	2	208.5	6.9	20	36.0*	3,3
OH (280)	1	319.9	10.9	15	22.4*	2,2
	2	307.3	9.4	15	22.56*	2,2

The sublot variances were not equal for IL (range of standard deviations 1.5 to 6.4-mm), MN (1.5 to 9.1-mm), or OH (3.0 to 18.0-mm). Therefore, the sampling plan and bias and precision calculations should be project-specific. The maximum standard deviation among subsections is used: IL (6.4 mm), MN (9.1 mm), and OH (18.0 mm).

Table 61 lists ranges of required sample sizes per sublot for a range of bias values for each project.

Table 61. Required Sample Size to Attain Specific Bias Values for Core Thickness, mm

Bias	IL	MN	OH
5.0	54	77	153
7.5	24	34	68
10.0	14	20	38
12.5	9	13	24
15.0	6	9	17
17.5	4	7	12
20.0	3	5	10

Table 62 lists the bias values for the sample sizes used in the study. As an example, IL lot 1, with a sample size of 15, yields a bias of 9.7 mm.

Table 62. Bias Values for Core Thickness, mm

Site	Lot	n	Bias
IL	1	15	4.1
	2	15	10.7
MN	1	21	6.2
	2	20	8.5
OH	1	15	39.9
	2	15	27.3

GPR Thickness

Table 63 lists summary and test statistics for GPR thickness for the equality of variance test among sample areas. The ratio of variances is presented in the table; an * indicates that the variances were not equal at the 5-percent level of significance.

Table 63. Descriptive Statistics for GPR Thickness, mm

Site	Lot	Sublot	Mean	Std. Dev.	n	Var. Test	df
IL	1	1	246	2.3	36	3.61*	8,8
		2	246	4.6	160	14.89*	39,39
		3	246	2.3	160	2.934*	39,39
		4	246	3.0	160	13.14*	39,39
		5	246	3.6	156	2.85*	39,39
	2	1	249	1.0	160	1.58	39,39
		2	249	1.5	160	7.08*	39,39
		3	249	2.0	156	9.83*	39,39
		4	249	2.5	145	21.75*	39,39
		5	249	1.5	157	1.6	40,40
MN	1	1	198	9.1	160	2.72*	39,39
		2	190	13.5	153	3.65*	39,39
		3	196	19.0	156	5.61*	39,39
		4	193	11.4	160	3.75*	39,39
		5	193	14.0	158	3.61*	39,39
	2	1	190	9.4	160	17.21*	39,39
		2	180	13.5	160	4.12*	39,39
		3	188	9.1	160	7.22*	39,39
		4	185	11.2	160	20.90*	39,39
		5	188	12.2	160	29.4*	39,39
OH	1	1	297	9.7	120	1.5	39,39
		2	302	8.4	120	3.61*	39,39
		3	307	12.2	120	1.7	39,39
		4	305	15.2	120	3.24*	39,39
		5	310	11.7	120	1.094	39,39
	2	1	302	11.4	120	4.86*	39,39
		2	300	22.6	120	12.25*	39,39
		3	305	11.7	120	2.82*	39,39
		4	310	10.2	120	2.38*	39,39
		5	305	8.9	108	1.39	39,39

The sublot variances were not equal for IL (range of standard deviations 1.0 to 4.6 mm), MN (9.1 to 19.0 mm), and OH (8.4 to 22.6 mm). Therefore, the sampling plan and bias and precision calculations should be project-specific. The maximum standard deviation among sublots is used: IL (4.6 mm), MN (19.0 mm), and OH (22.6 mm).

The sampling plan allowed an analysis of variability both longitudinally and transversely. The maximum longitudinal variability tended to be at 3.2 m from the lane edge and the minimum (when it was not zero) at 1.4 m from the lane edge. Measurements at 0.5 m often were constant and hence showed no variability. At this transverse position (0.5 m), one might consider recommending a sample size of two. A comparison of the means between transverse offsets for each sublot indicates that the means between transverse offsets are significantly different for all of the sublots on all of the projects.

Table 64 lists ranges of required sample size per lot and transverse position for a range of bias values for the 5-percent level of significance with 80-percent power of test for each project.

Table 64. Required Sample Size to Attain Specific Bias Values for GPR Thickness, mm

Bias	IL	MN	OH
5	16	154	65
10	4	38	17
15	3	17	8
20	2	10	4
25	2	6	3

Table 65 lists the bias values for the sample sizes used in the study. As an example, IL lot 2, with a sample size of 40, yields a bias of 3.3.

Table 65. Bias Values for GPR Thickness, mm

Site	Lot	n	Bias
IL	1	9	6.9
	2	40	3.3
MN	1	40	9.9
	2	40	9.9
OH	1	40	6.5
	2	40	6.5

A regression analysis was performed on the GPR data to discern the effect of transverse offset on pavement thickness. Table 66 provides the analysis of variance results from this analysis. Table 67 provides the results for each of the parameters included in the model. As can be seen, the linear regression is significant. The F ratio (15144) indicates that a significant amount of the variability in all of the thickness data is explained by the differences in projects, lots, and transverse offsets. The "Prob > F" indicates the probability of getting an F ratio of this size if very little of the variation was explained by the differences in projects, lots, sublots, and transverse offsets.

Table 66. Analysis of Variance Results from Linear Regression of GPR Thickness for PCC Projects

Source	df	Sum of Squares	Mean Square	F Ratio
Model	5	7412476.0	1482495	15144.06
Error	3819	373852.7	98	**Prob > F**
Total	3824	7786328.8		0.0000

Table 67. Results for Parameter Estimates Used in Linear Regression of GPR Thickness for PCC Projects

Term	Estimate	Std Error	t Ratio	Prob > |t|
Intercept	245.0	0.3	754.79	0.0000
Site[IL-OH]	-1.4	0.2	-6.15	<0.0001
Site[MN-OH]	-55.1	0.2	-236.6	0.0000
Site[IL-OH]*Lot[1-2]	-1.5	0.2	-7.12	<0.0001
Site[MN-OH]*Lot[1-2]	2.2	0.2	9.52	<0.0001
Transverse Offset	0.7	0.1	11.88	<0.0001

As for the HMAC OK1 project, the use of cores and GPR allows for comparisons between thicknesses obtained by the two techniques. Table 68 contains the F and t statistics of each sublot. The F-test was used to examine the differences in the variances between sublots, and the t-test was used to examine the differences in the means between sublots. The asterisk indicates significant differences in either the variance or the means. The degrees of freedom also are provided for each of the sublots. The numbers are presented for the F statistic; hence, the number on the left is the number of degrees of freedom for the numerator of the F ratio and the number on the right is the number of degrees of freedom for the denominator of the F ratio. By adding these two numbers together, the degrees of freedom for the t-statistic are obtained. As seen in the table, there are numerous differences between the two measurement types, particularly between the means, indicating that the cores and GPR often estimate different populations.

Another analysis was undertaken to further discern the differences between the two measurement types. All three of the projects were compiled and a linear regression was performed between the averages for the individual sublots for the GPR thickness versus the core thickness. The results from this analysis are provided in figure 21. As can be seen, when the data from all three projects are used, the GPR does a good job of estimating the pavement thickness with an r-value

Table 68. Results from Comparisons of GPR and Core Thickness for the PCC Projects

Project	Lot	Sublot	df	F	t
IL	1	1	2,35	3.48	3.08*
		2	159,2	9.92	0.26
		3	2,159	3.02	5.66*
		4	2,159	1.68	1.46
		5	2,155	2.14	3.28*
	2	1	2,159	42.12*	0.16
		2	2,159	16.02	0.27
		3	2,155	1.97	2.48*
		4	144,2	3.03	1.80
		5	156,2	1.10	3.68*
MN	1	1	159,3	17.74*	7.87*
		2	152,3	10.31	1.10
		3	155,3	9.15	0.52
		4	159,3	14.33*	7.76*
		5	157,4	2.45	3.58*
	2	1	159,3	6.02	5.02*
		2	159,3	2.09	4.04*
		3	159,3	8.04	4.99*
		4	159,3	46.88*	19.76*
		5	159,3	9.59	1.92
OH	1	1	2,119	2.77	4.52*
		2	2,119	1.46	2.81*
		3	2,119	2.18	1.86
		4	119,2	15.50	1.03
		5	116,2	8.96	1.85
	2	1	2,122	1.58	0.09
		2	119,2	4.81	0.17
		3	119,2	15.53	0.69
		4	119,2	1.97	0.41
		5	107,2	4.09	2.17*

of 0.98. As mentioned previously, an r-value of 0 indicates that no correlation exists, while a value of 1.0 indicates a perfect correlation.

Because the correlation of all projects was good, the projects were examined individually. The same linear regression was performed for the GPR thickness versus the core thickness using the average values for the sublots for each project. Figure 22 contains the results for the IL project. The r-value was 0.74. The root mean square error indicates that the error in the regression is approximately 1 mm. These values indicate that the GPR can do a fairly good job of estimating the pavement thickness. However, when the parameter estimates are reviewed, it can be seen that the GPR thickness needs to be corrected through a correlation with cores or other measures to provide a more accurate estimate of the thickness.

Figures 23 and 24 contain the results for the MN and OH projects, respectively. Both of these projects had very low r-values, indicating the lack of a correlation; regressions returned were statistically significant. These poor results indicate that the GPR did not provide the same estimated thickness as the cores. As stated for the HMAC analysis, further evaluation needs to be undertaken to fully determine the potential of GPR.

CONFORMAL INDICES FOR PCC PROPERTIES

As for the HMAC projects, the data collected in this study allowed for the computation of CI values for some properties on the three projects. The CI values for the quality characteristics measured on the PCC projects are shown in table 69. The lowest CI values were those for plastic air content. This is an indication that concrete producers can control this property very well. Both cylinder and core strengths produce comparable CI values although they were measured on two different projects. For thickness, overall, both core and GPR values tended to be comparable in magnitude but both varied appreciably from project to project.

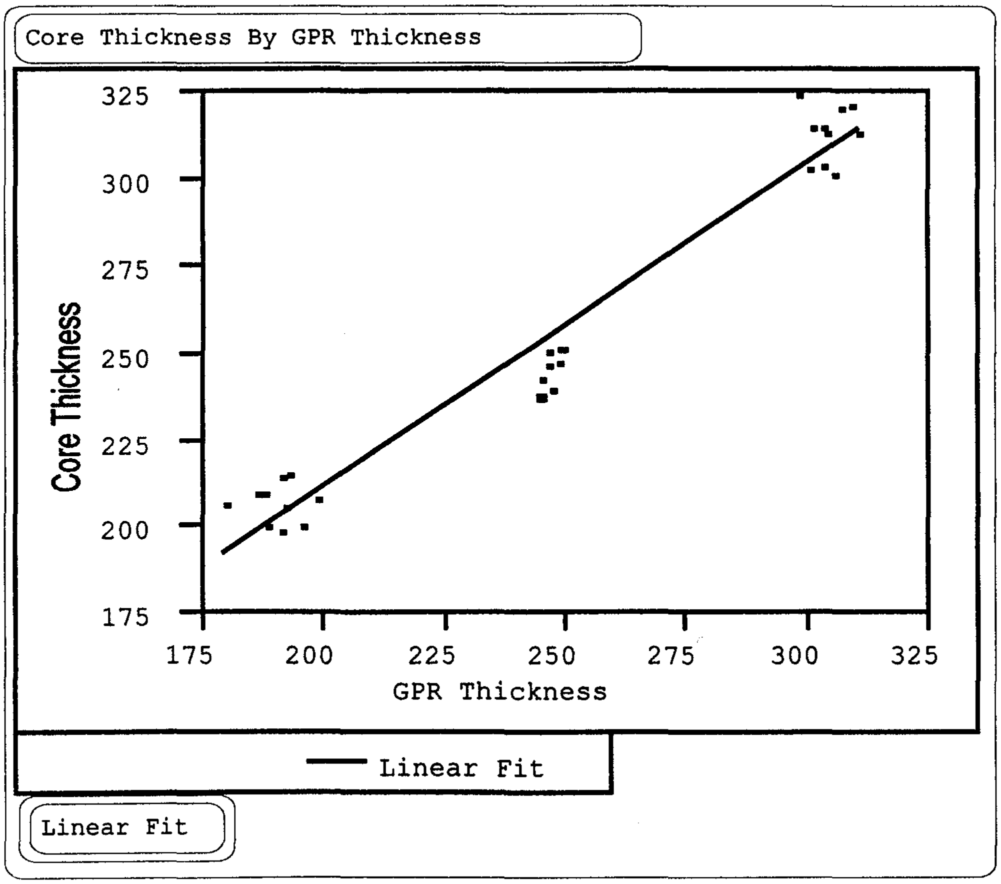


Figure 21. Linear Regression of GPR Thickness Versus Core Thickness for All PCC Projects

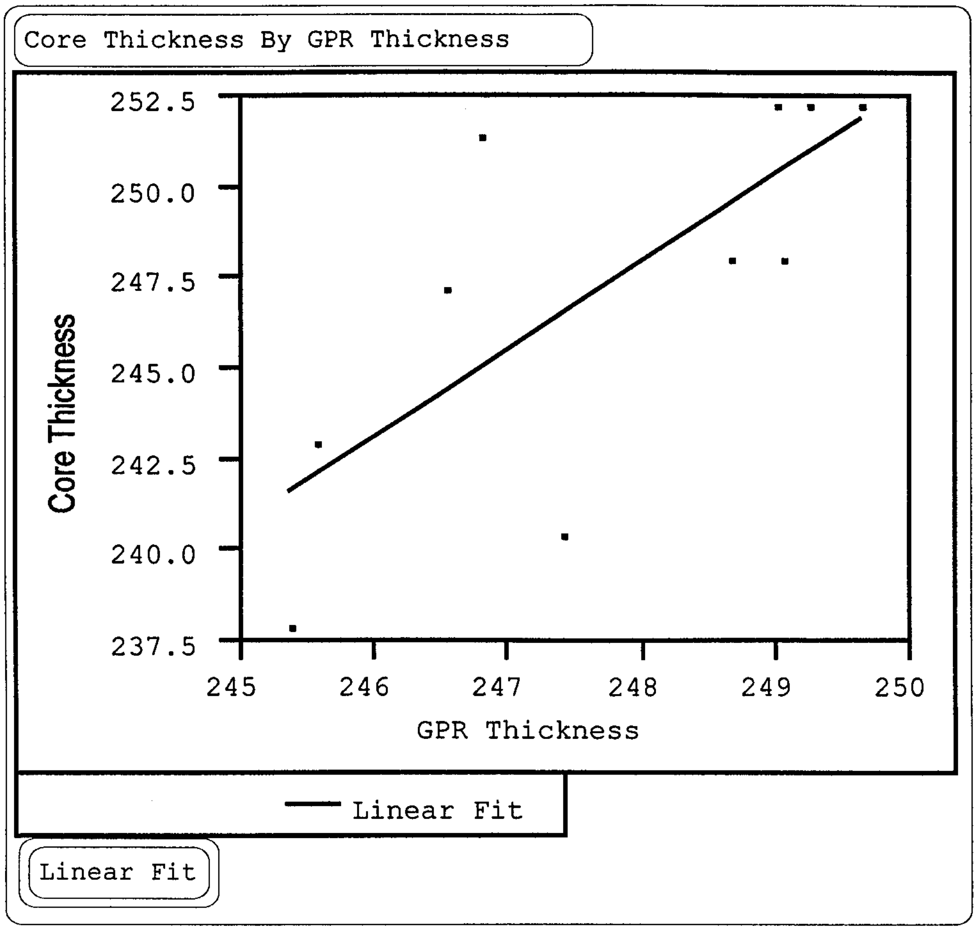


Figure 22. Linear Regression of GPR Thickness Versus Core Thickness for Project IL

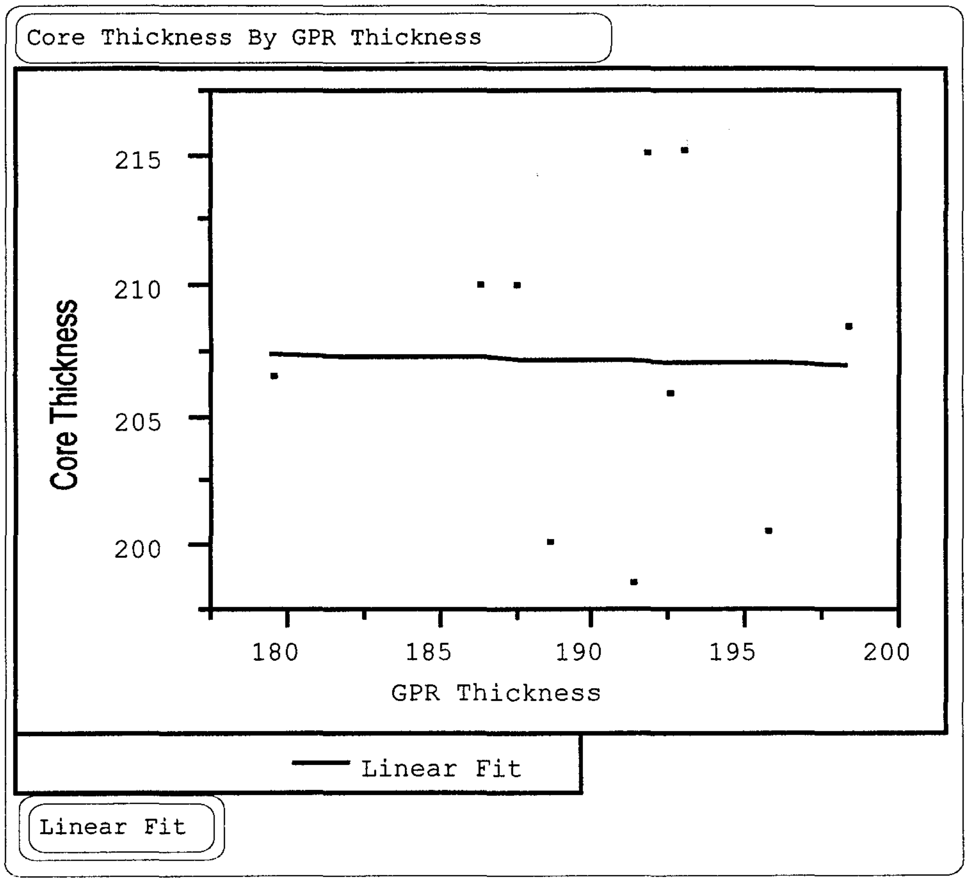


Figure 23. Linear Regression of GPR Thickness Versus Core Thickness for Project MN

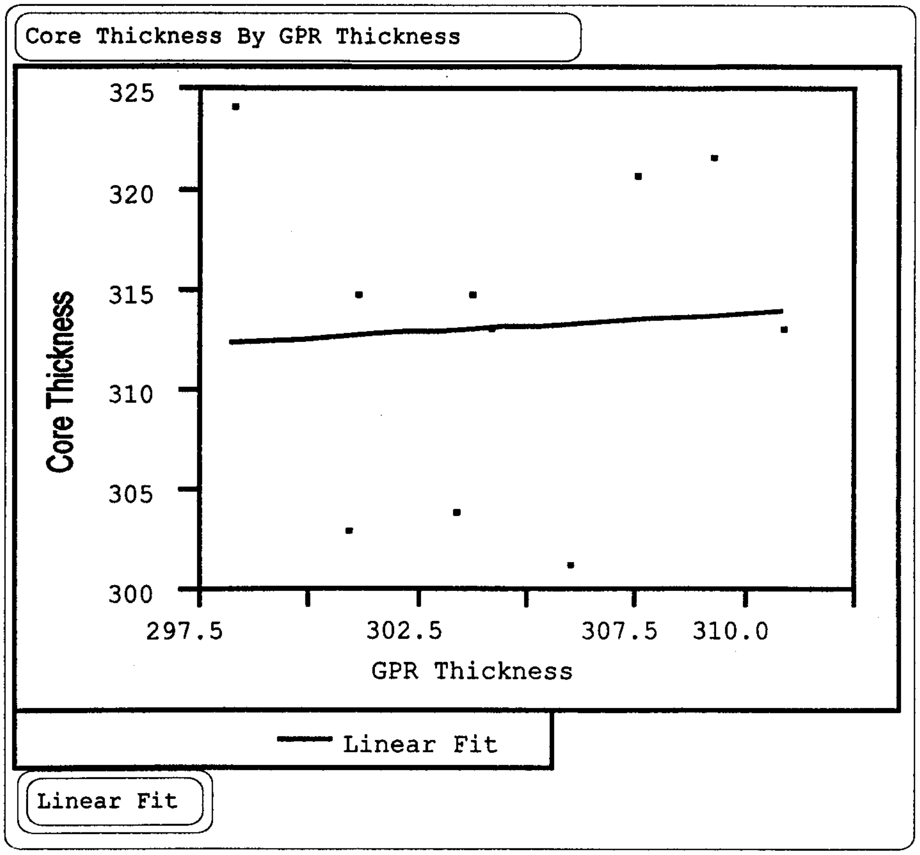


Figure 24. Linear Regression of GPR Thickness Versus Core Thickness for Project OH

Table 69. Conformal Indices for PCC Projects

Project	Lot	Plastic Air Content, %	Cylinder Compressive Strength,MPa	Core Thickness, mm	Core Compressive Strength, MPa	GPR Thickness, mm
IL	1	1.0	10.5	6.4	NR	6.1
	2	2.1	17.9	10.2	NR	8.0
MN	1	0.7	NR	8.5	17.0	16.9
	2	0.5	NR	8.5	13.7	14.8
OH	1	1.4	NR	41.0	NR	27.5
	2	0.8	NR	29.4	NR	29.3

Note: NR = No Requirement

CORRELATION OF INTERRELATED PROPERTIES OF PCC PROJECTS

As was done in the HMAC analysis, a correlation between each of the PCC properties was done. There were a few properties that produced reasonably good correlation coefficient (r) values. For this analysis the degrees of freedom vary from property to property because the number of values vary. For the smallest degrees of freedom (n = 8), an r-value of 0.765 indicates a significant correlation at an α-value of 1 percent; for 28 degrees of freedom, the significant r is 0.46; for 40 degrees of freedom and higher, the significant r is, for all practical purposes, 0.4. The properties that have r-values equal to or greater than these are shown in table 70.

Among the strongest correlations were, as expected, those between cylinder and core compressive strengths on the IL and MN projects. It is not known why a reasonable correlation coefficient (only -0.021) was not found on the OH project. Also, unit weight correlated well with core strength on both the IL and MN projects, but not with cylinder strength on any project. Thickness had a reasonable correlation with cylinder strength but not core strength on the MN project. Finally, air content produced a significant r-value with cylinder strength, core strength, and unit weight on the IL project, and with thickness on the OH project.

Table 70. Significant Correlation Coefficient Values of PCC Variables

Variable	Core Strength	Thickness	Air Content
Project IL			
Cylinder Strength	0.84		-0.56
Air Content	-0.65		
Unit Weight	0.66		-0.77
Project MN			
Cylinder Strength	0.74	-0.63	
Unit Weight	0.79		
Project OH			
Air Content		-0.62	

RELATIONSHIP OF PAVEMENT VARIATION WITH PERFORMANCE

One of the objectives of this work was to relate the variability of the pavement properties with pavement performance. The sampling and testing plan included a round of FWD testing on the finished surface of each of the three HMAC test sections. No monitoring has been performed on these sections since they were constructed; therefore, the only measure of performance available is the FWD testing.

A linear regression was performed comparing sensor 1 minus sensor 3 of the FWD data to the thickness of the HMAC surface layer. It is believed that sensor 1 minus sensor 3 is representative of the stiffness of the HMAC surface layer. Correlations have been made between the base curvature index (BCI) and the subgrade stiffness and between surface curvature index (SCI) and granular base stiffness, which were highly significant.[90] The correlation coefficient of this analysis was a fairly low value of 0.48. However, the regression was statistically significant, with an F Ratio of 8.2983 and a probability of 0.75 percent. This significance indicates that there is a relationship between the sensor 1 minus sensor 3 and the thickness of the HMAC even if the thickness of the HMAC surface layer does not explain all of the variation seen in sensor 1 minus sensor 3. The slope of the regression line was a positive 3.33, indicating that as the thickness increases, the difference between sensor 1 and sensor 3 increases.

A further comparison was made between sensor 1 minus sensor 3 and the coefficient of variation of the thickness of the HMAC layer. The correlation coefficient for this regression was better than the one for the previous analysis at 0.64. The regression also was statistically significant, with an F Ratio of 19.6010 and a probability of 0.01 percent. The slope of the regression line was -4.88, indicating that as the coefficient of variation in thickness increases, the difference between sensor 1 and sensor 3 decreases. The two regressions indicate that while the FWD may be a good test of the variability of the pavement for thicker pavements, it should not be used for thin pavements.

A similar set of analyses performed on the PCC sections produced opposite results. The first regression compared the difference between sensor 1 and sensor 3 of the FWD results with the thickness of the PCC layer. The correlation coefficient for this analysis was a moderately successful 0.69. The F ratio for the regression was 24.9049, with a probability value of less than 0.01 percent. The slope of the regression line was -0.08, indicating that as the thickness increases, the difference between sensor 1 and sensor 3 (or the stiffness of the PCC layer) decreases. However, a slope this small is probably within the precision limits of the FWD equipment. While the statistics may be significant, the result probably is not significant.

The second regression between the difference between sensor 1 and sensor 3 of the FWD results and the coefficient of variation of the thickness of the PCC layer yielded a correlation coefficient of 0.39 with an F ratio of 5.1268 and a probability value of 3.15 percent. The slope of the regression line was 0.81, indicating that as the coefficient of variation of the thickness of the PCC layer increases, the difference between sensor 1 and sensor 3 increases. In other words, as the variability of the thickness increases, the stiffness of the PCC layer decreases.

PROPOSED PROCEDURE FOR SEQUENTIAL ADDITION TO LOTS

Before the days of sophisticated statistical computer programs, the comparison of the magnitude of the mean to that of the median was used as an approximation of normality. This comparison is examined in this study in an attempt to find a "quick and easy" way of estimating when a distribution is approximately normal. This may lead to a quick field test of normality on a lot-by-lot or sublot-by-sublot basis.

Large lot sizes have the advantage of containing relatively large sample sizes which allow the reduction of both buyer and seller risks. A primary disadvantage of large lots is that the likelihood of changes increase as the project continues, which increases the probability that the population will become non-normal. The procedure proposed here is a method to quickly check the normality of the sublot or lot population and, when found to be reasonably normal, to allow additional samples to be added to the previous sublot or lot. It can be used with sublots whenever the sublot contains three or more test results; otherwise, it would apply to a lot being combined with a previous lot. The discussion below uses the assumption that three or more results are available in a sublot. If this is not applicable, "lot" will replace "sublot" in the application.

Two steps are required in the procedures; the first step is to compare the median of a sublot of a set of data with the mean, and if the median is within 2 percent of the mean, to assume normality within the sublot; the second step is to compare the mean of the sublot with that of the previous sublot, and if the means are within 2 percent, to allow the sublot data to accumulate in the lot. If in the first step the sublot is non-normal, the decision of specification compliance for the sublot will have to be based on the statistical parameters of that individual sublot. Although this procedure is somewhat crude, it is amenable to field usage without sophisticated statistical analyses and can be refined if desired. To keep the procedure sufficiently simple to use in a field application, each sublot (lot) is judged individually; determining the median of a large sample size in the field becomes onerous. The values of ±2 percent were tested for one quality characteristic that appeared to contain a significant shift in properties between lots.

The example of this procedure is provided for the nuclear density tests from the left wheelpath of the LA project. A sublot of n = 5 was chosen because it generally produces a reasonable bias estimate for most quality characteristics and it allows for a quick determination of the median. The analysis for this proposed procedure is shown in table 71.

Table 71. Example of Proposed Procedure for Approximating Normal Distribution by Sublot

Lot No	Sublot No	Sublot Mean	Sublot Median	% Diff Median	Add to Lot	% Diff Mean	Add to Lot	Lot	
								Mean	Std Dev
1	1	95.8	96.4	0.6	Y	-	-	95.8	1.69
	2	95.6	95.0	0.6	Y	0.2	Y	95.7	1.70
	3	96.7	96.4	0.3	Y	1.1	Y	96.0	1.48
	4	95.5	94.6	0.9	Y	1.3	Y	95.6	1.54
2	1	88.1	88.7	0.7	Y	8.4	N	88.1	1.60
	2	88.4	89.6	1.4	Y	0.3	Y	88.4	1.93
	3	87.6	87.5	0.1	Y	0.1	Y	86.7	1.94
	4	87.5	87.5	0.0	Y	0.1	Y	87.5	1.53

The explanation for table 71 is as follows:

- The percent difference median is the absolute value of the mean minus the median divided by the mean for the current sublot.
- Add to lot is the decision as to whether either of the percent differences of the two comparisons exceed 2.0 percent.
- The percent difference mean is the absolute value of the difference of the preceding sublot mean from the current sublot mean divided by the sublot mean.

The first four sublots do not exceed the 2.0 percent limit so each is accumulated in the lot mean and standard deviation statistics. When the fifth sublot is tested, the percent difference median does not exceed the criterion, meaning that the sublot is judged to be relatively normal; however, the percent difference mean is greater than 2.0 percent, indicating a significant shift in the mean from the preceding sublot. Therefore, this sublot cannot be accumulated with the preceding sublot and a new lot must be started.

While this procedure looks promising, additional analysis under varied conditions should be done.

CHAPTER 7. CONCLUSIONS AND RECOMMENDATIONS

The following conclusions are drawn from this study.

- Most quality characteristics measured in this study tend to be normally distributed. This is confirmed through such tests as the Shapiro-Wilk test, through normal distribution plots, and by the relatively small magnitude of skewness and kurtosis values found.
- For several of those cases in which the quality characteristic was not normally distributed, there tended to be an explanation. One reason was that the mean of the characteristic was close to a physical limit, which allowed only one tail of the distribution to develop. Another reason was the existence of apparent outliers. There were some distributions that tended to be bimodal, which can be caused by a shift in the process.
- The computer program SKEWBIAS, developed in this study, has the capability to establish the bias when estimating PD for populations that exhibit various degrees of skewness. SKEWBIAS indicates that skewness values as large as about 1.0 produce relatively small bias values. On the other hand, skewness values of 1.5 or higher tend to produce larger bias values.
- Correlations between variables indicated several interrelated properties, especially for HMAC mixtures.
    - For HMAC mixtures, several adjacent sieve sizes were found to be interrelated, indicating that only one sieve in a range of gradation is necessary for acceptance testing. Although not specifically addressed in this study, the use of more sieves for QC may be advisable to produce sufficient information on which to direct corrective action when the process gets out of control.
    - For PCC materials, the variables that had the strongest statistical correlations on most projects were cylinder and core compressive strength, and unit weight and core compressive strength.
- This study quantified the relationship between sample size and bias for many quality characteristics. The analysis quantified that relatively large sample sizes are required to

detect smaller biases and that small sample sizes can detect only large biases. This study has produced typical sample sizes and biases that SHA's can use to assess the cost-effectiveness of this relationship.

- The CI was used to compare the ability to hit the target consistently.
  - For HMAC mixtures, the most notable differences in CI values between projects were in GPR thickness and gradation, particularly the 2.36-mm and 1.18-mm sieves.
  - For PCC, the low CI for plastic air content indicates that concrete producers can control this property well. Both core and cylinder compressive strengths, although obtained on different projects, had very similar CI values. CI values of thickness, whether measured by cores or GPR, tended to have similar magnitudes but varied appreciably from project to project.
- In establishing specification limits for various material quality characteristics, the precision of the standard test procedure used to measure the characteristics of interest should be taken into account so as not to set too narrow of a target range that cannot be practically achieved given limitations of test procedures.

The following recommendations are made on the basis of findings of this study and the consideration of needs in the measurement and specification of construction quality.

- The procedures suggested in this study to assess normality of lots and sublots in order to allow data to be accumulated in larger lots needs to be tested under more conditions than was possible in this study. A computer program that uses Monte Carlo sampling can be written to further study this procedure.
- The computer program SKEWBIAS should be examined more extensively for conditions such as specifications with double limits.
- Nondestructive tests that are quick have been sought for many years. Consideration should be given to the application of more nondestructive testing procedures such as GPR, FWD, and, although not used in this study, laser measurement of rideability for the measurement of construction quality. Specifically for GPR thickness, further

examination is essential because the correlation with core thickness was not good. However, the FWD testing provided promising results for its use in measuring pavement variability, specifically for the PCC pavements and for thicker HMAC pavements.

- Although not addressed in this study but found in the literature search, previous studies have provided a list of quality characteristics that have the greatest apparent impact on performance. Among these are air voids, HMAC thickness, base thickness, and compaction for HMAC pavements; and thickness, joint spacing, and percent steel for PCC pavements. Further study needs to be conducted to determine how variations in these and possibly other characteristics affect pavement performance.

REFERENCES

1. *AASHTO Quality Assurance Guide Specification*, American Association of State Highway and Transportation Officials, AASHTO Quality Construction Task Force, February 1996.

2. *Implementation Manual for Quality Assurance*, American Association of State Highway and Transportation Officials, AASHTO Quality Construction Task Force, February 1996.

3. "Glossary of Highway Quality Assurance Terms," *Transportation Research Circular 457*, Transportation Research Board, Washington, D.C., 1996.

4. Von Quintus, H.L., T.W. Kennedy, and J. Epps, *Operational and Performance Characteristics of Drum Mix Plants*, Report No. FHWA-TS-84-212, Federal Highway Administration, October 1984.

5. Hughes, C.S., "Incentive and Disincentive Specification for Asphalt Concrete Density," *Transportation Research Record No. 986*, Transportation Research Board, Washington, D.C., 1984.

6. Hughes, C.S., "Variability in Highway Pavement Construction," *NCHRP Synthesis of Highway Practice 232*, Transportation Research Board, National Research Council, Washington, D.C., 1996.

7. Weed, R.M. *Highway Quality Assurance: The Need for Science-and-Knowledge-Based Decision Making*, Report No. NJ-97-001-7490, New Jersey Department of Transportation, Trenton, NJ, 1997.

8. Von Quintus, H.L., C.S. Hughes, J.A. Scherocman, and T.W. Kennedy, *NCHRP Report 338: Asphalt Aggregate Mixture Analysis System - AAMAS*, Transportation Research Board, January 1990.

9. Hudson, S.B., F.T. Higgins, and F.J. Bowery, *Determination of Statistical Parameters for Bituminous Concrete*, Research Report 68-14, Pennsylvania Department of Transportation, Harrisburg, PA, October 1972.

10. Gonsalves, G.F.D. and J. Eisenberg, *Implication of Statistical Quality Control of Portland Cement Concrete*, Arizona Department of Transportation Research Report No. HPR-1-12-P152, May 1975.

11. Okamoto, P.A., C.L. Wu, S.M. Tarr, M.I. Darter, and K.D. Smith, *Performance-Related Specifications for Concrete Pavements - Volume III*, Report No. FHWA-RD-93-044, Federal Highway Administration, Office of Engineering and Highway Operations R&D, November 1993.

12. New Hampshire DOT, *Special Provision Amendment to Section 106 - Control of Materials*, Section 106.03.01, May 3, 1994.

13. California DOT, *Special Provisions for Construction on State Highways*, Section 10-1.15, January 11, 1993, Page 29.

14. South Dakota DOT, *Special Provisions Regarding Quality Control/Quality Assurance Specifications for Asphalt Concrete Pavement, Part I, Contractor-Furnished Quality Control Program*, January 24, 1994, Page 1-2.

15. Kandahl, P.S., R.J. Cominsky, D. Maurer, J.B. Motter, "Development and Implementation of Statistically-Based End Result Specifications for Hot-Mix Asphalt in Pennsylvania," *Transportation Research Record 1389*, Transportation Research Board, Washington, D.C., 1993.

16. New Hampshire DOT, *Special Provision: Amendment to Section 401 - Plant Mix Pavements*, Section 403.3.13.1, June 29, 1994.

17. California DOT, *Special Provisions for Construction on State Highway*, Section 39-8.02, January 11, 1993, Page 33.

18. Huft, D.L., "Analysis and Recommendations Concerning Profilograph Measurements on F0081(50)107 Kingsbury County," *Transportation Research Record 1348*, Transportation Research Board, Washington, D.C., 1991.

19. *Special Provisions for Furnishing Portland Cement Concrete Pavement (Quality Assurance)*, Michigan Department of Transportation, Bureau of Highways, Draft, November 1991.

20. *QC/QA Specification for Portland Cement Concrete, Division 500 - Rigid Pavements*, West Virginia Department of Transportation.

21. *Standard Specifications*, Washington State Department of Transportation, 1991.

22. *Quality Assurance for Portland Cement Concrete Pavements*, Report No. 501-R-187, Indiana Department of Transportation, 1990.

23. Ahlrich, R.C. "Investigation of Unstable Asphalt Concrete Airfield Pavement," *Transportation Research Record 1300*, Transportation Research Board, Washington, D.C., 1991.

24. Dager, J., "Batch Plants No Guarantee to Consistent Quality Mixes," *Roads and Bridges*, Volume 32, Number 5, Des Plaines, Illinois, May 1994, Page 34.

25. Brock, J.D., *Segregation Causes and Cures*, Report No. T-115, Astec Industries, Chattanooga, Tennessee, 1986.

26. Lavon, B., and M. Fradua, "When Should Concrete Cylinder Test Results be Questioned?" *Concrete International*, Volume 16, No. 12, December 1994, Pages 43-46.

27. Ksaibati, K., S. Asnani and T. Adkins, "Factors Affecting the Repeatability of Pavement Longitudinal Profile Measurements," *Transportation Research Record No. 1410*, Transportation Research Board, Washington, D.C., 1993.

28. Scofield, L.A., S.A. Kalavela, and M.R. Anderson, "Evaluation of California Profilograph," *Transportation Research Record 1348*, Transportation Research Board, 1992.

29. Darter, M.I., M. Abdelrahman, T. Hoerner, M. Phillips, K.D. Smith, and P.A. Okamoto, *Performance-Related Specifications for Concrete Pavements - Volume II*, Report No. FHWA-RD-93-043, Federal Highway Administration, Office of Engineering and Highway Operations R&D, November 1993.

30. *The Special Report 61B: AASHO Road Test: Report 2 — Materials and Construction*, Highway Research Board, Publication 951, National Research Council, 1962.

31. Shook, J.F., "Significance of Test Results Obtained From Random Samples," *Statistical Methods for Quality Control of Road and Paving Materials*, American Society of Testing and Materials STP No. 362, Philadelphia, PA, 1963.

32. Adam, V., and S.C. Shah, "Quality Control Analysis of Asphaltic Concrete," *Proceedings of Highway Conference on Research and Development of Quality Control and Acceptance Specifications*, Bureau of Public Roads, Office of Research and Development, April 1965.

33. *Highway Quality Assurance, Demonstration Project # 42*, USDOT, FHWA Demonstration Projects, Region 15, Washington, DC, 1976.

34. Schmitt, R.L., A.S. Hanna, J.S. Russell, and E.V. Nordheim, "Pavement Density Measurement Comparative Analysis Using Core and Nuclear Methods," Paper Presented at the Annual Meeting of the Association of Asphalt Paving Technologists, Salt Lake City, Utah, March 1997.

35. Cominsky, R.J., and P.S. Kandahl, *Statistical Acceptance of Bituminous Paving Mixtures*, PA-79-92, Pennsylvania Department of Transportation, Harrisburg, Pennsylvania, May 1982.

36. Shapiro, S.S., *How to Test Normality and Other Distributional Assumptions*, American Society for Quality Control, Statistics Division, Milwaukee, Wisconsin, 1990.

37. "Recommended Practice for Evaluation of Compression Test Results of Field Concrete," *ACI Proceedings*, ACI 214-57, July 1957.

38. Cummings, A.E., "Strength Variations in Ready-Mixed Concrete," *ACI Proceedings*, April 1955.

39. Julian, O.G., "Discussion of Paper by A.E. Cummings," *ACI Proceedings*, American Concrete Institute, April 1955.

40. Schleider, R.H., *A Study of Concrete Pavement Beam Strengths*, Construction Division, Texas Highway Department, 1959.

41. *Highway Quality Control Program - Statistical Parameters*, Research Report R-572, Michigan Department of Highways, March 1966.

42. Wood, S.L., *Evaluation of Long-Term Properties of Concrete*, Research and Development Bulletin RD102T, Portland Cement Association, Skokie, Illinois, 1992.

43. *AASHTO Design Procedures for New Pavements - Participant's Manual*, National Highway Institute, Federal Highway Administration, August 1993.

44. Balagru, P., "Comparison of Normal and Lognormal Frequency Distribution for Representing Variations of Concrete Strength," *ACI Materials Journal*, Volume 92, No. 2, March-April 1995.

45. Parker, F., Jr., E.R. Brown, R.L. Vecelio, "Development of New Criteria for Control of Hot-Mix Asphalt Construction," *Transportation Research Record 1389*, Transportation Research Board, Washington, D.C., 1993.

46. Elliott, R.P., "Quality Assurance: Specification Development and Implementation," *Transportation Research Record 1310*, Transportation Research Board, Washington, D.C., 1991.

47. Brakey, B.A., *HBP QA/QC Pilot Projects Construction in 1992*, CDOT-DTD-R-93-14, Colorado Department of Transportation, Denver, Colorado, June 1993.

48. Benson, P.E., "Comparison of End-Result and Method Specifications for Managing Quality," *Transportation Research Record 1491*, Transportation Research Board, Washington, D.C., 1995.

49. Scherocman, J.A., and H.L. Von Quintus, *The Investigation of Contractor Quality Control and End Result Acceptance Specifications of Bituminous Paving Mixtures for South Carolina*, South Carolina Department of Highways and Public Transportation, Columbia, South Carolina, March 1993.

50. Minnesota DOT, *Supplemental Specifications to the 1988 Standard Specifications for Construction*, Section 2331.3H1c, January 2, 1991.

51. South Dakota DOT, *Special Provision Regarding Quality Control/Quality Assurance Specifications for Asphalt Concrete Pavement, Part I, Asphalt Concrete Quality Control Testing*, January 24, 1994, Page 4.

52. Hughes, C.S., "State-of-the-Art: State Acceptance of Bituminous Concrete Production and Construction Using Quality Assurance Specifications," *Quality Assurance in Pavement Construction*, American Society of Testing and Materials, STP 709, Philadelphia, PA, 1978.

53. Weingarten, H., "Confidence Intervals for the Percent Nonconforming Based on Variables Data," *Journal of Quality Technology*, Vol. 14, No. 4, October 1982.

54. Darter, M.I., M. Abdelrahman, P. Okamoto and K.D. Smith, *Performance-Related Specifications for Concrete Pavements — Volume I*, Report No. FHWA-RD-93-042, Federal Highway Administration, November 1993.

55. McMahon,T.F., et al, *Quality Assurance in Highway Construction*, FHWA Report No. FHWA-TS-89-038, Washington, D.C., October 1990.

56. Miao, B.Q., P.C. Aitcin, W.D. Cook, and D. Mitchell, "Influence of Concrete Strength on Insitu Properties of Large Columns," *ACI Materials Journal*, Volume 90, No. 3, May-June 1993, Pages 214-219.

57. Gray, R.J., "Variability of Compressive Strength Test Results," *Concrete Technology Today*, Volume 11, No. 2, Portland Cement Association, June 1990, Pages 1-4.

58. Hughes, C.S., "Compaction of Asphalt Pavement," *NCHRP Synthesis of Highway Practice 152*, Transportation Research Board, October 1989.

59. Hughes, C.S., "SUPERPAVE™ Field Control Tests," *Journal of the Association of Asphalt Paving Technologists*, Volume 62, 1993.

60. Hadley, W.O., "Material Characterization and Inherent Variation Analysis of Asphaltic Field Cores," *Transportation Research Record 1317*, Transportation Research Board, Washington, D.C., 1991.

61. Ford, M.C., Jr., "Pavement Densification Related to Asphalt Mix Characteristics," *Transportation Research Record 1178*, Transportation Research Board, Washington, D.C., 1988.

62. Shook, J.F., M.A. Diaz, M. Stroup-Gardiner, S.B. Seeds, *Performance-Related Specifications for Asphalt Concrete - Phase II, Final Report*, FHWA-RD-91-070, Federal Highway Administration, McLean, Virginia, 1993.

63. Elliott, R.P., M.C. Ford, Jr., M. Ghanim, Y.F. Tu,, "Effect of Aggregate Gradation Variation on Asphalt Concrete Mix Properties," *Transportation Research Record 1317*, Transportation Research Board, Washington, D.C., 1991.

64. Solaimanian, M., T.W. Kennedy, *Evaluation of Field Compaction, Density Variations and Factors Affecting Density Through 1987 HMAC Field Construction Data - Final Report*, FHWA/TX-92+486-4F, Texas State Department of Highways and Public Transportation, Washington, D.C., 1989.

65. Huber, G.A., and T.S. Shuler, "Providing Sufficient Void Space for Asphalt Cement: Relationship of Mineral Aggregate Voids and Aggregate Gradation," *Effects of Aggregates and Mineral Fillers on Asphalt Mixture Performance*, ASTM STP 1147, Philadelphia, Pennsylvania, 1992.

66. Simpson, A.L., B.R. Rauhut, P.R. Jordahl, E. Owusu-Antwi, M.I. Darter, O.J. Pendleton, Y.H. Lee, *Sensitivity Analyses for Selected Pavement Distresses*, SHRP-P-393, Strategic Highway Research Program, National Research Council, Washington, D.C., 1994.

67. Eltahan, A.A., A.L. Simpson, and J.B. Rauhut, *Common Characteristics of Good and Poorly Performing AC Pavements*, Draft Report, FHWA Contract DTFH61-96-C-00003, June 1997.

68. Barnett, T.L., M.I. Darter, and N.R. Laybourne, *Evaluation of Maintenance/Rehabilitation Alternatives for CRCP*, FHWA/IL/UI-185, Illinois Department of Transportation, 1981.

69. Smith, K.D., D.G. Peshkin, M.I. Darter, A.L. Mueller and S.H. Carpenter, *Performance of Jointed Concrete Pavements - Volume 1*, Report No. FHWA-RD-89-136, Office of Engineering and Highway Operations R&D, Federal Highway Administration, March 1990.

70. Darter, M.I., J.M. Becker, and M.B. Snyder, *NCHRP Report 277: Development of a Concrete Pavement Evaluation System (COPES) - Volume 1*, Transportation Research Board, 1984.

71. Tayabji, S.D. and D. Whiting, "Field Evaluation of Concrete Pavement Consolidation," *Transportation Research Record 1110*, Transportation Research Board, Washington, D.C., 1987.

72. Mattacchione, A., and L. Mattachione, "Correlation Between 28-Day Strength and Density," *Concrete International*, Volume 17, Number 3, March 1995, Pages 37-41.

73. Bartlett, F.M., and J.G. MacGregor, "Equivalent Specified Concrete Strength From Core Test Data," *Concrete International*, Volume 17, Number 3, March 1995.

74. Okamoto, P.A., "Field Evaluation of Dowel Placement Along a Section of Interstate 45 in Texas," *Transportation Research Record 1186*, Transportation Research Board, 1988, Pages 16-34.

75. Ross, F.R., and T.S. Rutkowski, *A Designed System of Load Transfer Dowels for Joints in PCC Pavement*, Applied Research Section, Materials Division, Wisconsin Department of Transportation, January 1989.

76. "Standard Test Method for Bulk Specific Gravity and Density of Compacted Bituminous Mixtures Using Saturated Surface-Dry Specimens," ASTM D2726-90, *Annual Book of ASTM Standards*, Volume 4.03, Philadelphia, Pennsylvania, 1993, Pages 317-319.

77. "Standard Test Method for Theoretical Maximum Specific Gravity and Density of Bituminous Paving Mixtures," ASTM D 2041-91, *Annual Book of ASTM Standards*, Volume 4.03, Philadelphia, Pennsylvania, 1993, Pages 257-263.

78. "Standard Test Methods for Quantitative Extraction of Bitumen From Bituminous Paving Mixtures," ASTM D 2172-92, *Annual Book of ASTM Standards*, Volume 4.03, Philadelphia, Pennsylvania, 1993, Pages 284-295.

79. "Standard Test Method for Air Content of Freshly Mixed Concrete by the Pressure Method," ASTM C231-91b, *Annual Book of ASTM Standards*, Volume 4.02, Philadelphia, Pennsylvania, 1992, Pages 138-145.

80. "Standard Test Method for Specific Gravity, Absorption and Voids in Hardened Concrete," ASTM C642-90, *Annual Book of ASTM Standards*, Volume 4.02, Philadelphia, Pennsylvania, 1992, Pages 320-321.

81. "Standard Test Method for Compressive Strength of Cylindrical Concrete Specimens," ASTM C39-86, *Annual Book of ASTM Standards*, Volume 4.02, Philadelphia, Pennsylvania, 1992, Pages 20-24.

82. "Standard Test Method for Obtaining and Testing Drilled Cores and Sawed Beams of Concrete," ASTM C42-90, *Annual Book of ASTM Standards*, Volume 4.02, Philadelphia, Pennsylvania, 1992, Pages 27-30.

83. *Manual for FWD Testing in the Long Term Pavement Performance Study - Operational Field Guidelines, Version 2.0*, Federal Highway Administration, Washington, D.C., February, 1993.

84. Russell, J., A.S. Hanna, H.U. Bahia, R.L. Schmitt, and G.A. Jung, "Summary of Current Practices in QC/QA for Hot-Mix Asphalt," Paper Submitted for Presentation and Publication at the 1998 Annual Meeting of the Transportation Research Board.

85. *SAS Procedure's Guide*, Release 6.03 Edition, SAS Institute, Cary, North Carolina, 1988.

86. Snedecor, G.W., and W.C. Cochran, *Statistical Methods*, Iowa State University Press, Sixth Edition, 1967.

87. Volk, W., *Applied Statistics for Engineers*, Chemical Engineering Series, McGraw-Hill Book Company, Inc., 1958.

88. *Compilation of Statistics from the AASHTO Materials Reference Laboratory Proficiency Sample Program, From 1966 through 1996*, National Institute for Standards and Technology, Gaithersburg, MD, 1997.

89. *Compilation of Statistics from the Cement and Concrete Reference Laboratory Proficiency Sample Program, From 1966 through 1997*, National Institute for Standards and Technology, Gaithersburg, MD, 1997.

90. Von Quintus, H.L., and A. Gonzales, *Flexible Pavement Design Guidelines for Streets and Roadways; Volume 1: Data Collection and Analyses*, Bexas County and Metropolitan Planning Organization, San Antonio, TX, 1987.

BIBLIOGRAPHY

Afferton, K.C., "New Jersey's Thickness Specification for Bituminous Pavement," *Transportation Research Board Circular No. 172*, 1975.

Asphalt Pavement Profilograph Summary - 1993, Kansas Department of Transportation Interim Report, March 1994.

Determination of Statistical Parameters for Highway Construction, Research Report No. 18, State Road Commission of West Virginia, July 1968.

Greer, W.C., "Evaluation of Strength Tests and Acceptance of Concrete Pavements," *Proceedings of the Fourth International Conference on Concrete Pavement Design and Rehabilitation*, Purdue University, West Lafayette, Indiana, 1989.

Hot Bituminous Paving QA/QC Pilot Projects Constructed in 1992, Interim Report No. CDOT-DTD-R-93-14, Colorado Department of Transportation, June 1993.

Hudson, S.B., F.T. Higgins, and F.J. Bowery, *Determination of Statistical Parameters for Bituminous Concrete*, Materials Research and Development, Inc., October 1972.

Irick, P.E., S.B. Seeds, M.G. Myers, and E.D. Moody, *Development of Performance-Related Specifications for Portland Cement Concrete Pavement Construction*, Final Report, ARE Inc., PB91-174151, May, 1990.

Newlon, H.H., Jr., "Variability of Portland Cement Concrete", *Proceedings of National Conference on Statistical Quality Control*, 1966.

Ozyildirim, C., *Comparision of Air Void Content in Fresh Versus Hardened Concretes*, Final Report, VTRC 90-R23, March, 1990.

Patel, A.J., "Development of Fatigue Based Pay Adjustment Factors in Performance-Related Specifications for Full-Depth Asphalt Concrete Pavements," Ph. D. Thesis, University of Illinois, 1996.

Patel, A., M. Thompson, E. Harm, and W. Sheftick, "Development of QC/QA Specifications for Hot Mix Asphalt Concrete in Illinois," Presented to the 1997 Annual Meeting of the Transportation Research Board.

Poblete, M., R. Salsilli, R. Valenzuela, A. Bull, and P. Spratz, "Field Evaluation of Thermal Deformations in Undoweled PCC Pavement Slabs," *Transportation Research Record 1207*, Transportation Research Board, Washington, D.C., 1988.

Recommended Practice for Evaluation of Strength Test Results of Concrete, ACI 214-77, American Concrete Institute, 1989.

"Standard Specification for Ready-Mixed Concrete," ASTM C94-92, *Annual Book of ASTM Standards*, Volume 4.02, Philadelphia, Pennsylvania, 1992.

Troxell, G.E., H.E.Davis, J.W. Kelly, *Composition and Properties of Concrete*, Second Edition, McGraw-Hill Book Company.

WASHTO, *Model Quality Assurance Specifications*, Western Association of State Highway and Transportation Officials, WASHTO QA Task Force, Draft Final Report, October 1990.

Weed, R.M. *Managing Quality: Time for a National Policy*, Report 96-004-7490, New Jersey Department of Transportation, Trenton, NJ, 1996.

Willenbrock, J.H., and P.A. Kopac, *The Development of Tables for Estimating Percentage of Material Within Specification Limits*, Pennsylvania Department of Transportation, Bureau of Materials Testing and Research, Project No. 74-27, Research Report No. 2.