NCHRP
REPORT 447

NATIONAL
COOPERATIVE
HIGHWAY
RESEARCH
PROGRAM

Testing and Inspection Levels for Hot-Mix Asphaltic Concrete Overlays

TRANSPORTATION RESEARCH BOARD

NATIONAL RESEARCH COUNCIL

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

NCHRP REPORT 447

Testing and Inspection Levels for Hot-Mix Asphaltic Concrete Overlays

JEFFREY S. RUSSELL
AWAD S. HANNA
ERIK V. NORDHEIM
University of Wisconsin–Madison
Madison, WI

ROBERT L. SCHMITT
University of Wisconsin–Madison
Platteville, WI

SUBJECT AREAS

Planning and Administration • Pavement Design, Management, and Performance • Materials and Construction

Research Sponsored by the American Association of State Highway and Transportation Officials
in Cooperation with the Federal Highway Administration

TRANSPORTATION RESEARCH BOARD — NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. — 2001

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Systematic, well-designed research provides the most effective approach to the solution of many problems facing highway administrators and engineers. Often, highway problems are of local interest and can best be studied by highway departments individually or in cooperation with their state universities and others. However, the accelerating growth of highway transportation develops increasingly complex problems of wide interest to highway authorities. These problems are best studied through a coordinated program of cooperative research.

In recognition of these needs, the highway administrators of the American Association of State Highway and Transportation Officials initiated in 1962 an objective national highway research program employing modern scientific techniques. This program is supported on a continuing basis by funds from participating member states of the Association and it receives the full cooperation and support of the Federal Highway Administration, United States Department of Transportation.

The Transportation Research Board of the National Research Council was requested by the Association to administer the research program because of the Board's recognized objectivity and understanding of modern research practices. The Board is uniquely suited for this purpose as it maintains an extensive committee structure from which authorities on any highway transportation subject may be drawn; it possesses avenues of communications and cooperation with federal, state and local governmental agencies, universities, and industry; its relationship to the National Research Council is an insurance of objectivity; it maintains a full-time research correlation staff of specialists in highway transportation matters to bring the findings of research directly to those who are in a position to use them.

The program is developed on the basis of research needs identified by chief administrators of the highway and transportation departments and by committees of AASHTO. Each year, specific areas of research needs to be included in the program are proposed to the National Research Council and the Board by the American Association of State Highway and Transportation Officials. Research projects to fulfill these needs are defined by the Board, and qualified research agencies are selected from those that have submitted proposals. Administration and surveillance of research contracts are the responsibilities of the National Research Council and the Transportation Research Board.

The needs for highway research are many, and the National Cooperative Highway Research Program can make significant contributions to the solution of highway transportation problems of mutual concern to many responsible groups. The program, however, is intended to complement rather than to substitute for or duplicate other highway research programs.

Note: The Transportation Research Board, the National Research Council, the Federal Highway Administration, the American Association of State Highway and Transportation Officials, and the individual states participating in the National Cooperative Highway Research Program do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

NCHRP REPORT 447

Project D10-39A

ISSN 0077-5614

ISBN 0-309-06663-8

Library of Congress Control Number 00-136358

Price $40.00

NOTICE

The project that is the subject of this report was a part of the National Cooperative Highway Research Program conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council. Such approval reflects the Governing Board's judgment that the program concerned is of national importance and appropriate with respect to both the purposes and resources of the National Research Council.

The members of the technical committee selected to monitor this project and to review this report were chosen for recognized scholarly competence and with due consideration for the balance of disciplines appropriate to the project. The opinions and conclusions expressed or implied are those of the research agency that performed the research, and, while they have been accepted as appropriate by the technical committee, they are not necessarily those of the Transportation Research Board, the National Research Council, the American Association of State Highway and Transportation Officials, or the Federal Highway Administration, U.S. Department of Transportation.

Each report is reviewed and accepted for publication by the technical committee according to procedures established and monitored by the Transportation Research Board Executive Committee and the Governing Board of the National Research Council.

Published reports of the

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

are available from:

Transportation Research Board
National Research Council
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

and can be ordered through the Internet at:

http://www4.nationalacademies.org/trb/homepage.nsf

Printed in the United States of America

FOREWORD

By Staff
Transportation Research Board

This report contains the findings of a study to develop a rational method for determining the minimum level for both agency and contractor testing and inspection activities necessary to satisfactorily construct hot-mix asphaltic concrete (HMA) overlays using the AASHTO *Quality Assurance Guide Specification* and the AASHTO *Implementation Manual for Quality Assurance* and to apply the findings of this research to other construction activities. The minimum level of testing and inspection is defined in this report as the minimum testing and inspection resources that should be allocated for a given project. Satisfactorily constructing an HMA overlay is defined as meeting the specifications which are, in turn, defined by test properties and compliance measures. The contents of this report, therefore, will be of immediate interest to highway professionals responsible for planning, administering, and financing highway improvements; those concerned with pavement design, management, and performance; as well as those involved in materials and construction issues.

The University of Wisconsin—Madison was awarded a contract to conduct NCHRP Project 10-39A, "Testing and Inspection Levels for Hot-Mix Asphaltic Concrete Overlays." The Asphalt Institute assisted the research team as a subcontractor. The Arizona, Florida, Kentucky, Minnesota, Ohio, and Wisconsin DOTs and their contractors provided project-specific testing and inspection data for the 16 HMA overlay projects used to measure statistical properties of construction test data and document current inspection practices. The objectives of this research were limited in scope by those data that an agency can collect from construction tests to determine whether an HMA overlay has been satisfactorily constructed. Actual performance data and life-cycle cost data were not included within the scope of the research.

The report includes recommendations for both testing and inspection levels of HMA overlays. Specific recommendations are given for quality control, quality acceptance, and verification testing levels. Inspection level recommendations are provided for important tasks and for those projects where a limited number of agency inspectors are available. The report also provides implementation guidelines and a case application for the research. The case application uses data from an actual project, giving specific examples for quality control, acceptance, and verification testing levels. Development of a list of important inspections for a project, based on attributes and available inspectors, and description of practical applications to consider during implementation are provided in the body of the report. A rational method is also described for determining testing and inspection levels in other areas of highway construction, based on the results of this research project.

As of publication, recommendations in this report may not be completely compatible with FHWA regulations (23 CFR 637 Subpart B). The CFR applies to Federal-aid highway projects on the National Highway System and requires independent-sampling for verifying the quality of materials. The samples tested for verification must be independent of quality control samples; consequently, splitting a sample would not be permitted. The CFR does allow splitting samples for the evaluation of testing personnel, procedures, and equipment. This report recognizes both independent- and split-sampling, but encourages split-sampling for all scenarios, principally based on existing guides and practice.

CONTENTS

AUTHOR ACKNOWLEDGMENTS

The research reported herein was performed under NCHRP Project 10-39A by the University of Wisconsin–Madison, Department of Civil and Environmental Engineering, and the Asphalt Institute. The University of Wisconsin–Madison was the contractor for the study. The work undertaken at the Asphalt Institute was under a subcontract with the University of Wisconsin–Madison.

Jeffrey S. Russell, Associate Professor of Civil and Environmental Engineering, University of Wisconsin–Madison, was the principal investigator. Co-investigators were Awad S. Hanna, Associate Professor of Civil and Environmental Engineering, Erik V. Nordheim, Professor of Statistics, and Hussain U. Bahia, Assistant Professor of Civil and Environmental Engineering, all from the University of Wisconsin–Madison. Research Assistants were Robert L. Schmitt, former Research Assistant at the University of Wisconsin–Madison, now Assistant Professor of Civil Engineering at the University of Wisconsin–Platteville, and Galadriel S. Jung, Research Assistant, University of Wisconsin–Madison.

The work was performed under the supervision of Professor Russell. The work at the Asphalt Institute was performed under the supervision of R. Michael Anderson with the assistance of Dwight D. Walker and Gary S. Irvine.

The research team would like to acknowledge the efforts of the External Advisory Board during this study (refer to Appendix B of the research team's final report for a list of names). Members of the External Advisory Board offered valuable guidance and recommendations during the course of the study.

The research team would also like to acknowledge the agencies and contractors that participated in the questionnaire survey and personal interviews. A special thanks to the states and contractors that provided project-specific data: Arizona, Florida, Kentucky, Minnesota, Ohio, and Wisconsin. Their time and efforts are greatly appreciated by the research team, for without such time and efforts, this research project would have not been possible.

TESTING AND INSPECTION FOR HOT-MIX ASPHALTIC CONCRETE OVERLAYS

SUMMARY

This report has been developed in order to assist agencies and contractors in developing a rational method to determine the minimum level of testing and inspection activities necessary to satisfactorily construct hot-mix asphaltic concrete (HMA) overlays. In this report, the "rational method" is defined as the steps necessary to evaluate and prepare the testing portion of the specification. "Minimum level of testing and inspection" is defined as the minimum testing and inspection resources that should be allocated for a given project. "Satisfactorily constructing an HMA overlay" is defined as meeting the specifications which are defined by test properties and compliance measures. Existing AASHTO quality assurance (QA) publications, specifically the *Implementation Manual for Quality Assurance* and the *Quality Assurance Guide Specification (1,2)*, provided guidance during this effort.

Current practices in QA contracting for HMA overlays have been synthesized through a review of literature, surveys from 42 states and 61 contractors, and 40 state QA specifications. In reviewing the specifications, the researchers consistently found three fundamental measures for acceptance testing: (1) mix properties, (2) density, and (3) smoothness. These measures are then linked to the pay equations, which are used to determine the amount of contractor payment. Specifications are written that define test properties and compliance measures and corresponding levels that are perceived to yield a quality pavement. Most states have developed pay factors to relate construction quality with performance by financially penalizing contractors when design targets are not met, there is excessive variation, or both.

Data were collected from 16 HMA overlay projects in six states during the 1997 construction season. An experiment was conducted on these projects to measure the mean and variation between agency and contractor laboratories with 24 split-sample tests. Contractor quality control test data and agency acceptance and/or verification test data were collected. Test property data included coldfeed aggregate gradation, hot-mix plant-produced properties (aggregate gradation, asphalt content, mix volumetrics), and density. Inspection practices and the work of individual project inspectors were documented on 13 projects.

The concept of obtaining a representative sample is reviewed with respect to HMA overlay construction. For plant-mix samples, it is recommended that samples be collected

behind the paver. All samples must be collected in strict accordance with random sampling principles. If the sampling frequency is very high and there are insufficient resources to test the samples, then it may be desirable to modify the sampling rate to adhere to random sampling principles. Samples should be collected with the presumption that they will be tested. Agencies and contractors should obtain time estimates for their specific testing procedures when determining a reasonable sampling frequency.

For contractor quality control, a quantity-based testing frequency is recommended for controlling coldfeed or hot-bin aggregate gradation so as to minimize aggregate particle size variation within the final mixture. A statistical comparison should be conducted at the start of production, between the input and final output aggregate gradation, to determine if both sampling locations provide statistically similar test results. A formal analysis of variance was conducted to measure sources of variation found in plant-mixing and density data. The data analysis concluded that between-day variation is significant for both processes and that the contractor should ensure that hot-mix and density samples be collected and tested within each day of production, rather than high-frequency testing on periodic days. As a starting point, it was recommended that a minimum of n equal to 3 complete hot-mix tests be performed per day. The number of daily tests may be increased to n equal to 4 if the same technician can perform the tests without increasing costs. A rational method of determining the number of daily density tests is provided using a standard statistical equation, which describes an increase in confidence in the daily average as the number of tests increases.

For acceptance testing levels, an analysis has been conducted to determine how specification limits could be evaluated using contractor field data. It is recommended that an agency establish a range for specification limits that results in a desired level of in situ performance. As performance data are collected, they should be compared with the level of contractor variability used to construct the project, and specification limits should be adjusted accordingly. It is recommended that an analysis of normality be performed for Quality Level Analysis (QLA) specifications using state-specific data to ensure that the lot distribution satisfies statistical requirements. Combining material from different days of production into a single lot, which has different statistical characteristics, can produce lots that are not normally distributed. Many states specify quantity-based sampling using material from different days, and it is recommended that an analysis of normality be performed so that both the agency and contractor are aware of the potential consequences. The number of samples within a lot can affect risk levels for both the agency and contractor. It is recommended that states specify percent within limits as a compliance measure to evaluate sample sizes for the lot and acceptable levels of risk.

For agency verification of contractor acceptance tests, it is recommended that the agency perform split-sample testing, rather than independent-sample testing. Split-sample verification testing reduces the number of comparison tests and removes unnecessary project effects during the verification, such as materials, production, and sampling variation. Equation 4.6 (discussed in Chapter 4) can be used to establish the testing tolerances between the contractor and the state laboratories. An implementation procedure was developed for the verification process that evaluates the variability between laboratories at project start-up and compares this variability with statewide project variability. This procedure should be performed on all projects at the start of production.

Inspection levels should be designed using an evaluation of those inspections thought to influence pavement performance significantly. The number of agency inspectors for a particular project should be determined from the project average annual daily traffic, tonnage, and whether agency or contractor data are used for acceptance. Contractor recommendations are based on the availability of two technicians—one at the plant, the other at laydown operations—with limited time allowed for inspection in addition to routine testing functions.

Guidelines are provided that incorporate key decisions made during the planning, design, construction, and operation and maintenance phases of a project that affect the selection of testing and inspection levels. Practical issues to consider when implementing the research results are also recommended. Finally, recommendations for future research are given.

CHAPTER **1**

INTRODUCTION

1.1 BACKGROUND

Along with various other responsibilities, state highway agencies (agencies) must provide the public with the best highway facilities possible, given available resources. Traditionally, agencies have engaged in extensive testing and inspection efforts to ensure construction quality and, thereby, the satisfactory performance of the highway facility. Although serving the public well, such efforts consume an appreciable amount of resources. Agencies throughout the country are reexamining current levels of testing and inspection and the manner in which these quality assurance (QA) efforts are accomplished.

New AASHTO Joint Construction/Materials Quality Assurance Task Force publications (the AASHTO *Implementation Manual for Quality Assurance* and the AASHTO *Quality Assurance Guide Specification*) have been developed to improve product quality and make effective use of resources *(1,2)*. The researchers on this project used these documents to develop a rational method for determining staffing levels for inspecting and testing a specific construction activity and to provide agencies with an ability to apply these findings to other construction activities.

There are standard and emerging contracting methods now available for highway construction, including Method, QA, Design/Build, Design/Build/Operate, Warranty, and Multi-Parameter. Each of these methods influences testing and inspection levels and the degree of agency involvement in the design and construction phases. This study was concerned with QA contracting and the testing and inspection levels for agencies and contractors necessary to satisfactorily construct hot-mix asphaltic concrete (HMA) overlays.

1.2 RESEARCH PROBLEM STATEMENT

Achieving a high-quality HMA overlay is a primary objective of state highway agencies and contractors. An appropriate level of testing and inspection is required to ensure that a high-quality pavement has been constructed. High levels of testing and inspection provide relatively large amounts of data to assess pavement quality; however, this effort may be time consuming and may not be cost-effective in the long term. Low levels of inspection and testing appear to be more economical, but insufficient data may make it difficult to assess the quality of the constructed HMA overlay.

1.3 RESEARCH OBJECTIVE

The objectives of this research were (1) to develop a rational method for determining the minimum level for both agency and contractor testing and inspection activities necessary to satisfactorily construct HMA overlays using the AASHTO *Implementation Manual for Quality Assurance* and the AASHTO *Quality Assurance Guide Specification* and (2) to apply the findings of this research to other construction activities.

The rational method was defined as the steps that should be taken when developing the testing and inspection specification. The minimum level of testing and inspection was defined as the minimum testing and inspection resources that should be allocated for a given project. Satisfactorily constructing an HMA overlay was defined as meeting the specifications with specified test properties, test methods, compliance measures, and pay equations. When a contractor received 100-percent payment, it was assumed that the contractor constructed the overlay as specified.

The objectives of this research are limited in scope by those data an agency and contractor can collect from construction tests to determine whether an HMA overlay has been satisfactorily constructed. Actual performance data and life-cycle cost data are not included in the scope of this research.

Figure 1.1 presents a summary of the current recommended testing levels found in the AASHTO publications. An objective of this research is to provide a rational method for determining the testing levels shown in Figure 1.1. These recommended levels are supported by data; however, in the absence of data, a methodology based on engineering judgment and consultation with industry professionals has been applied in the form of recommended guidelines. The methodology developed in this report can be evaluated by the agency and contractor and adjusted to meet the project needs prior to implementation.

1.4 RESEARCH APPROACH

This research was conducted in two phases. In the first phase, the research team developed a work plan for determining testing and inspection levels. Steps necessary to develop the work plan included (1) review of relevant literature and research in Quality Assurance (QA) for highway construction and other

		RECOMMENDED	
	Testing Property	Quality Control	Acceptance
Materials	Aggregate Gradation	Coldfeed or Hot Bins Plant Set-up: 1 per 500 tons or C	N/A
	Aggregate Properties (Fracture)	1 per 1,000 tons	N/A
	Asphalt Binder Properties	N/A	N/A
Plant Mixing	Aggregate Gradation	1 per 1,000 tons	A
	Asphalt Content	1 per 500 tons or C	A
	Mixture Properties (Air Voids, VMA, etc.)	1 per 1,000 tons	A
Mix Transport and Laydown	Temperature of Mix	1 per hour	N/A
	Temperature of Base of Air	As needed	N/A
	Tack/Prime	Load or Half Day, whichever comes first	N/A
	Pavement Application Rate	C	N/A
Compaction	Density	1 per 500 tons	A
	Temperature of Mat	1 per hour	N/A
	Thickness	C	N/A
	Smoothness	C	Sublot: 0.1 Lane-Mile Lot: Project

C = As needed to control operations
A = To be determined by the agency (sublot and lot sizes)
N/A = No testing frequency provided

Figure 1.1. Testing frequencies recommended by AASHTO.

industries, (2) review of current practices of QA programs, (3) interviews of those individuals having experience in QA contracting, (4) identification of tests and inspections thought to ensure pavement performance, and (5) development of a procedure for determining minimum testing and inspection levels.

The second phase of research was implementing the work plan through actual field data collection and determining minimum testing and inspection levels for HMA overlays. Implementing the work plan involved project selection, data collection and coordination with project field staff, synthesis of project data, and analysis of the data. Recommendations were then formulated from the data analysis to develop a process for determining testing and inspection levels for HMA overlays and to describe how this process can be used for determining levels in other types of highway construction.

1.4.1 Literature Review

A literature review was conducted to identify, collect, review, and synthesize literature and research on HMA quality control and acceptance, including the unpublished NCHRP Interim Report on Construction Testing and Inspection Levels (performed under NCHRP Project 10-39), and the AASHTO Joint Construction/Materials Quality Assurance Task Force publications (*Implementation Manual for Quality Assurance* and *Quality Assurance Guide Specification*) *(1,2).* This review focused on these documents as well as current practices used in measuring and assessing HMA construction quality. A glossary of terms relevant to this research project is provided as Appendix A of this report. Established quality control (QC) and QA literature external to highway construction was also reviewed to gain a greater perspective on methods for quality control and quality assurance.

A survey was conducted to understand issues relevant to the frequency of testing and inspection by agencies and contractors on HMA overlay projects. Survey forms were developed with the primary objective of understanding both quantitative and qualitative aspects of current QA programs, including those factors affecting testing and inspection levels for both agency and contractor perspectives.

QA specifications were collected from 40 states in conjunction with the surveys. Many of the specifications were dated

1996 and were either initial drafts or revisions. A follow-up survey was sent in 1997 to update and clarify information received from the initial survey. These specifications were thoroughly reviewed to gain perspective on current practices using QA contracting to construct HMA overlays.

1.4.2 Procedure Development

A procedure was developed to provide a method for determining appropriate levels of testing and inspection for both agencies and contractors. The review of the literature, surveys, and current QA specifications was supplemented with 19 interviews with individual agency and contractor representatives during this process. An optimization equation to determine this level was not used, but rather the elements of the specification that determine whether satisfactory construction has been achieved were analyzed and integrated into a process. Because the determination of testing levels is a highly quantitative issue, several statistical methods were used within this process.

An External Advisory Board (participants listed in Appendix B of the research team's final report) was created to provide technical assistance during the development of this procedure. Current practices identified in the surveys and specifications were reviewed with the Board to develop a comprehensive understanding of current testing and inspection levels and to select a range of HMA tests thought, based on the collective experience of the Board, to influence pavement quality significantly.

1.4.3 Work Plan Design

A work plan was designed to collect field construction data from an adequate number of actual HMA overlay projects. The work plan specified that the construction testing and inspection data be collected from six states, each with at least two projects. Project selection was limited to HMA overlay projects having sufficient project tonnage to provide at least 24 hot-mix companion split-sample tests between the contractor and agency. Many QA specifications compare agency and contractor test results—this component of the work plan allowed an evaluation of the relationship among test results from multiple laboratories. The work plan was not designed to collect long-term performance data and correlate these data with QA construction test data. Rather, the focus was on collecting and analyzing construction data used to determine whether an HMA overlay has been satisfactorily constructed.

1.4.4 Interim Report

The purpose of the interim report was to summarize the findings from the literature review, procedure development, and work plan design. The NCHRP panel provided a review of the report and offered recommendations for modifying the work plan to better achieve the research objectives. Revisions were made to the work plan prior to starting the second phase of the project.

1.4.5 Work Plan Implementation

The work plan was implemented to collect construction-specific test and inspection data from a wide range of HMA overlay projects. Sixteen projects were identified for the study, with some states having more than two projects. Six states were identified for the study: Arizona, Florida, Kentucky, Minnesota, Ohio, and Wisconsin. At least 24 split-sample test results were collected between the agency and contractor on these projects. A three-way split-sample comparison was made on six projects between the agency, contractor, and the Asphalt Institute. Inspection practices were also observed and recorded for 13 projects.

1.4.6 Documentation and Recommendations

The testing and inspection data from 16 projects were documented and analyzed. Several statistical methods were used in this effort in order to provide insight into the behavior of the testing data during construction. Inspection data were analyzed in order to understand the relationship between input of resources and quality output. Recommendations were developed using the data to select appropriate testing and inspection levels for constructing HMA overlays. A methodology to determine appropriate testing and inspection levels for other construction activities was developed.

1.4.7 Implementation

Recommendations were used to develop an implementation plan that follows a format to implement the findings of this research into current practice. Guidelines were developed and a case application was provided. Practical applications that the agency and contractor must be aware of to overcome barriers to implementing this research were given.

1.5 SUMMARY

The research team began the effort by assessing the background of the research problem and stating the research objectives. The AASHTO QA publications were cited as important documents that agencies and contractors can refer to for methods for determining testing and inspection activities necessary to satisfactorily construct HMA overlays. Steps taken in the research included literature review, procedure development, design and implementation of a work plan, documentation, recommendations, and implementation. The next objective of the research effort was to understand current practices in quality control and quality assurance for constructing HMA overlays.

CHAPTER 2

CURRENT PRACTICE

2.1 INTRODUCTION

This chapter presents findings from current practices in testing and inspection for HMA construction, as well as established QC methods. The literature was reviewed specific to testing, inspection, and methods of implementing QA into a field-level format. Survey techniques were used to gather pertinent information from agencies and contractors concerning current practices in QA programs. Specifications from 40 state highway agencies were reviewed in order to understand the working elements of the QA program. Personal interviews were conducted with 19 practitioners from eight states to gain a greater understanding of the survey responses and specifications, thereby allowing the research team to develop a more detailed understanding of current QA practices for HMA construction.

2.2 LITERATURE REVIEW

A literature and research review was conducted to find all relevant information on testing and inspection levels for HMA overlays. A separate review of testing literature, inspection literature, and literature on field-level implementation was performed. Testing provides a quantitative description of the work product, while inspection provides a qualitative assessment of various processes, such as the condition of production equipment and the methods used to produce a high-quality construction product.

2.2.1 Testing Literature

Numerous QA and acceptance reports and papers using statistical methods were collected and reviewed. Several of the documents discussed operating characteristic curves for assigning both seller (contractor) and buyer (owner) risks during the acceptance or rejection of work. Literature from HMA course manuals and industry trade publications were also consulted. Many highway agencies refer to investigations and informal studies in an attempt to define statistical quality control and quality assurance, but most of this work is neither documented nor published. Although these documents provided detailed descriptions of QA and acceptance procedures, none specifically addressed testing levels.

The AASHTO *Implementation Manual for Quality Assurance* discusses the unit of measure for testing *(1)*. Frequency schedules for testing may be derived in terms of time or quantity or both. A schedule based on quantity may yield more samples than desired when production rates are high and fewer samples when production rates are low or intermittent. Schedules based on time yield fewer samples for high production and more samples for low production.

Determining levels of variation of the materials and testing, with respect to pay adjustments, is an important element in the specification. With respect to variation, the specifying agency should understand contractor process capability, state-of-the-art recommendations, and actual product performance history and then set acceptable ranges based on these factors *(1)*. Pay factors are designed to encourage quality and be related to the loss or gain in service life of the product. Incentive payments should be outlined to contractors to achieve proper control of the construction process and greater longevity of the pavement.

AASHTO Designation R9-90 has defined the variability-known and variability-unknown methods for accepting work quality *(3)*. Both methods use fundamental sampling plans to estimate the mean and standard deviation of the work. The variability-known procedure assumes that the variability is known and constant and evaluates the lot mean on the basis of acceptance criteria developed using an assumed variability for the lot. The variability-unknown acceptance plan assumes the variability of the lot to be unknown *(3)*.

AASHTO has recommended the use of Quality Level Analysis (QLA) specifications in both of the QA publications *(1,2)*. QLA-type specifications are thought to be the most statistically rigorous approach to measure compliance. They are thought to be more beneficial to the agency and contractor, because they provide the ability to quantify risk levels during acceptance. A potential disadvantage of QLA is ensuring that the sample in question meets the assumptions of a normal distribution.

A common measure for QLA-type specifications is Percent Within Limits (PWL), which estimates the percentage of the lot falling within lower and/or upper specification limits. Percent Defective (PD) can also be used in QLA specifications, where the percentage of the lot falling outside the specification limits is estimated (e.g., PWL equaling 100 − PD).

The level at which an agency is willing to accept the HMA product at full payment is referred to as the Acceptable Quality Level (AQL). The level at which an agency is willing to reject the HMA product is the Rejectable Quality Level (RQL). Values between the AQL and RQL are accepted, but at reduced payment. State-of-the-practice suggests that AQL values of PWL equalling 90 and RQL values of PWL equalling 60 are commonly specified by agencies, and the RQL value can vary from a low value of PWL equalling 25 to a high value of PWL equalling 80. For PWL values between the AQL and RQL, the work is thought to have lower quality and a pay adjustment is typically assigned.

Determining the expected life and life-cycle costs of an HMA overlay are very difficult. Prediction models have been developed for determining the expected life, or pavement performance, from numerous input variables. *NCHRP Report 332,* "Framework for Development of Performance-Related Specifications for Hot-Mix Asphaltic Concrete," used prediction models to determine pavement performance *(4)*. Pavement performance predictions were found with fundamental mixture response variables as a function of simulated materials and construction (M&C) variables. The Simplified Rational Pavement Design (SRPD) model developed from AASHTO road test data was used in the analysis *(5)*. The SRPD model was supplemented with an Asphalt Institute equation developed to determine the dynamic modulus from basic mix properties *(6)*. The result of one performance prediction concluded that conformance levels of PWL equal to 72 and 100 percent PWL for asphalt content did not provide a good indicator of how the constructed pavement would perform. Although the scope of the study was not to produce models for predicting pavement performance, it offered conceptual framework for their development. The study revealed that one of the most difficult components of performance-related specifications is the economic quantification of predicted pavement performance *(4)*.

The FHWA report, "Cost Effectiveness of Sampling and Testing Programs," provided a means of establishing priorities among QC tests and optimizing sampling frequencies for each test, based on the effects of material properties measured on the long-term performance of the pavements *(7)*. Appropriate procedures were developed, including critical considerations and limitations resulting from lack of suitable stochastic models to predict performance and contractor behavioral response to changes in testing frequency. Results based on limited models indicated that higher frequencies of testing than commonly used would be cost-effective, decreasing the equivalent annual pavement costs by much more than additional testing costs.

A study in Florida evaluated the cost-effectiveness of testing procedures for HMA construction *(8)*. An analysis was performed to establish the probability of density test failure, and the corresponding margin of failure, for different levels of testing and lengths of projects. A reduction in apparent structural strength because of density test failure was computed on the basis of relationships established between elastic moduli and density. This structural deficiency was corrected by an additional thickness of material sufficient to reduce the pavement surface deflection to the same level as that encountered in a properly constructed pavement. An elastic layer computer program was used to determine these additional thicknesses. The cost-effectiveness of any particular testing frequency was based on the cost of testing plus the cost of additional material to correct for deficient density. Results indicated that current density testing frequencies of 12 tests per mile are generally cost-effective for projects 3 to 10 mi long. Projects greater than 10 mi should reduce the frequency of density testing.

The purpose of NCHRP Project 9-7, "Field Procedures and Equipment to Implement SHRP Asphalt Specifications," was to establish appropriate levels of quality control for asphalt mixtures constructed using Superpave™ equipment *(9)*. The study included field sampling and testing on six projects in five states. A recommendation of the study was to perform additional sampling and testing 2 days prior to production, using a procedure known as pre-control. Initial production pre-control is intended to provide sufficient information to allow the plant and project personnel to evaluate the production of the design mixture. Within this initial phase of control, the QC personnel attempt to control the volumetric properties of the as-designed asphalt mixture. If the properties do not match the original design but are acceptable, then changes may be made to the target values for future control. During normal production, recommended QC testing frequencies are 5 sublots per 5,000 ft for density. No specific sampling/testing rate is recommended for the plant-produced mix; however, the contractor should "periodically" determine laboratory air voids for control purposes. Recommended acceptance testing frequencies for both plant-produced mix properties and density were specified as 5 samples per 10,000 tons, using stratified random sampling procedures.

The FHWA's "Practical Applications of Statistical Quality Control in Highway Construction" was one of the first comprehensive publications that specifically addressed the use of statistics to control highway construction quality *(10)*. The publication, primarily a course notebook for engineers and technicians, presents principles and examples of applying statistical QC methods and describes the use of statistical process control in highway construction. It begins by reviewing the collection, organization, and analysis phases of QC data. Well-established sampling, analysis, and control methods provide the reader with easily understood examples and procedures. Methods for sampling are presented using a workshop format that identifies features of certain sampling methods. For example, stratified random sampling is defined in terms of sublot sampling, because sublots are typically used for collecting construction data. Features of the normal probability distribution are discussed, with suggestions for several possible inferences, if, in fact, this type of distribution is present in the data. Procedures for developing statistical process control charts are given, assuming the data are normally distributed.

2.2.2 Inspection Literature

Inspection as it relates to the contractor QC plan, is discussed in the AASHTO *Implementation Manual for Quality Assurance*. This manual indicates that inspection is an activity as important to QC as it is for acceptance, for both production facilities and field observations *(1)*. Inspection minimizes problems in the production facility that are visually detectable, such as stockpile or equipment maintenance practices or needs, that may eventually affect the quality of the material produced. The QC plan usually indicates inspection activities that will be performed by the contractor's QC personnel. Agency personnel should be made aware of the importance of this part of the program. According to world quality leaders, such as J.M. Juran, inspection can be defined as the determination of whether a product conforms to a specification *(11)*.

From the review of available information, this discussion of inspection is provided for the QC plan, and not for quality assurance. The scope of inspection becomes a function of the QC plan by definition (quality assurance monitors the contractor's QC plan). Inspection is also a function of contractor performance. For example, a highly variable process is typically subject to a greater degree of inspection. Field personnel typically "tighten inspection" when the construction process does not appear to be under control. Tightened inspection usually results in more frequent testing.

Several sources were reviewed to create a comprehensive listing of current inspection practices. The literature review included manuals from AASHTO, the Federal Aviation Administration (FAA), FHWA, the National Asphalt Pavement Association (NAPA), and the U.S. Army Corps of Engineers. These sources were selected because they were most applicable to the HMA industry and provided comprehensive descriptions of the HMA construction process. The literature review provided practical applications for quality control, as well.

The first document reviewed was the *Hot-Mix Asphalt Paving Handbook* developed jointly by AASHTO, FAA, FHWA, NAPA, and the U.S. Army Corps of Engineers, to describe the production and placement of asphalt mixtures from a practical viewpoint *(12)*. The handbook is intended for both agency and contractor personnel and provides an overview of inspection levels of HMA production from processes to inspection tasks.

The *Hot-Mix Asphalt Paving Handbook* was used as a guide in developing the *Hot-Mix Asphalt Construction Participant Manual*. This manual, however, is only specific to the laydown process of HMA construction and gives specific instructions and guidelines on important considerations for laydown crews. This manual is the HMA construction portion of a four-part National Asphalt Training Program developed by AASHTO and FHWA and includes (1) Design, Rehabilitation, and Maintenance of HMA Pavements, (2) HMA Materials and Mix Design, (3) HMA Production, and (4) HMA Construction *(13)*. This manual is intended to provide field personnel with recommended practices to minimize difficulties during construction laydown.

NAPA literature was also reviewed, including four specific handbooks: (1) *Quality Control for Hot Mix Asphalt Operations*, (2) *Roller Operations for Quality*, (3) *Paver Operations for Quality*, and (4) *Field Management of Hot Mix Asphalt*. The purpose of *Quality Control for Hot Mix Asphalt Operations* is to assist HMA producers in establishing a QC system that will ensure a high probability of compliance with the specifications *(14)*. This handbook gives information specific to implementing a quality program in HMA production. *Paver Operations for Quality (15)* and *Roller Operations for Quality (16)* include information on key factors during paver and roller operations to provide high-quality pavements that meet specified density.

Factors affecting the volumetric properties of HMA are discussed in *Field Management of Hot Mix Asphalt (17)*. This handbook focuses on the material properties of the mix during plant production and their effects on the final pavement properties. A unique component of this handbook is a rating system, developed by a panel of NAPA Quality Improvement Committee members and FHWA, to rate the importance of specific factors affecting volumetric properties of the mix. The rating system is shown in Table 2.1.

According to NAPA, the properties and characteristics of the aggregate are the most important factors that affect the HMA mix *(17)*. One of the aggregate properties rated by NAPA, aggregate shape, was given a rating of 10, because it affects the air voids and voids in mineral aggregate (VMA) in both the laboratory- and plant-produced mixes *(17)*.

2.2.3 Field Implementation Literature

Field-level implementation requires the use of several practical applications to ensure a QA program is implemented successfully. QC plans, checklists, daily diaries, and feedback systems were researched through a review of nationally accepted QC literature and HMA trade publications.

TABLE 2.1 NAPA rating system *(17)*

Rating	Significance
0	No Effect
1	Minimal Effect
2	Little Effect
3	Minor Effect
4	Some Effect
5	Moderate Effect
6	Appreciable Effect
7	Significant Effect
8	Principal Effect
9	Major Effect
10	Dominant Effect

2.2.3.1 QC Plans

QC planning is an integral part of enhancing awareness for all levels of personnel involved in a project. QC plans are considered an important part of quality programs by both nationally accepted QC practices and the HMA industry. Specifically for HMA construction, planning appropriate control is one of the most effective practical applications that a contractor can use to place a quality pavement. A QC plan should state the quality policies, practices, organization, and activities that will be conducted to produce a quality product for a project. The plan should meet product specifications through process management and inspection *(18).*

A formal planning process using a QC plan allows the contractor to specify key areas during production that will require personnel awareness. With increased awareness, it is more likely that these areas will receive the required attention by project staff. QC plans will change depending on agency philosophy and project-specific requirements, but the format should not change considerably for different projects. A QC plan should be created at the beginning of a project and all project personnel should be aware of its contents. Using customer satisfaction as a goal, QC plans are a way for the seller to demonstrate a commitment to quality to the customer *(14)*.

2.2.3.2 Checklists

Checklists provide a constant reminder of inspection tasks to be completed. A checklist is defined as a tool to ensure that all important steps or actions in an operation have been taken *(19)*. These steps or actions are listed so that inspectors know which inspection tasks must be performed at a specific location. This approach is useful for HMA inspection because, in most cases, inspection processes do not require the attention of an inspector for an entire day. It is possible to perform inspections during one process for a short period and then move on to another process. A "patrol beat" can be used over multiple production areas to better allocate the time of one inspector *(11)*. A patrol beat consists of several inspection tasks during different processes that do not require an inspector to monitor production at all times. Because the inspection tasks usually require a brief visual inspection, one inspector can cover several areas. This type of inspection is needed for projects where only one agency inspector is available and time must be split between the plant and laydown. Checklists are especially helpful for this type of inspection to ensure that no required inspection tasks are overlooked.

For a checklist to be useful, it must be in the proper format. According to QC literature and HMA trade publications, some of the suggested checklist characteristics are as follows *(12,14)*:

- Standardized form;
- Clear and simple form to ease recording;
- Appropriate spacing for recording;
- Clear directions for correct use of the form; and
- Ample space for recording the project number and location, weather conditions, signature, date, and remarks.

2.2.3.3 Daily Diaries

Daily diaries are necessary for field-level implementation of a quality program. They provide a way to record information regarding daily project conditions that enable project staff to later understand the project conditions of a specific day. Daily diaries require a systematic approach to ensure that documented information can be used effectively at a later time. Diaries should also include items such as changes that occur during operation, different or unusual events on a project, visitors to a project, and reasons for paving delays (e.g., breakdown or weather). Diaries should be updated twice per day, usually at the middle and end of the work day. The recorded data should be detailed and include the date and location of paving, names and titles of people involved in any discussion, the topics discussed, and outcomes of the discussions *(12)*. The diaries should be employee-specific, and the level of detail noted should depend on an employee's duties and location.

Checklists and daily diaries provide a standardized way to document data regarding the HMA construction process. They are important for reference in cases of contractor and agency disputes, as well as follow-up research to understand pavement performance. The data recorded can then be used by project staff for the feedback that is necessary for process control and improvements during construction.

2.2.3.4 Feedback System

A feedback system is important in implementing a QC and QA program. Decision-makers need to know what is occurring, to have "feedback," in order to decide if corrective action should be taken *(11)*. Feedback requires project personnel to communicate project concerns with one another. If changes to an earlier process in production are required, based on the output of a later process, field personnel aware of the required changes must provide the feedback needed for correction. For example, if segregation is a problem during laydown, there is probably a problem in the plant-mixing or mix-transport processes. The paving inspectors would communicate this information, which could result in an evaluation of the truck loading method or plant-mixing operations to determine the source of segregation.

Feedback is also input from a customer relative to the effect of the product. A feedback loop is a systematic sequence for communicating information on process performance as an input to maintenance of process stability *(20).* For HMA construction, the customer could be any member of the project staff who receives output from another process. For example, a customer could be a paving crewmember who receives mix from the truck as the mix is delivered. Any input that the paving crewmember would communicate regarding the quality of the mix received would be feedback and part of the feedback loop.

For feedback to have an effect on HMA production, a system must be established. Figure 2.1 shows a feedback loop that relates information about outputs back to the input stage for process analysis *(21)*. This figure could be applied to HMA construction to identify the source of segregation during production. Input would be the HMA mix as it is delivered to the paver; however, the process would be laydown and the output would be the mix directly behind the paver. If segregation is detected, the feedback loop should be completed by the paving inspector who would inform the person who can adjust the input (delivered mix) and eliminate the segregation before it is delivered to the paver. It is also possible that the segregation is caused by paver stoppage, which would require informing the paver operator of the problem.

Without feedback from personnel at laydown, mix plant, or trucking operations, there is no way of correcting the segregated mix. If feedback is used effectively, all parties involved in the quality system should have up-to-date information on defects and corrective actions. This will lead to improvements in the process, resulting in a better quality product. Improved feedback can lead to improved HMA construction, resulting in a higher quality pavement.

2.3 SURVEY FINDINGS

A survey was conducted to collect information about current QC and QA practices among agencies and contractors. The survey requested both quantitative and qualitative aspects of current QA programs. A mailing list was established for 50 state highway agency representatives by using the membership list of the AASHTO Highway Subcommittee on Materials, and a list of 269 contractors was provided from 14 state HMA contractor associations. In some cases, the same contractor was listed, but a separate company division or asphalt plant was provided. Sample surveys are provided in Appendix C of the research team's final report.

A total of 42 agency and 61 contractor surveys were returned. Table 2.2 summarizes three basic attributes from the agency surveys: (1) contractor requirements, (2) project resources, and (3) acceptance testing. Contractor data from a similar survey are summarized in parallel to the agency data in Table 2.2.

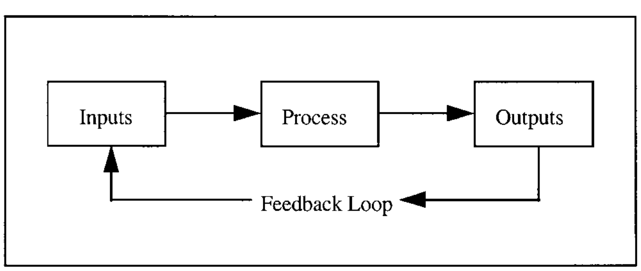


Figure 2.1. Feedback loop (21).

2.3.1 Contractor Requirements

A greater percentage of states are requiring certification of contractor technicians (36 of 42 states). Of these 36 states, 27 use agency programs, 3 use the National Institute for Certification in Engineering Technologies (NICET), 3 use the New England Transportation Technological Certification Program (NETTCP), 2 have contractor association programs, and 1 uses a university program. Although certification may seem an obvious component of a QA program, the importance of formal training and education for those responsible for field-level decision making must not be overlooked.

Most states are requiring the contractor to perform mix designs (i.e., 36 of 42) and provide QC plans (i.e., 34 of 40). The QC plan serves as a critical component of the QA program by providing the agency and contractor with a document outlining those tests or production processes that will be conducted and monitored during construction. A typical QC plan contains the types and frequencies of tests and inspections, methods for material storage and handling, a list of personnel responsible for various QC functions, and methods to ensure that testing equipment complies with testing standards.

2.3.2 Project Resources

With the current trend of agency downsizing, time and cost initially were thought to be a major factor in determining testing levels. However, only 5 of 35 states identified time and cost as factors. Agency-estimated contractor costs for quality control ranged from \$0.30 to \$2.00 per ton of mix, depending on construction markets and project size measured in tons. Contractor cost as a percentage of the total project cost ranged from 0.5 to 10 percent, with an average of 2 percent provided in the surveys.

Agency staffing levels in 11 of 42 states have been used to set testing levels. Current staffing levels for a typical QA project were provided by 31 agencies and 58 contractors. Agencies are generally staffing two field personnel—with one at the mix plant and the other at laydown operations. These agency personnel work from 6 to 8 hr per day, while contractor personnel work about 10 hr per day. Contractors are staffing two to four field personnel on the average QA project, with a minimum of one person assigned to the mix plant (QC laboratory) and one person to laydown operations to measure pavement density. An additional person may be included to assist with laboratory testing and monitoring of aggregate stockpiles. Most contractors have a QC manager who supports the plant and laydown personnel.

TABLE 2.2 Agency surveys

Attribute	Yes	No	Total Responding[a]
Contractor Requirements			
Technician certification required	36	6	42
Contractor provides mix design	36	6	42
Contractor QC Plan required	34	6	40
Project Resources			
Required time and cost determine testing levels	5	30	35
Staffing determines testing levels	11	31	42
Waive testing for small tonnage (< 500 tons)	22	8	30
Adjust testing levels during production	10	31	41
Acceptance Testing			
Contractor tests used for acceptance	27	15	42
Split-samples used for verification	29	13	42
Independent-samples used for verification	20	22	42
Pay adjustments	39	2	41
Dispute Resolution System	33	9	42

[a]When the total number responding is less than 42, this means that some agencies did not provide a response.

Four project factors were evaluated as factors to waive or increase current testing and inspection levels: (1) traffic volume, (2) project tonnage, (3) overlay thickness, and (4) existing base material. Most of the states (i.e., 22 of 30) waive testing and inspection requirements for small tonnage projects (i.e., less than 500 tons), while three responses were given for the other factors. Contractors responded that 500 to 1,000 Annual Average Daily Traffic (AADT), tonnage less than 500 tons, overlay thickness less than 25 mm, and concrete crack-and-seat projects should waive testing. Contractors also thought that the four project factors should increase testing and inspection, but the quantitative measures of these attributes varied considerably. Typically, more testing is conducted during the beginning of HMA production, and either agencies specify the increase (as was true for 10 of 31) or contractors initiate an increase in testing.

2.3.3 Acceptance Testing

As agencies have continued downsizing, there has been an increased use of contractor test results for acceptance. The surveys found that 27 of 42 states (64 percent) are specifying contractor acceptance tests. A smaller number of states have preferred to keep the roles of contractor QC testing and agency acceptance testing independent of each other. In either case, the agency is responsible for overall quality assurance and acceptance of the work.

Two methods are available to verify test results between the contractor and agency: (1) split-sampling and (2) independent-sampling. The surveys found that some states are specifying both methods for verifying test results, as shown by the high frequency of responses. Many of the states are using split samples for verification (i.e., 29 of 42), and a near-equal number are using independent samples (i.e., 20 of 42). The benefit of split-sampling is to avoid unnecessary project effects when comparing tests, such as the effects of sampling from different locations within a truck box or mat and sampling at different times during production. Independent-sampling allows the contractor and agency to conduct the sampling process independent of each other.

Another important finding was the widespread use of pay adjustments, where 39 of 41 states (95 percent) are implementing some form of contractor pay adjustment. Most of the states have implemented dispute resolution systems for the acceptance and pay adjustment process (i.e., 33 of 42). These systems provide a methodical, expedient approach to resolve discrepancies with testing data.

2.4 QA SPECIFICATIONS

QA specifications were collected from 40 states in conjunction with the surveys. Many of the specifications were dated 1996 and were either initial drafts or revisions. A follow-up survey was sent in 1997 to update and clarify information received from the initial survey. The information from these specifications was divided into two categories: (1) acceptance testing and (2) pay adjustments.

2.4.1 Acceptance Testing

After reviewing the specifications, the research team found that the specifications consistently included three fundamental measures for acceptance testing: (1) mix properties, (2) density, and (3) smoothness. These measures describe overall pavement quality by measuring, respectively, the HMA material composition, the densification of the material to withstand repeated loading, and ride quality. Whether viewed independently or collectively, these measures describe the quality level achieved during the construction. A synthesis of raw data collected from the state specifications is provided in Appendix D of the research team's final report.

Five different measures are being used to determine specification compliance: (1) Average, (2) Quality Level Analysis, (3) Average Absolute Deviation, (4) Moving Average, and (5) Range. Table 2.3 provides the characteristics of these compliance measures along with supporting equations. Each of these measures uses different statistical characteristics to estimate the material properties.

Sublots and lots are used to define the amount of material for each acceptance decision. The lot is composed of several equal-size sublots, to allow efficient sampling under often-changing construction conditions. Figure 2.2 provides an example of four 1,000-ton sublots used to create one 4,000-ton lot. Within each sublot, two samples are randomly chosen.

2.4.1.1 Mix Properties

Table 2.4 provides the acceptance attributes for mix properties for 40 states. There are three primary tests for mix properties: (1) aggregate gradation, (2) asphalt content, and (3) mix volumetrics (e.g., air voids and VMA). Most states are using tonnage to define sublot and lot sizes. Sublot sizes range from 1 test per 500 tons to 1 test per 2,000 tons. The next most common measure for sublots and lots is time, where sublots are in increments of 3 hr and lot sizes range from 1 to 4 samples per day. A variable lot size is specified

TABLE 2.3 Characteristics of compliance measures

Compliance Measure	Characteristics	Equation
Average	• Arithmetic average of tests • Variation must be known because it determines how accurately the average can be estimated from a given sample size. • A confidence interval should be constructed to describe the interval of the mean that can be found at a specified probability level.	C. I. = $z_{\alpha/2}\sqrt{\dfrac{\sigma^2}{n}}$ where, C.I. = Confidence Interval of mean; $z_{\alpha/2}$ = standardized statistic; σ^2 = known variance; and n = number of tests.
Quality Level Analysis	• Estimate PWL or PD using the sample mean and standard deviation. • Using the interrelationship of the mean and standard deviation to estimate PWL or PD develops a distribution of the process. • Quality indexes for the upper and lower specification limits are first calculated (Q_U and Q_L) then applied to statistical tables to determine the estimated PWL or PD.	$Q_U = \dfrac{(USL - \overline{X})}{s}$ $Q_L = \dfrac{(\overline{X} - LSL)}{s}$ where, USL = Upper Specification Limit; LSL = Lower Specification Limit; $\overline{X}$ = sample mean; and s = sample standard deviation.
Absolute Average Deviation	• Average of absolute deviations from a target value, typically the JMF design value. • Specifications are currently structured to allow greater cumulative deviations from the target for smaller sample sizes.	$\Delta = \dfrac{\Sigma \lvert X - TV \rvert}{n}$ where, Δ = average absolute deviation; X = individual test result; TV = Target value; and n = number of tests.
Moving Average	• Measures the arithmetic moving average of several consecutive tests. • Evaluates changes or trends in the moving average relative to target values or specification limits.	$\overline{X} = \dfrac{\Sigma X}{n}$ where, $\overline{X}$ = sample mean; X = individual test result; and n = number of tests.
Range	• Measures the arithmetic range of tests. • Compares the range of values to specification limits, but does not compute the distribution of this range.	Range = Max – Min where, Max = Maximum test value; and Min = Minimum test value.

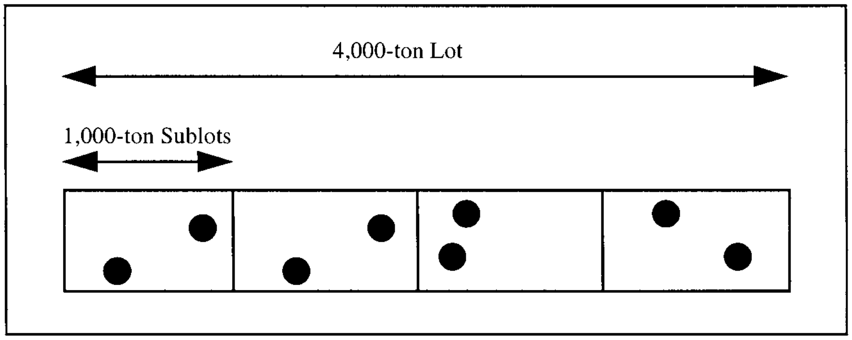


Figure 2.2. Configuration of sublot and lot for acceptance.

TABLE 2.4 Mix property acceptance attributes for 40 states

Attribute	Number of States Specifying		
	Aggregate Gradation	Asphalt Content	Mix Volumetrics
Sublot Size			
1 per 500 tons to 1 per 900 tons	12	17	9
1 per 1,000 tons to 1 per 2,000 tons	12	10	15
1 per 3 hours	2	6	5
4 samples	1	1	0
Variable	2	2	2
Lot Size			
1 per 500 to 5 per 6,000 tons	13	17	15
1 to 4 per day	7	11	7
4 Sublots	2	2	1
Project	1	1	0
Total per Mix Design	4	4	3
Variable	2	2	2
Cumulative	0	1	1
Continuous	0	1	0
Sampling Location[a]			
Coldfeeds or Hot Bins	17	0	0
Plant Discharge	4	7	4
Truck	15	19	15
Windrow	1	2	1
Volume Analysis	1	2	0
Mat	9	11	9
Asphalt Content Testing Methods[b]			
Extraction	-	20	-
Nuclear Gauge	-	20	-
Ignition Oven	-	18	-
Plant Record	-	11	-
Tank Stickings	-	9	-
Specific Gravity	-	4	-
Compliance Measure[c]			
Quality Level Analysis	13	14	10
Absolute Average Deviation	7	8	7
Moving Average	6	7	6
Average	5	6	2
Range	3	3	3

[a]States may specify multiple locations for aggregate gradation and asphalt content.
[b]States may specify multiple testing options for asphalt content.
[c]One or more compliance measures may be specified within a state (i.e., may vary by property being tested).

by some states, where the amount of material produced for a given day defines the lot. Asphalt content is more frequently tested than aggregate gradation or mix volumetrics for both sublots and lots.

Hot-mix samples taken from the truck box are the most commonly specified location for sampling. Seventeen states are specifying coldfeed belt samples for acceptance of aggregate gradation. More of the states are specifying the aggregate gradation test for hot-mix samples at various sampling locations. Fewer states are sampling material for aggregate gradation and asphalt content at plant discharge (after mixing and before either storage or placement in haul trucks). Laydown operations are a frequent location for collecting hot-mix samples, with 2 states using the windrow and 11 states choosing a location behind the paver.

Equal numbers of states are specifying solvent extraction and the nuclear asphalt content gauge (NACG) test procedures to determine asphalt content (20 states each). It is expected that the industry will move more toward greater use of the ignition oven (18 states) because of environmental and safety concerns when using chemical solvents, a significant reduction in testing time, and reported improvements in testing precision. Additionally, the ignition oven has an advantage over the NACG because aggregate gradation can be determined from the burned sample. Plant recordation (11 states) and tank stickings (9 states) continue to be specified as methods to determine asphalt content. Back calculation of asphalt content using the specific gravity of the mix is used in four states.

Quality Level Analysis was the most frequently specified compliance measure for the three primary mix properties. Absolute Average Deviation was followed closely by Moving Average as the next most common methods for measuring specification compliance. Average and Range methods were less frequently used.

TABLE 2.5 Density acceptance attributes for 40 states

Attribute	Number of States Specifying
Sublot Size	
1 per 80 to 1 per 1,500 tons (1,360 Mtons)	19
330 to 660 yards (300 to 600 meters)	5
1 to 5 per day	4
Square yards	1
Square meters	1
Variable	1
None	9
Lot Size	
1 per 400 to 1 per 6,000 tons	17
5 to 10 per day	11
330 to 1760 feet (300 to 1,500 meters)	5
Total Per Mix Design	4
1 per shift	1
Cumulative	1
Variable	1
Sampling Method	
Nuclear Gauge	16
Core	15
Nuclear Gauge corrected to Core[a]	9
Reference[b,c]	
Theoretical Maximum Density	32
Laboratory Maximum Density	9
Test Strip	8
Compliance Measure[b,d]	
Quality Level Analysis	20
Average	8
Range	4
Moving Average	3
Average Absolute Deviation	3

[a]Number of cores for correcting nuclear readings ranged from 3 to 12.
[b]When the total number is less than 40, this means that some agencies did not provide a response.
[c]States may specify multiple options as a density reference.
[d]States may specify multiple options for compliance.

2.4.1.2 Density

Table 2.5 provides the acceptance attributes for density among 40 states. Similar to plant-produced mix properties, more of the states are using tonnage for sublots (19 states) and lots (17 states). Sublot sizes range from one test per 80 tons to one test per 1,500 tons. Other specified sublot and lot sizes include length, time, and area. Sublot length varied from 330 to 660 yards, time for sublots and lots ranged from 1 to 5 per day, and area included a 2,000-yd^2 (1,672-m^2) sublot and 5,000-yd^2 (4,200-m^2) lot.

Methods to sample pavement density include core samples (15 states), nuclear density readings (16 states), and correcting the nuclear density readings to core samples (10 states), with the number of correction tests ranging from 3 to 12. ASTM D2950-91, "Standard Test Method for Density of Bituminous Concrete in Place by Nuclear Methods," recommends that at least seven core densities and seven nuclear densities be used to establish a conversion factor *(22)*. The standard recommends that a new conversion factor be established any time a change is made in the paving mixture or in the construction process.

Density is calculated using several procedures, including Theoretical Maximum Density (TMD), Laboratory Maximum Density (LMD), test strip, and roadway voids. TMD is specified in 32 states, while 9 states specify LMD, and 8 states use test strip density. It is expected that more states will use TMD because Superpave testing protocols use this value for mix volumetric analysis. As is true for plant-produced mix properties, Quality Level Analysis is the most common compliance measure for density (20 states). The Average method is next most common (8 states), followed by Range (4 states), Absolute Average Deviation (3 states), and Moving Average (3 states).

2.4.1.3 Smoothness

Table 2.6 lists the acceptance attributes for pavement smoothness for 26 states (14 of the 40 states did not provide

TABLE 2.6 Smoothness acceptance attributes for 26 states

Attribute	Number of States Specifying[a]
Sublot Size	
1 per 0.1 mile	9
1 per 100 feet (100 meters)	3
1 per day	1
10 per day	1
1 per 200 tons	1
Section	1
Lot Size	
Total Project	12
1 per day	3
1 per mile	2
1 per 2,500 feet	1
1 per 160 feet	1
1 per 1,000 tons	1
Sampling Method[b]	
California Profilograph	12
Regular Straightedge	6
Profilometer	3
Rolling Straightedge	2
Mays Ride Meter	1
Compliance Measures	
Profile Index	16
Surface Variation	6
Quality Level Analysis	1

[a]When the total number is less than 26, this means that some agencies did not provide a response.
[b]States may use multiple methods for sampling.

smoothness acceptance specifications). States are primarily specifying length as a sublot size (12 states) and the total project as the lot size (12 states). The 1/0.1-mi sublot size is most common (9 states); this size is also found in the AASHTO *Quality Assurance Guide Specification (2)*.

The most common methods to measure acceptance of pavement smoothness are the California Profilograph (12 states) and regular straightedge (6 states). Other specified methods include profilometer (3 states), rolling straightedge (2 states), and Mays Ride Meter (1 state). The Profile Index is the most common measure for determining smoothness compliance, with 16 states specifying this method. Surface Variation measures are also specified (6 states), where the rate of defects over a certain distance are determined. Quality Level Analysis has been specified as a smoothness compliance measure in one state.

2.4.2 Pay Adjustments

Pay adjustments have become an integral part of many QA specifications. Table 2.7 lists attributes from pay factors from 40 state specifications. The surveys found that 95 percent of states apply some type of pay adjustment to the level of quality measured by the test results. In theory, pay adjustments are the difference between life-cycle costs from design and actual life-cycle costs from as-built construction. It is assumed that the pay adjustment quantifies the difference in reduced service life and an increase in the life-cycle costs.

There are many attributes to consider when making pay adjustments, including the type of adjustment and the method of combining or selecting the various construction tests. Pay adjustments have been developed for plant-produced mix properties (aggregate gradation, asphalt content, and mix volumetrics) and construction tests (pavement density and smoothness). The most common aggregate sieve size used for payment was the 75 µm sieve (25 states). The next most commonly specified sieve sizes are 4.75 mm and 2.36 mm, where these sieves define the particle size between coarse and fine aggregates. Many states specify both sieve sizes in their specifications. Percent of TMD is primarily used for payment of in-place density (29 states), agreeing with an earlier finding in the specifications where 32 states used TMD as a density reference value. Profile Index is used in 16 states, concurring with an earlier review of smoothness specifications.

TABLE 2.7 Pay adjustment attributes for 40 states

Attribute	Number of States Specifying
Type of Adjustment	
Factor	36
Fixed Rate	4
Bonus	21[a]
Aggregate Gradation Sieve Sizes	
12.5mm (1/2")	15
9.5mm (3/8")	15
4.75mm (#4)	17
2.36mm (#8)	18
2.07mm (#10)	10
1.18mm (#16)	7
600 m (#30)	10
450 m (#40)	10
300 m (#50)	10
75 m (#200)	25
Asphalt and Mixture Properties	
Asphalt Content	31
Air Voids	16
Voids in Mineral Aggregate	7
Stability	2
Voids Filled with Asphalt	1
Asphalt Penetration	1
Anti-Strip Additive	1
Moisture Content	1
Theoretical Maximum Density	1
Density	
Percent Theoretical Maximum Density	29
Percent Test Strip Density	6
Percent Laboratory Maximum Density	4
Smoothness	
Profile Index	16
Rolling Straightedge	1
Profilometer/Mays Meter	1
Method of Combination	
Weighted[b]	25
Minimum[c]	12

[a]Bonus provision is contained within the Factor or Fixed Rate.
[b]Weights summing to 1.0 are multiplied to each property then summed.
[c]Minimum individual pay factor of all measured properties is used.

The two primary methods for making a pay adjustment are the factor (or multiplier) and the fixed rate. The factor was the most common (i.e., 36 of 40 states) and applies a predetermined pay percentage to the bid price, based on measured quality from the test results. Table 2.8 provides an example of a pay factor table for density where 10 cores are sampled for each daily lot.

Four of 40 states use a fixed-rate adjustment that varies with the measured quality level, but does not use a percentage to adjust payment. A fixed-rate adjustment is applied based on the measured quality level, regardless of bid price. Table 2.9 provides an example of a fixed-rate table for air voids. Slightly more than one-half of the reviewed specifications (i.e., 21 of 40) have bonuses. Bonus adjustments are typically found with QLA-type specifications where work having an estimated PWL greater than 90 is awarded a bonus.

Several states are implementing pay factor equations, where a linear or non-linear equation is developed between the level of quality and payment. These equations are common for QLA-type specifications that measure a percentage of material within lower and/or upper specification limits (PWL). Table 2.10 provides several linear pay factor equations from the specifications, including the test property, sample size, and RQL assigned to the equations.

Most states are using weighted values to determine the pay factor for a lot (25 of 38), where weights that sum to 1.0 are multiplied to individual pay components then added. However, there has been variation of the coefficients assigned and no consensus given to the individual equation components. AASHTO refers to these weighted-type pay adjustments as Composite Pay Factors *(1)*. The minimum individual pay adjustment within a lot is specified by 12 states. For example, if asphalt content and density are a minimum PWL equalling 90 and laboratory-measured air voids are PWL equalling 80, the minimum value of PWL equalling 80 is assigned for payment.

2.5 PERSONAL INTERVIEWS

Personal interviews were conducted in order to characterize the working practices for applying QA specifications. The interviews allowed more "why" questions to be asked than are possible with questionnaire surveys and QA specifications. Given the wide variety of specifications and current practices, personal interviews provide focus on a few agency and contractor representatives with substantial experience in QA contracting.

TABLE 2.8 Pay factor for density using a multiplier (Ohio DOT)

Average of 10 Cores within a lot, %	Pay Factor
97.0 or greater	0.60
96.0 to 96.9	0.94
92.0 to 95.9	1.00
91.0 to 91.9	0.94
90.0 to 90.9	0.88
89.0 to 89.9	0.70
Less than 89.0	0.60

TABLE 2.9 Pay factor for air voids using a fixed rate (Arizona DOT)

Percent Within Limits, %	Fixed Rate, $/ton
100	+1.00
95-99	+0.50
90-94	0.00
85-89	-0.25
80-84	-0.50
75-79	-0.75
70-74	-1.00
65-69	-1.25
60-64	-1.50
55-59	-2.00
50-54	-2.50

Telephone and field interviews were conducted in 8 states with representatives from 5 agencies and 14 contractors. The Midwest interviews were conducted with representatives from Illinois, Indiana, Iowa, Michigan, and Ohio. In the Southeast, interviews were conducted with representatives from Alabama, Florida, and Georgia. These two regions were chosen because they produced a high concentration of survey responses from both agencies and contractors, both parties had working experience with their respective QA programs, and considerable information was attainable from concentrated areas within research time limitations.

Both subjective and objective information was collected from these interviews, including opinions on the current program, successes and setbacks, current staffing levels, cost data, testing time data, and historical production data. The interviews conducted in the Midwest and Southeast are summarized in Appendix E of the research team's final report. Important findings from these interviews were that agencies are limited by the number of on-site inspectors and the capacity to perform tests. Contractors generally assign one or two laboratory technicians at the plant and one density technician at laydown operations for each project. Contractors provide a full-time QC Manager to manage the company QC program and oversee the project technicians.

2.6 SUMMARY

Several areas of current practices in testing and inspection for HMA construction were researched. It was determined during the research that a common group of construction tests and inspections are thought to affect pavement quality. Specifications are written to require that these construction tests conform to levels thought to ensure pavement performance.

TABLE 2.10 Pay factor equations

State	Pay Equation	Test Property	Sample Size, n	RQL, PWL
New Jersey	$PF = 102 - 0.2 \times PD$ $PF = 10 + 1.0 \times PWL$[a]	Density	5	50
New Mexico	$PF = 55 + 0.5 \times PWL$	AG, AC, AV, Density	3 (minimum)	60
New York	$PF = 21.7 + 0.833 \times PWL$ ($PWL \geq 94$) $PF = 57.8 + 0.499 \times PWL$ ($PWL < 94$)	Density	4	5[b]
South Dakota	$PF = 55 + 0.5 \times PWL$	AG, AC, AV, VMA, Density	5	60
Vermont	$PF = 83 + 0.2 \times PWL$	AV	3 (minimum)	50
Virginia	$PF = 55 + 0.5 \times PWL$	AC, AV, VMA	4	40

AG = Aggregate Gradation
AC = Asphalt Content
AV = Air Voids
VMA = Voids in Mineral Aggregate
[a]Equation given as an example in the specification only.
[b]Remove and replace for material PWL < 5.

The specifications also encourage contractors to achieve design target levels with minimal variation. Pay factors have been developed to relate construction quality with performance by financially penalizing contractors when design targets are not met and/or there is excessive variation. Information gathered during the review of current practices provided necessary background for developing the work plan to determine minimum testing and inspection levels.

CHAPTER 3

WORK PLAN AND DATA COLLECTION

3.1 INTRODUCTION

The objective of this research was to develop a rational method to determine the minimum level of testing and inspection activities necessary to satisfactorily construct HMA overlays. Earlier, the rational method was defined as the steps that should be taken when developing the testing specification. The minimum level of testing and inspection was defined as the minimum testing and inspection resources that should be allocated for a given project. Satisfactorily constructing an HMA overlay was defined as meeting the specifications with specified test properties, test methods, and compliance measures. This chapter describes the data collection process used to address the research objective and these definitions.

Data to determine minimum testing and inspection levels were collected and analyzed for those elements of the specification thought to be indicators of satisfactory construction. However, as illustrated in the review of current practices, satisfactory construction and the precise definition of construction quality vary among state highway agencies and contractors. A review of 40 state QA specifications found that different test properties are specified (i.e., aggregate gradation, asphalt content, mix volumetrics, and/or density), different test methods are used for these properties (i.e., specifying either core samples or nuclear density gauges to determine density), and different compliance measures are used to describe the work statistically and to assign an appropriate pay factor. As an illustration, one state may believe a satisfactory level of quality is achieved when tests for the 75 µm sieve are within Job Mix Formula (JMF) limits of ±1.5 percent (South Dakota), while another state may use JMF limits of ±2.0 percent (Florida). Given this variation, the premise of achieving specification requirements appears to be a logical interpretation shared among agencies and contractors, since specifications describe in detail what is necessary to achieve the spirit of quality construction.

The work plan was designed to collect data from actual QA specifications. New QA specification data were needed because limited knowledge of inspection, sampling, and testing provided by historical data constrained the analysis capability. New data provided detailed accounts of inspection activities and sampling and testing procedures operating under existing QA specifications. Data collected during execution of the work plan were specific to construction.

New data provided a statistically sound experimental design to compare the mean and variation between agency and contractor laboratories (or a hired consultant working for either the agency or contractor). The review of current practices found that states specifying contractor data for acceptance perform verification testing for hot-mix samples using split samples, independent samples, or both. The work plan specified 24 split samples between the agency and contractor to determine the testing variation found when comparing split-sample test results on actual projects. This sample size provided a reliable estimate for the distribution of differences across a project. To further enhance the split-sample data, the Asphalt Institute was contracted to conduct three-way split-sampling on six projects.

3.2 PROJECTS

A total of 16 HMA overlay projects were obtained for data collection in six states. The basic attributes of each project, including mix design, existing base, length, and tonnage, are provided in Table 3.1. Appendix F of the research team's final report provides a detailed description of each project.

Other states contacted during the study were California, Colorado, Georgia, Maryland, North Carolina, South Carolina, South Dakota, Tennessee, Utah, and Virginia. Each state was evaluated in order to determine whether the agency and contractor had resources to participate in the study; several issues precluded the use of some states, such as interference with standard project functions and commitment of additional resources for increased testing. Each project included split-sample testing between the agency and contractor. Although many of the projects had the minimum of 24 split samples, some projects were unable to meet this requirement because of inherent project factors.

Project statistics are summarized in Appendix G of the research team's final report. The tables in Appendix G of the research team's final report provide the mean and standard deviation from the different laboratories on each project. Also provided are the summary statistics from the 24 split samples on each project. On some projects, 24 tests could not be obtained, while other projects had more than 24 tests.

The asphalt binder test data, conducted at the University of Wisconsin–Madison Bituminous Material Testing Laboratory,

TABLE 3.1 Projects selected for data collection

State	Project Index	City and Project Name	Mix Design	Existing Base	Length, km	Tons
Arizona	1	Phoenix I-10	Superpave	Milled Asphalt	10.9	100,000
	2	Benson I-10	Superpave	Milled Asphalt	9.6	71,000
	3	Topock I-40	Superpave	Milled Asphalt	3.7	25,000
	4	Wickenburg USH-60	Superpave	Existing Asphalt	6.8	25,000
	5	Wittmann USH-60	Superpave	Milled Asphalt	9.7	21,000
Florida	6	Lake City I-10	Superpave	Milled Asphalt	14.9	83,000
	7	Panama USH-231	Superpave	Milled Asphalt	14.3	50,000
Kentucky	8	Lexington I-64	Marshall	Rubblized PCC	6.6	60,000
	9	Florence STH-18	Marshall	Existing PCC	5.8	57,000
Minnesota	10	Winona I-90	Superpave	Existing PCC	12.8	36,000
	11	Minneapolis I-494	Superpave	Milled Asphalt	4.8	39,000
Ohio	12	Dayton I-75	Marshall	Existing PCC	6.4	50,000
	13	Findlay I-75	Superpave	Existing PCC	6.4	40,000
Wisconsin	14	Plainville STH-13	Marshall	Milled Asphalt	20.8	75,000
	15	Baldwin USH-12	Marshall	Milled Asphalt	11.0	76,000
	16	Milwaukee I-94	Superpave	Milled Asphalt	23.7	108,000

are summarized in Appendix H of the research team's final report. Asphalt binder samples were collected for 6 days of production on six projects; however, study resources limited testing to three projects. The raw data from asphalt binder testing are provided, as well as basic statistics.

3.3 TEST SELECTION

Tests were selected from the four construction processes that significantly influence quality. Testing and inspection of HMA can occur within the four construction processes provided earlier in Figure 1.1. These phases occur in the following sequence:

1. Materials,
2. Plant Mixing,
3. Mix Transport and Laydown, and
4. Compaction.

The selected tests were standard to construction, rather than non-standard tests such as L.A. abrasion and permeability. Table 3.2 summarizes the tests selected for the work plan.

3.3.1 Materials

3.3.1.1 Aggregates

Before production in the hot-mix plant, component materials are often tested to ensure that they have the same physical properties desired in the mix design. AASHTO recommends QC testing frequencies for aggregate gradation and aggregate properties *(2)*. Aggregate gradation frequencies are specified at 1 test per 500 tons during plant set-up and 1 test per 1,000 tons during normal production. Tests are specified at 1 per 1,000 tons for three aggregate properties: (1) fracture, (2) sand equivalent, and (3) Atterberg Limits.

A review of the agency survey responses revealed that most agencies were concerned with only two physical properties of the component aggregates: (1) gradation and (2) fractured faces. Other aggregate tests, such as L.A. abrasion, deleterious materials, sand equivalent, and insoluble residue, are used by some agencies; however, these tests are not routinely used during construction.

Aggregate stockpiles are tested to determine that the gradation is similar to that used in mix design. Much of this testing is accomplished by the contractor in routine QC operations. The agency can also require testing for specific aggregate properties that may be considered important to the perfor-

TABLE 3.2 Work plan tests

Construction Process	Testing Property	Test	Standard Test Method
Materials	Aggregate Gradation Asphalt Binder	Sieve Analysis Physical Properties	AASHTO T11 & T27 AASHTO MP1 & M226
Plant Mixing	Aggregate Gradation Asphalt Content Volumetric Properties	Sieve Analysis Extraction Nuclear Gauge Ignition Oven Bulk Sp. Grav. Max. Sp. Grav. Air Voids VMA VFA	AASHTO T11, T27, & T30 AASHTO T164 AASHTO T287 ASTM (under review) AASHTO T166 & TP4 AASHTO T209 AASHTO T269 & TP4 AASHTO TP4 AASHTO TP4
Transport & Laydown	Mix Temperature	Thermometer	N/A
Compaction	Density	Nuclear Gauge Core Samples	ASTM D2950 AASHTO T166

mance of the HMA pavement. An example of this type of test is the test for fractured faces of coarse aggregate (percent crushed particles).

The work plan included testing for aggregate gradation from stockpiles and coldfeed blends. It was initially planned that face fracture tests would be performed on 6 projects where the Asphalt Institute participated in sampling with the contractor and agency; however, this effort was unsuccessful because of lack of data.

3.3.1.2 Asphalt Binder

Quality assurance is needed to ensure that asphalt binders are consistent with the JMF and project requirements. Typically, asphalt binder assurance samples are taken daily. Although an agency may obtain several asphalt binder samples over the course of a project, only one sample may be tested from the entire project.

With continued implementation of Superpave, the asphalt industry will perform more certification of suppliers of performance-graded asphalt binders (PGAB). Currently, AASHTO specifies the certification procedures in Designation PP26-96. In the work plan, properties will be determined using the AASHTO M226 or the AASHTO MP1 PGAB specification.

The AASHTO M226 specification, “Viscosity Graded Asphalt Cement,” provides a standardized procedure to grade the viscosity of asphalt binders. AASHTO PP26-96 is a component of the M226 specification. Two fundamental tests are performed under M226—the Thin Film Oven Test (TFOT) and the Rolling Thin Film Oven Test (RTFOT). TFOT determines the viscosity of the asphalt binder prior to plant mixing, while RTFOT determines the viscosity of the asphalt binder after it has undergone short-term aging while mixing with hot aggregates.

AASHTO MP1, “Performance Graded Asphalt Binder,” is a comprehensive Superpave specification that details performance-based grading tests on the asphalt binder. Primary tests included in the MP1 specification are Dynamic Shear Rheometer (DSR), Rotational Viscometer (RV), Bending Beam Rheometer (BBR), Direct Tension Tester (DTT), RTFOT, and Pressure Aging Vessel (PAV). The DSR and RV tests measure the asphalt binder properties at high and intermediate temperatures, while the BBR and DTT measure properties at low temperatures. The RTFOT and PAV tests simulate aging and hardening of the asphalt binder.

Asphalt binder and coldfeed (hot bin) aggregates were analyzed at the University of Wisconsin’s Bituminous Materials Testing Laboratory. This portion of the study was primarily exploratory research for providing initial test results and statistics describing asphalt binders during production.

AASHTO T164, “Quantitative Extraction of Bitumen from Bituminous Paving Mixtures,” is a test specification to determine asphalt content using solvent extraction. Different methods exist for the solvent extraction procedure, including centrifuge, vacuum, and reflux extraction.

3.3.2 Plant Mixing

Production testing of HMA is where most testing and inspection resources are expended. Typically, the contractor will perform QC tests on the asphalt mixture. The agency will require a certain level of acceptance testing to ensure the quality of the mix. AASHTO recommends QC testing frequencies for aggregate gradation, asphalt content, and mixture properties *(1)*. Testing frequencies for aggregate gradation and mixture properties are specified at 1 test per 1,000 tons, while asphalt content is more frequent with 1 test per 500 tons.

Plant mix tests are separated into areas: (1) determining that the asphalt mixture components are similar to the design JMF and (2) determining the volumetric properties of the asphalt mixture. Standard tests for determining asphalt mixture components are asphalt content and aggregate gradation. Asphalt content has typically been measured using solvent extraction (AASHTO T164). In response to the need for alternative methods for determining asphalt content, two other methods, the nuclear gauge and ignition furnace, have been developed within the past 10 years in order to determine asphalt content.

Determination of volumetric properties is accomplished through different laboratory compaction techniques such as Marshall, Hveem, gyratory, and, recently, Superpave gyratory compaction. Each method may have minor modifications as determined by the state. Projects used for data collection used the Marshall and Superpave compaction methods. Emphasis was placed on projects using the Superpave gyratory compactor because Superpave is beginning to be implemented by agencies.

Collection of plant-mix samples occurred throughout production on each project. A two-way split sample between the contractor and agency occurred on these projects. A three-way split sample occurred between the Asphalt Institute, contractor, and agency on 6 projects for 24 samples. Agencies were requested to test all 24 split samples. Mutual aggregate gradation sieve sizes and asphalt content test methods were used.

3.3.3 Mix Transport and Laydown

Activities during the post-production phase of mix transport and laydown are typically confined to inspection activities. In this phase, the contractor and agency personnel may verify the temperature of the mix, the air and base temperatures, and the application of tack coat by inspection. The temperature of the mix is important. Depending on the binder type, overheating can damage the asphalt binder by premature aging, or, in the case of a modified asphalt, damage the added polymer. In these instances, mix temperature should be determined in the haul truck at the asphalt plant. The temperature of the mix prior to laydown is important only to the compaction process. A mix that is cool (i.e., less than 250°F) at laydown will require increased compactive efforts if inadequate density is to be prevented. Because the post-production activities of mix transport and laydown require inspection only, no formal tests were included in the work plan.

The air and base temperatures affect compaction. As is true of the temperature of the mat at laydown, cooler air and base temperatures can affect the compaction process, so that final compaction is inadequate.

3.3.4 Compaction

Determination of pavement density is a primary activity performed by either the contractor or the agency before acceptance of the completed pavement. Improper compaction can result in the pavement performing at lower levels than material properties would indicate. Inadequate compaction can enhance permeability and allow air to enter the pavement and prematurely age the mixture. The intrusion of water can also result in stripping in a moisture-sensitive mixture. AASHTO recommends QC testing frequencies for density at a rate of 1 test per 500 tons *(1)*. A primary purpose of density QC testing is to monitor and control the laydown and compaction processes to ensure that desired target levels are met with minimal variation.

Results of core sample and nuclear density tests from either the agency or contractor were collected. Additional density tests, other than those specified in the particular contract documents, were not collected. This research study focused on ensuring that mix properties and density data were collected. Although pavement smoothness is an important measure of HMA overlay quality, smoothness data were not collected and analyzed. There were concerns about how the smoothness data were sampled. Therefore, recommendations for smoothness testing are not provided in this report.

3.4 PROCEDURE FOR DATA ANALYSIS

A procedure was developed to analyze the data from the QA specifications by evaluating individual testing and inspection components. An integral component of the QA specifications that quantifies construction quality is the pay factor. As a practical matter, pay factors provide a way of describing the level of construction quality, assuming the pay factor has been structured correctly. Full payment or bonus payments indicate that the construction has a high level of quality. Reduced payments indicate that the work is substandard and it will result in diminished pavement overlay life. The pay factor is a function of many interrelated variables, as shown in Figure 3.1. It is necessary to evaluate several of these variables when establishing testing levels. Although all variables can be evaluated, those variables where data are available should receive priority.

Procedures were developed to evaluate test properties, variation, specification limits, sample size, and the actual pay factor. The Quality Level Analysis (QLA) method was cho-

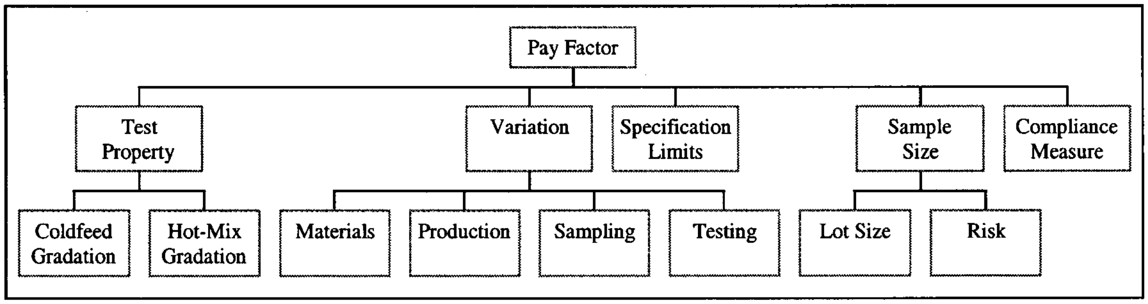


Figure 3.1. Pay factor variables.

sen as the compliance measure in the analysis of pay factors. The synthesis of current practices found that QLA was the most common compliance measure used to evaluate pay factors. QLA uses both the mean and standard deviation of the lot to estimate the percentage of material falling within specification limits (i.e., Percent Within Limits). QLA was chosen for the analysis of pay factors for several reasons:

- AASHTO recommends this method in both QA publications.
- Most state highway agencies have adopted this method.
- The QA literature indicates it to be more efficient than the other acceptance methods.
- This method combines the process mean and variation so that their interrelationship can be better understood.
- Sensitivity analysis can be easily conducted by adjusting the sample size and specification limits.
- Risk levels for both the contractor and agency can be readily estimated prior to construction.

Testable hypotheses were developed for several of the pay factor variables using traditional null and alternative hypotheses, designated as H_O and H_A, respectively. The null hypothesis is the hypothesis to be tested. A hypothesis is not proven, but rather, it is either accepted or rejected. A formal statistical test was conducted to accept or reject the null hypothesis.

3.4.1 Test Properties

The review of current practices revealed that final hot-mix aggregate gradation tests are thought to be an important indicator of construction quality and final pavement performance. An analysis was conducted to determine whether input aggregate gradation tests provide a reliable way to control final hot-mix aggregate gradation. Many contractors perform QC testing on aggregate coldfeeds or hot bins prior to mixing with asphalt binder. The purpose of coldfeed or hot-bin aggregate gradation testing is to provide the contractor with timely data to control the final aggregate gradation. In terms of acceptance testing, either input or final output gradation may be specified, however, there would be duplicate testing if both input and output aggregate gradation are specified on a project. If both tests provide similar information, it may be possible to eliminate one test and reduce testing levels.

Hypotheses tests were developed to determine if the mean and standard deviation are equivalent between blended coldfeed samples (input) and hot-mix samples (output) as follows:

Hypothesis #1

H_O: Mean of input and output aggregate gradation are equal ($\mu_{input} = \mu_{output}$).
H_A: Mean of input and output aggregate gradation are not equal ($\mu_{input} \neq \mu_{output}$).

Hypothesis #2

H_O: Variation of input and output aggregate gradation are equal ($\sigma^2_{input} = \sigma^2_{output}$).
H_A: Variation of input and output aggregate gradation are not equal ($\sigma^2_{input} \neq \sigma^2_{output}$).

Data collected from the field HMA overlay projects were then applied to the hypothesis tests. Analysis and calculations are given in Chapter 4.

3.4.2 Variation

Variation exists in HMA construction. The ability to stay within the specification limits is largely governed by the variability of the overall construction process, including materials, production, sampling, and testing. The industry can work to reduce variation, but this reduction costs money. Variation is a fundamental concern for both quality control and acceptance because it is used in determining the percentage of material within (or outside) specification limits. It is necessary to determine the reliability of the test results in the presence of variability, because test results are used for measuring construction quality. This research project offers methods to understand variation, how much specific items contribute to variation, and where it is cost-effective and worthwhile to expend more QC effort. It is assumed that variation is related to pavement quality and life-cycle costs.

A primary purpose of QC testing for plant mixing and density is to ensure that design target levels are met with minimal variation. Contractors use QC test data to understand the process and determine where changes are needed. If variation is to be controlled, contractors need to discern between random variation and variation assignable to controllable project factors. Testing levels can then be increased and allocated to those areas of production that have the greatest amount of variation. A practical way of interpreting variation for quality control is to determine whether variation exists between days.

A hypothesis test was developed to determine if variation *between* days is significant.

Hypothesis #3

H_O: Day-to-day variation is zero ($\sigma^2_{Day} = 0$).
H_A: Day-to-day variation is not zero ($\sigma^2_{Day} \neq 0$).

Data collected from the field HMA overlay projects were then applied to the hypothesis tests. Analysis and calculations are given in Chapter 4.

Successive samples taken during production will be subject to different degrees of materials, production, and sampling variation. These components of variation are further compounded with the introduction of testing variation when laboratory tests are performed on the sample. Many industry professionals think that testing variation is a small component of the total variation. It is also thought that testing variation is similar across several projects. If testing variation is a relatively high percentage of overall project variation, then testing levels should be increased for the testing component of the process.

A hypothesis test was developed to determine if testing variation is less than one-half of total project variation. This cutoff value was chosen to determine whether more individual samples, and testing of those samples, are necessary to measure and control materials, production, and sampling variation, or whether more testing on each individual sample is necessary to measure and control testing variation.

Hypothesis #4

H_O: Testing variation is less than one-half of total project variation.
H_A: Testing variation is greater than one-half of total project variation.

Data collected from the field HMA overlay projects were then applied to the hypothesis tests. Analysis and calculations are given in Chapter 4.

3.4.3 Specification Limits

Specification limits quantitatively define the allowable limits for typical HMA production processes and are considered determinates between acceptable and unacceptable quality. Development of specification limits for highway construction varies among the states, as described in *NCHRP Synthesis of Highway Practice 232*, "Variability in Highway Pavement Construction" *(23)*. A survey in *NCHRP Synthesis of Highway Practice 232* found that agencies are setting limits using "experience, engineering judgment, tolerances from other agencies, and standard precision statements more often than they use variability data from studies or projects" *(23)*. Some states are beginning to set specification limits using variability found in actual project data. In many states, specification limits used for acceptance are interchangeably used as QC limits.

Ideally, specification limits should be based on both the engineered tolerances of material properties, as measured by the resultant in situ performance, and the ability of the contractors to meet the limits. It has been difficult to accurately translate the relationship between construction variability and performance; therefore, a reasonable alternative is to set specification limits knowing a range of variability experienced in construction. Then, as contractors achieve a certain level of variability, the resulting performance can be better understood and a relationship developed. A procedure in the data analysis included an evaluation of specification limits using variability found in actual production data.

Specifications typically require that limits must be met for the individual lots, where sample sizes of mix properties may be n equals 4 and density sample sizes may be n equals 10. Sample size has a direct effect on the ability to meet specification limits when Statistical Quality Assurance (SQA) specifications are used, such as QLA-type that measure Percent Within Limits (PWL). QLA specifications use the standard deviation in the calculation for PWL, so standard deviations at different sample sizes were evaluated.

3.4.4 Sample Size

A fundamental issue for determining testing levels is defining an appropriate sample size and corresponding lot size to estimate the properties of the lot. In many cases, the sample size drives development of the lot size. For example, the time required to perform n equalling 4 hot-mix tests may be a common way to define a lot. Determining a lot size for tonnage is found by multiplying the number of samples by the sampling frequency.

The basis for constructing lots is governed by the four acceptance testing attributes: (1) practical, (2) economical, (3) statistical, and (4) equitable. The limited resources available, particularly time and people, apply practical constraints to any level of testing. The economic aspects of a testing specification must be considered. Specifying certain levels of testing, whether agency or contractor, requires time and money, and careful thought must be given when establishing testing levels. Statistical issues include the ability to characterize the test results using statistical principles. Finally, a primary goal when designing a testing specification is one that is fair and equitable to both the contracting parties.

The lot is a self-contained entity that makes a self-declaration of the work. Although lots within a project can be compared to ascertain any notable differences, an individual lot must stand on its own merit through estimated statistical properties. Thus, the lot must be designed correctly, have a sufficient sample size, and have reliable test results. One of the greatest challenges of acceptance specifications is accurately estimating material properties of the lot in the presence of variation from multiple sources. The true population properties for a given lot of material will never be truly known, and it is the responsibility of the agency to specify a sufficient number of acceptance samples. A large number of samples provides more data to estimate the desired test property. Sample size and lot size are used interchangeably; for example, 4 hot-mix tests per 4,000-ton lot or 5 density tests per 750-ton lot.

Another important consideration is whether to use time (days) or quantity (tons) to define a lot. Time-based lots assign tests to a given period of production, typically 1 day, while quantity-based lots assign tests to an arbitrary size, described earlier by tonnage, length, and area. The tests in a quantity-based lot can come from one or more days, depending on tonnage, length, or area produced. There are relative advantages and disadvantages when specifying either time (day) or quantity (tonnage) for sample sizes.

Constructing quantity-based lots using material from different days can influence the statistical estimates of the lot. Using PWL specifications encourages the contractor to reduce variability within lots to improve the likelihood of receiving full payment and/or incentive payments. Combining adjoining sublots from distinct populations into one lot could conceivably lead to multi-modal populations that are not normal, or lead to a normal distribution that has greater variability than that of the individual lots. When sublots are extrapolated across lots having different variation, a multi-modal distribution may be created using material from different populations. This will decrease the PWL because of variability between days. The acceptance specification must address this lot-to-lot variability issue.

First, a hypothesis test was developed to determine if test data are normally distributed within days. Then, a hypothesis test was developed to determine if test data are normally distributed between days. During the hypothesis testing, the effect of day-to-day variability was analyzed.

<u>Hypothesis #5</u>

H_O: Test data are normally distributed within days.
H_A: Test data are not normally distributed within days.

<u>Hypothesis #6</u>

H_O: Test data are normally distributed between days.
H_A: Test data are not normally distributed between days.

Data collected from the field HMA overlay projects were then applied to the hypothesis tests. Analysis and calculations are given in Chapter 4.

3.4.5 Pay Factors

Pay factors are an integral component of QA specifications. Agencies are developing pay factors that attempt to equate construction quality with life-cycle costs. Critical projects having high life-cycle costs may have severe pay factors, while less critical projects may have more lenient pay factors. Continuous pay equations were selected for an analysis of pay factors to remove the effects of “stepped” or “tiered” pay schedules where material falling on either side of the tier, by random chance, will obscure the outcome. Three pay factor equations identified during the synthesis of current practices were used in the analysis to provide a practical aspect to the results. These equations, shown in Table 3.3, were classified by severity of financial penalty: lenient, moderate, and severe.

The moderate and severe equations produce 100 percent payment at PWL equalling 90, but the values in the penalty range differ because of the slope in the equation. For example, at a measured PWL equalling 70, the severe pay factor is PF equals 80 while the moderate PF equals 90. Assuming a PWL greater than 90, PWL values in the bonus range are greater for the severe pay factor (maximum PF equalling 110) and less for the moderate (PF equalling 105) and lenient (PF equalling 103). Figure 3.2 illustrates the pay factor severity for the estimated PWL value.

A lenient pay factor may be more appropriate where the effect of construction quality has less effect on the life-cycle costs. For example, a temporary pavement or a highway with low AADT could be classified with lenient pay factors. Severe pay factors would be assigned to projects having high life-cycle costs, such as an interstate highway with high AADT. Projects classified under a severe pay factor must perform, and there is little margin for substandard quality. The effect of failure is significant in terms of user delays, maintenance costs, premature replacement, or reconstruction. The contractor is encouraged to provide high-quality construction and is awarded a larger bonus if the work is of excellent quality; however, the penalties are much greater than for other projects. Using different classes of pay factors can permit an efficient comparison among highway project classifications

TABLE 3.3 Pay factor used in data analysis

Pay Factor Severity	Pay Factor Equation
Lenient	PF = 83 + 0.2×PWL
Moderate	PF = 55 + 0.5×PWL
Severe	PF = 10 + 1.0×PWL

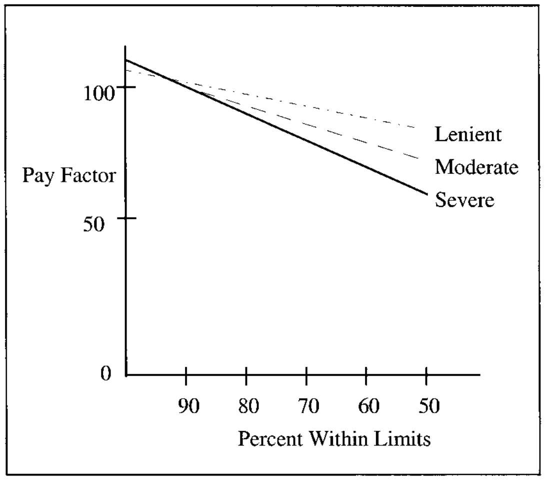


Figure 3.2. Relationship of pay factor severity and percent within limits.

and provide a rational basis for increasing testing levels on critical projects and reduced testing on less critical projects.

Different sample sizes have correspondingly different risk levels. During development of sample sizes for a lot, risks to both the agency and contractor should be evaluated for different sample sizes. Both the agency and contractor share risk during the acceptance process, designated as the α and β risks. AASHTO has defined these risks as follows *(1)*:

> Seller's Risk (α): The risk of rejecting "good" material. In highway construction, this is associated with the risk of a contractor having good material rejected by the owner. This risk is computed at the Acceptable Quality Level (AQL).
>
> Buyer's Risk (β): The risk of accepting "bad" material at reduced or full payment. In highway construction this is associated with the owners risk of accepting what is actually bad material. The risk is computed at the Rejectable Quality Level (RQL).

The α risk affects the contractor, because it is probable that the agency may reject what is, in fact, acceptable work. The β risk affects the agency, because it is probable that the agency may accept what is, in fact, rejectable work. These risks are a function of several attributes including (1) sample size, (2) AQL and RQL, and (3) the estimated PWL for the acceptance decision.

Operating characteristic (OC) curves are vital to developing a sound acceptance specification. OC curves illustrate the relationship of actual PWL values from construction with the Probability of Acceptance. With a developed OC curve from specified acceptance attributes, the agency and contractor will know how the acceptance program will operate prior to construction. A practical approach to interpret risk is to apply the pay factor to the Y-axis of the OC curve. This provides a visual way of relating the expected pay factor with the actual value for PWL. The expected pay factor is defined as the average of individual pay factors that would be expected with the specification. Figure 3.3 provides an example of an Expected Payment (EP) curve.

OC and EP curves can be developed from statistical tables, computer simulation, or commercial software. An FHWA-sponsored software package has been developed to construct OC and EP curves *(14)*. The software offers tremendous benefits to those developing and evaluating QA programs—it has eight operational programs and several practical applications.

One of the operational programs specifically developed to create OC and EP curves is OCPLOT. OCPLOT enables the user with a rudimentary knowledge of statistical quality assurance to develop curves for either pay adjustment or pass/fail

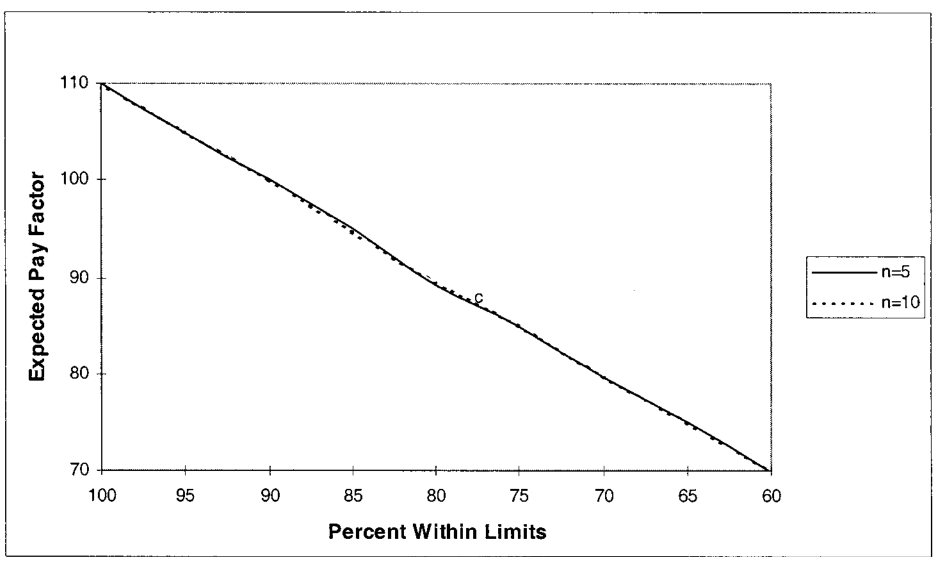


Figure 3.3. Expected payment curve for severe pay factor.

TABLE 3.4 Summary of developed hypotheses tests

Hypothesis	Description
1	H_O: Mean of input and output aggregate gradation are equal ($\mu_{input} = \mu_{output}$). H_A: Mean of input and output aggregate gradation are not equal ($\mu_{input} \neq \mu_{output}$).
2	H_O: Variation of input and output aggregate gradation are equal ($\sigma^2_{input} = \sigma^2_{output}$). H_A: Variation of input and output aggregate gradation are not equal ($\sigma^2_{input} \neq \sigma^2_{output}$).
3	H_O: Day-to-day variation is zero ($\sigma^2_{Day} = 0$). H_A: Day-to-day variation is not zero ($\sigma^2_{Day} \neq 0$).
4	H_O: Testing variation is less than half of total project variation. H_A: Testing variation is a greater than half of total project variation.
5	H_O: Test data are normally distributed within days. H_A: Test data are not normally distributed within days.
6	H_O: Test data are normally distributed between days. H_A: Test data are not normally distributed between days.

acceptance plans typically used in highway construction *(14)*. The sample EP curve in Figure 3.3 was developed for the severe pay factor using OCPLOT. The figure shows that the sample sizes of n equals 5 and n equals 10 will have a similar expected payment at both the AQL (PWL equals 90) and RQL (PWL equals 60).

A goal of developing statistically based QA specifications is to understand the effects of sample size on the pay factor and to minimize (within practical limits) the risks to both the agency and the contractor. A procedure in the data analysis was to analyze pay factors and risks so that an equitable specification can be developed.

3.5 SUMMARY

The work plan was designed to collect data from actual QA specifications. A procedure was developed to analyze the data from the QA specifications by evaluating individual testing and inspection components. Procedures were developed to evaluate test properties, variation, specification limits, sample size, and the actual pay factor. Testable hypotheses were developed for test properties, variation, and whether the data are normally distributed. Table 3.4 provides a summary of the six testable hypotheses developed to analyze the data. These testable hypotheses, along with specification limits and pay factors, are next analyzed using actual project field data.

CHAPTER **4**

PROCESS FOR DETERMINING TESTING LEVELS

4.1 INTRODUCTION

The process for determining testing levels begins by choosing whether to use contractor or agency data for acceptance. Once this decision has been made, methods of obtaining samples are given for either alternative. Next, the samples are tested for quality control and/or acceptance. Methods are given to determine both QC and acceptance testing levels. Testing levels have been defined as the frequency of sampling and testing material properties and ways of interpreting and arranging the test data for quality control and acceptance. Finally, a method is given for making a comparison between contractor and agency tests. Figure 4.1 illustrates how a systems approach was taken to determine testing levels. The following sections are found in this chapter: (1) Agency or Contractor Data for Acceptance, (2) Sampling, (3) Contractor QC Testing Levels, (4) Acceptance Testing Levels, and (5) Verification Testing Levels.

4.2 AGENCY OR CONTRACTOR DATA FOR ACCEPTANCE

A fundamental decision that must be made during development of any QA program is whether to use agency or contractor tests for acceptance. This decision will directly affect the level of resources required for both the agency and contractor on any project. Although many QA programs began by maintaining separate QC and acceptance functions, many programs are electing to combine QC data with acceptance data because of limited agency resources.

AASHTO states that the frequency of agency testing depends on whether the agency has decided to use contractor test results as a part of the acceptance decision *(1)*. If the agency does not use contractor results, then the agency's testing frequency will be relatively high and will remain at the same level throughout the project. If the contractor QC test results are used for acceptance decisions, then the required amount of agency testing can be reduced to the amount necessary to validate the contractor's results. Under this scenario, contractor testing frequencies are typically greater than, or equal to, agency testing frequencies. The contractor performs standard quality control while maintaining a mandated level of acceptance testing. The agency typically validates the reliability of the contractor tests through periodic verification samples (e.g., one verification test per every four contractor tests). As an owner and agent of the public, the agency must be assured that contractors have correctly obtained acceptance samples and produced reliable test results. When agency tests are used for acceptance, the contractor has greater flexibility for quality control.

An important concern when using contractor data for acceptance is developing trust and cooperation between the agency and contractor. Many agencies have elected to use independent sampling to verify the contractor's process, while others have used split-sampling. Each method has its own merits; however, the limitation of each must be understood. If the intent is to verify the contractor's entire process (material variability, production changes, sampling practices, and testing procedures), then independent-sampling may be more appropriate. Agencies and contractors must realize that independent verification makes it very difficult to decipher which source of variation causes differing readings between the agency and contractor. If the intent is to verify only the contractor's test results, then split-sampling is more appropriate. Thus, the decision of whether to use independent or split-samples to verify the contractor has different capabilities.

What a contractor elects to do for quality control is the contractor's choice. If the contractor data generated for quality control are used for acceptance, the agency will require a more formal and rigid plan. In either case, the agency maintains responsibility for overall quality assurance by evaluating contractor performance and accepting the work. From a resource perspective, the testing alternative to minimize the cumulative number of contractor and agency tests on a given project is by using contractor data for acceptance. Table 4.1 lists considerations when selecting agency or contractor tests for acceptance.

Using agency or contractor data for acceptance creates different combinations of testing on a given QA project. The states selected for the study have specified different testing requirements for agency or contractor acceptance data for hot-mix samples. Figure 4.2 provides testing alternatives from the six states in the study where 16 sequential samples are collected and are either tested by the contractor, agency, or both. Tonnage sizes for the 16 sequential samples are not equal among the states.

Arizona and Florida specified agency data for acceptance—contractor samples were used only for quality control.

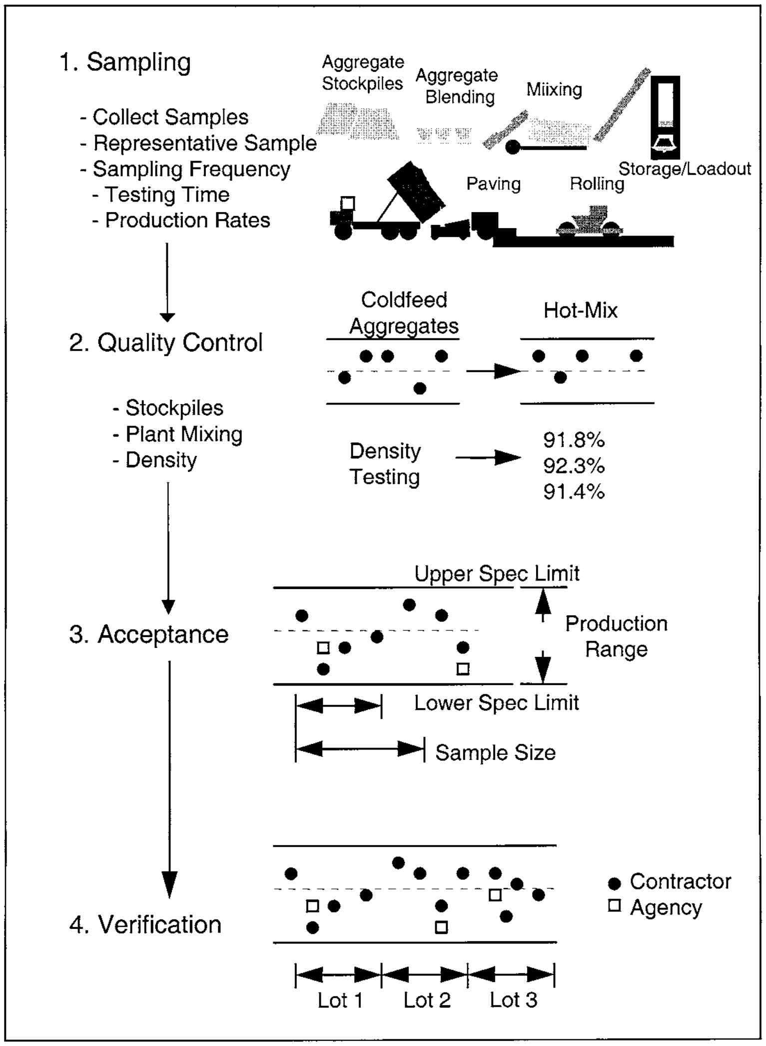


Figure 4.1. Graphical representation of HMA construction and testing components.

On some Arizona projects, the contractor tested all split samples collected with the agency, while on other projects certain samples were tested for quality control. In Florida, the agency acceptance samples and contractor QC samples were collected independently. Contractor QC sampling was time-based, while the agency sampling was quantity-based.

The remaining four states in the study used contractor data for acceptance, where the agency periodically tested a split sample to verify that the contractor tests were within a predetermined tolerance. A commonly specified procedure was one agency verification test per four contractor QC acceptance tests. Wisconsin verified the contractor QC acceptance tests at a rate of 1 test per every 10 contractor tests. In some states, the agency and contractor chose to compare several split-sample test results at the start of production to understand expected differences later in the project.

Different alternatives using agency or contractor data for acceptance highlight the effect this decision has on testing levels. Clearly, if contractor data are used for acceptance, the agency can reduce testing levels, but the agency may also increase testing at the agency's discretion. The implications of collecting split samples (companion) or independent samples must be considered and will be addressed later in the report.

4.3 SAMPLING

4.3.1 Principles

The purpose of sampling is to select and observe some part of the population so that an estimate can be made about the entire population *(24)*. The integrity of obtaining a *representative sample* must be firmly established and must not be compromised by a conventional approach of simply acquiring material to perform tests. With careful attention to the sampling design, unbiased estimates can be obtained for

TABLE 4.1 Considerations in choosing agency or contractor data for acceptance

Agency Acceptance Tests	Contractor Acceptance Tests
• Agency must have sufficient resources to achieve testing frequency. Also, tests must be performed when work is performed, versus after the project is complete. • Contractor and agency can use different types of test without dictating each other's testing procedures. For example, contractor uses coldfeed input gradation for control purposes while agency uses hot-mix gradation for acceptance. • Reduces possibility of contractor changing the production process with prior knowledge of when the test will be taken. • Quality control can occur as necessary without a mandated acceptance testing level, thereby providing greater contractor flexibility for conducting quality control. • Dispute Resolution System required if contractor questions agency test results.	• Agency has worked in partnership with the contractor and has reasonable assurance of contractor's honesty and integrity. • Agency can allocate limited resources across a greater number of projects. • Contractors must have sufficient resources for achieving lot size frequency. • Time-lag consideration for contractor QC testing frequency can influence desired sampling frequency. • A system of checks and balances is needed to determine the consistency of contractor test results. • Contractor and agency tests should be similar to have high degree of confidence. • Inconsistencies can occur between contractor and agency tests that affect assurances of the acceptance process. • Agency may be required to take additional tests to compare with contractor's initial results. • Contractor test results disqualified for acceptance and substituted with agency tests may require that the agency have available resources for increased substitute testing. • A Dispute Resolution System is necessary if inconsistencies develop between tests.

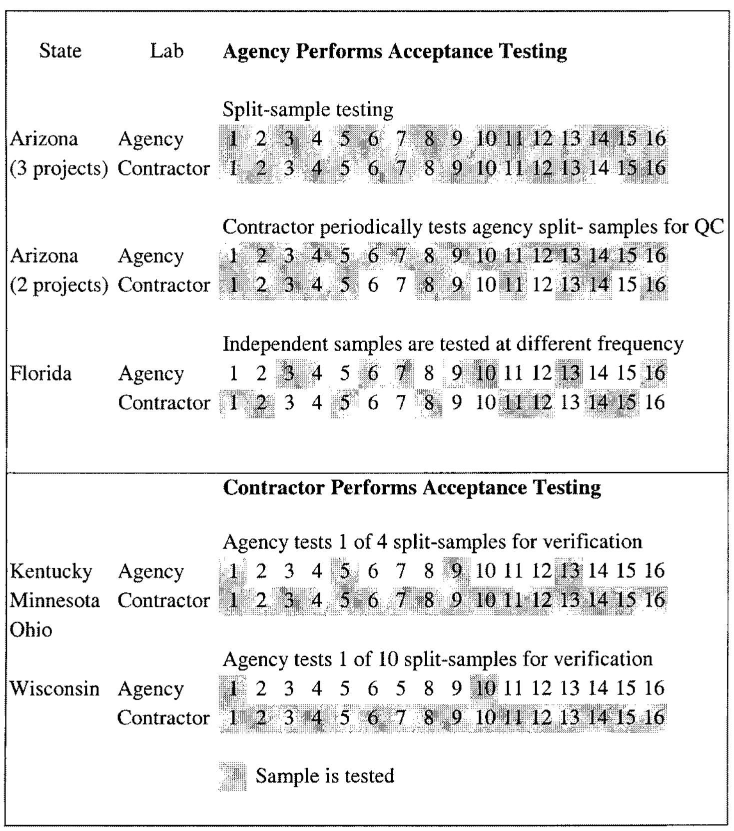


Figure 4.2. Relationship of agency and contractor data.

population quantities, such as the population mean and standard deviation.

Randomization is extremely important to the sampling process. Although many QA specifications have specified procedures for random sampling, the essential purpose of randomization must not be overlooked. Randomization ensures that each part of the population has an equal chance of being selected and it protects against unsuspected sources of bias. Violation of the randomization principle can produce biased samples that will inaccurately reflect the characteristics of the population.

Violation of the random sampling principle may occur when an overabundance of samples are collected and certain samples are electively discarded or not tested. Samples could be taken constantly during production; however, samples should be collected with the presumption they will be tested. The tested samples must align with the principles of simple random sampling or a modified version (e.g., stratified random sampling). Discarding certain samples from the population because of insufficient testing resources can severely violate the randomization principle. If samples are discarded, the random sampling procedure must be taken into account.

Figure 4.3 provides an example where a high level of samples are collected using stratified random sampling. A 4,000-ton lot is developed with equally sized 1,000-ton sublots, with four samples randomly collected within each sublot (Figure 4.3 (a)). The agency decides that it is unable to test all samples and chooses to discard several samples from testing. The incorrect method of discarding samples is shown in Figure 4.3(b). In this case, the number of samples within each strata is unequal and the principle of random sampling has been violated. Figure 4.3(c) shows the samples correctly removed by random selection in accordance with sampling principles.

An inconsistent sampling/testing frequency can show the estimation of population properties and may not conform with principles of random sampling or stratified random sampling. If the sampling frequency is very high, and there are insufficient resources to test the samples, then the sampling rate may need to be modified to avoid violation of randomization principles.

4.3.2 Location

AASHTO recommends that hot-mix acceptance samples be collected at the "roadway before compaction" *(2)*. There are two primary locations to sample hot-mix material for acceptance: (1) at the plant and (2) at laydown operations. Because obtaining representative samples is a requirement for acceptance, a comparative evaluation can be helpful in

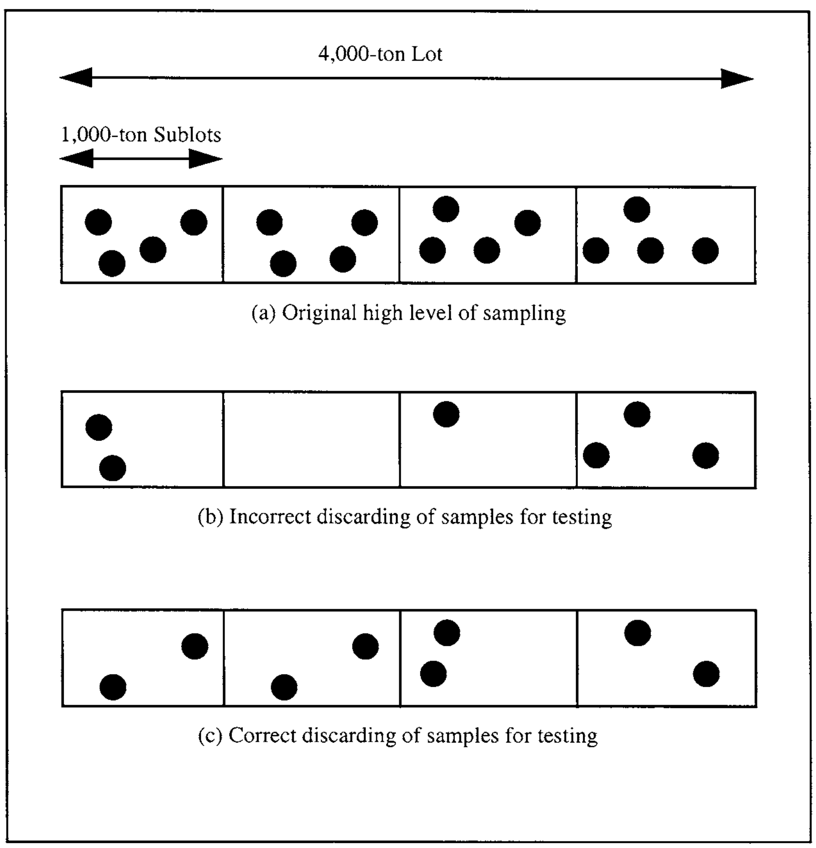


Figure 4.3. Examples of discarding samples.

understanding the advantages and disadvantages of each location. Table 4.2 lists attributes that should be considered when choosing between samples collected at the plant or laydown operation.

On projects where samples are taken at the mix plant, bias can be introduced if there is prior knowledge of when a sample will be taken. In the worst case, there exists a possibility for the plant operator to interact with the sampling technician and change plant settings immediately before or after sampling that can greatly affect the ability to collect a representative sample. Taking samples within a truck box presents a safety issue that can diminish the ability to collect a representative sample, although this is the most common sampling location for hot-mix. A sampling technician may lean over the edge of the truck box and collect only that material within reach, rather than entering the truck box and sampling material from several locations, as stipulated in many material sampling specifications.

Sampling at the laydown operation ensures that a more representative sample will be obtained because (1) material characteristics are believed to be more representative prior to compaction, (2) dependence between the sampling process and construction operation is removed, and (3) safety is improved. An increase in resource levels may be necessary to collect samples at the laydown operation, since an additional technician and vehicle may be required to handle and transport samples.

It is recommended that agencies and contractors evaluate these attributes and any other state-specific attributes when selecting a sampling location. Based on an evaluation of these attributes, and from observations during data collection, it is recommended that hot-mix acceptance samples be collected at the laydown operation. Contractor QC samples not used for acceptance may be sampled from the plant to expedite the sampling process. However, consideration must be given to the effects of obtaining a representative sample. Also, if the contractor and agency want to make a comparison of test results, sampling in different locations will introduce additional sources of variation.

Obtaining representative samples for density testing is also important. AASHTO recommends that density be sampled at the "roadway" *(2)*. Although there are no other locations at which to sample density, it is possible that core samples or nuclear density test sites may not be selected in strictly random fashion. A randomly chosen location may be substituted with a non-random location in an effort to "try to be fair." Such practices severely limit the ability to collect unbiased samples and defeat the principle of collecting representative samples, which, in turn, can lead to misinformation from the test results.

4.3.3 Collecting Acceptance Samples

The agency must be assured of the conditions surrounding the sampling of the material and that the integrity of the QA effort is maintained. AASHTO recommends that when split samples are used, the qualified agency testing personnel be

TABLE 4.2 Attributes for hot-mix sampling at plant and laydown operations

Attribute	Plant	Laydown Operation
Material Characteristics	• Material may segregate within the truck box at the sample face and introduce bias. • Sample may not represent AC absorption found immediately before compaction.	• Segregation possible in paver hopper or on mat. • Sample has time to absorb AC and is considered to be more representative for compaction.
Resources	• Requires a reduced amount of resources. • Less time for overall sampling cycle.	• Requires more resources (technician time and vehicle to transport sample). • More time for overall sampling cycle.
Interaction of Construction Operations and Sampling	• Possibility of changing plant settings immediately before or after sampling (over-conscientious or over-eagerness to sample), creating dependency between plant operations and the sampling process. • Selecting a "more representative" truck in "trying to be fair" can introduce bias.	• Trucking and paving operations are independent of sampling (bias of sampling process is minimized). • Removes (1) opportunity to change process immediately before or after sampling and (2) any dependency between plant operations and sampling process.
Safety	• Climbing into the truck box to obtain a representative sample presents a safety problem. • Unsafe sampling environment may influence the ability to obtain a representative sample.	• Improved safety because sampling is made at ground level. • Safety concerns exist from moving equipment and projects paved under traffic.

witnesses when the contractor personnel obtain the sample and split it *(1)*. The agency personnel will witness or assist in the sampling and splitting procedure and will take immediate possession of the agency portion of the split sample to ensure the validity of the split-sample comparison *(1)*. The agency can choose later whether to test the sample, particularly when contractor tests are used for acceptance, but they must first obtain the sample.

Trust and cooperation must be established between the agency and contractor during the sampling process. A dispute may arise over the integrity of the samples and having both parties witness the sampling event can reduce the likelihood of this happening. Agency witnessing of contractor acceptance sampling also removes the remote chance of a contractor substituting standard project samples with cataloged samples having predetermined material properties. It is recommended that agencies either collect their own samples or witness the contractor collect and split the sample and then have the agency field representative immediately take possession of the agency's portion of the split sample.

4.3.4 Sampling Frequency

Sampling frequency is a fundamental component when determining testing levels. Issues to consider include time constraints and whether to use time or quantity for sampling frequency.

4.3.4.1 Time Constraints

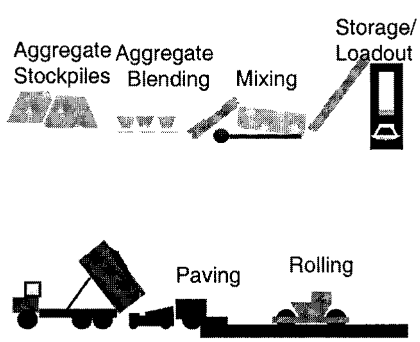
As a practical matter, time and staff to collect and test the samples is a major consideration in determining sampling and testing frequencies. Current HMA testing technology, governed largely by physical tests, has imposed time constraints on the number of possible samples collected and tested during a given production day. A limited amount of test data are available because of the inherent time to perform the entire sampling and testing cycle, as shown in Figure 4.4. A maximum of n equalling 4 hot-mix tests (e.g., aggregate gradation, asphalt content, Rice, and bulk gravities) is possible within a 10-hr production day, given standard resources found during the synthesis of current practices. It was possible to collect 10 samples within a day on three projects in the study. However, it required 2.5 days to complete testing at an equivalent staffing level. Doubling the testing resources could accommodate more sampling and testing.

Some laboratories, however, may only be capable of n equalling 3 tests per day. New Superpave™ technology can increase the testing time estimates for plant mixing because of greater material quantities and increased cooling durations. Density testing frequencies are much larger than mix properties. Twenty cores can be sampled and tested within a 10-hr work day, while as many as 50 nuclear tests can be collected (assuming one 4-min test duration at each test site).

It is possible that a larger number of samples could be tested than that routinely collected. A staged sequence of sampling/testing could be developed where a high frequency of samples would be tested in an assembly-line manner. An overlap of testing might also occur with multiple testing equipment and operators.

4.3.4.2 Time or Quantity for Sampling Frequency

A fundamental decision to consider when determining testing levels is whether to use time or quantity to sample the material. AASHTO discusses the unit of measure for contractor testing in the *Implementation Manual for Quality Assurance*, when frequency schedules may be derived in terms of time, quantity, or a combination of the two *(1)*. A schedule based on quantity may yield more samples than desired when production rates are high, and fewer samples than desired may result when production rates are low or intermittent. Schedules based on time yield fewer samples



Test	Time, hours	Number of tests per 10-hour workday
Coldfeeds	1.75	5
Hot Bins	2.00	5
AC Extraction	1.25	8
AC Ignition	1.00	10
Agg. Gradation	2.50	4
Volumetrics	2.25	4
Cores	0.50	20
Density	0.20	50

Figure 4.4. Minimum time requirements for typical HMA tests.

for high production and more samples for low production. For example, if 4 samples per day are specified and 1,000 tons are produced at low production, each sample would represent 250 tons. A quantity-based sampling frequency of 1 per 1,000 tons would only yield one sample per day.

The six states selected for the study specified both time and quantity for sampling frequency. Tables 4.3 and 4.4 provide a summary of sampling specifications for mix properties and density, respectively. The sampling frequency was identified interchangeably with sublot size for all states.

Table 4.5 provides the relative advantages and disadvantages when specifying either time or quantity for sampling. A time-based sampling frequency has the advantage of evaluating a given production time, where material and project characteristics may be better understood within time periods. A typical lot size from time-based sampling is 1 day.

With quantity-based sampling, material from different periods is assembled into one lot—the materials from these periods may have different production features. Quantity-based sampling has several advantages over time-based sampling. A given quantity of material can be traced through plant mixing and laydown operations, allowing for a discrete evaluation of the material throughout construction. Mix storage times may be an additional reason for selecting production as a testing frequency denominator, because material in storage for an extended period may not be tested with a time-based sampling frequency. An important characteristic of quantity-based sampling is that both small and large contractors must sample material at the same rate—this produces a testing specification that is fair to all parties.

Agencies have developed different units of measure for mix properties and density as identified in the synthesis of current practices. Currently, mix property units vary within sublot and lot sizes, where sublots are (1) *time* in hours and (2) *production* in tons; lots are (1) t*ime* in days and (2) *production* in tons. Units for density also vary for sublot and lot sizes, where sublots are (1) *production* in tonnes, (2) *length* in meters, and (3) *area* in meters^2; density lots are (1) *time* in days or shifts, (2) *production* in tonnes, (3) *length* in meters, and (4) *area* in meters^2. Tonnage is a measurement unit common to both mix properties and density for sublot and lot sizes. For either plant mix properties or density, the sampling frequency often becomes interchangeable with sublot size. For example, 1 sample taken per 1,000 tons translates to a sublot size of 1 test per 1,000 tons.

Referring again to Figure 1.1, AASHTO states that sublot and lot sizes for acceptance of plant mixing and density are "to be determined by the agency" *(2)*. The following sections provide a rational method for choosing frequencies for time-based and quantity-based sampling.

4.3.4.3 Time-Based Sampling Frequency

Time-based sampling frequencies can be easily obtained by using the testing time requirements provided earlier or state-specific time estimates. Aligning the testing time with sampling frequency ensures that a maximum number of samples can be collected and reasonably tested during production. Using the longest testing duration yields a minimum testing

TABLE 4.3 Mix property specifications

State	Measured Properties	Sampling Location	Sublot Size (Sampling Frequency)	Lot Size
Kentucky[c]	AC, AV, VMA	Truck	1 per half day (≤ 1996) 1 per 900 Mtons (1997)	2 per day (≤ 1996) 4 per 3,600 Mtons (1997)
Arizona[a]	AG (19.5mm, 9.5mm, 2.36mm, 75µm), AC, AV	Mat	1 per quarter day	4 per day
Florida[a]	AG (4.75mm, 2.36mm, 300µm, 75µm), AC	Truck	1 per 900 Mtons	4 per 3,600 Mtons
Wisconsin[c]	AG (12.5mm, 9.5mm, 4.75mm, 2.36mm, 600µm, 75µm), AC, AV, VMA	Truck	1 per 600 tons, then increases cumulatively by 300 tons: 900 1,200 1,500	4-sample Moving Average
Minnesota[c]	AG (12.5mm, 9.5mm, 4.75mm, 2.36mm, 75µm), AC, AV, VMA	Windrow	1 per 1,000 tons	4-sample Moving Average
Ohio[c]	AG (12.5mm, 4.75mm, 2.36mm, 75µm), AC, AV	Truck	1 per half day	3- or 4-sample Moving Average

[a] Agency performs acceptance testing.
[c] Contractor performs acceptance testing.

TABLE 4.4 Density specifications

State	Sampling Method	Sublot Size (Sampling Frequency)	Lot Size
Kentucky	Cores	1 per 2,500 feet	1 day (total varies on length paved)
Arizona	Cores	No specified size	10 per day
Florida	Cores	1 per 300 meters	5 per 1,500 meters
Wisconsin	Nuclear Gauge	No specified size	5 per 750 tons[a] and 7 per 750 tons[c]
Minnesota	Cores	No specified size	2 per 300 tons (lot increases for increased daily tonnage)
Ohio	Cores	No specified size	10 per day

[a] Agency performs acceptance testing.
[c] Contractor performs acceptance testing.

frequency that is robust to worst-case conditions. If an average testing time is used, it may not be possible to sample and test the material and keep pace with production. More data are desired to develop more reliable estimates, but this creates a situation where allowable sampling and testing time exceeds the sampling rates, resulting in an abundance of samples that may never be tested. Testing could be performed in a staged sequence. A greater number of resources (staff and equipment) would be necessary to categorize and manage the samples.

Seldom is the sampling frequency spaced at discrete intervals to match the testing time requirement. The random selection of hot-mix samples during production creates the possibility of two samples being collected within several minutes of each other. Such an occurrence highlights the need to design a sampling frequency while considering the practical effects of testing time.

4.3.4.4 Quantity-Based Sampling Frequency

A practical way of determining a quantity-based sampling frequency is to match sampling and testing time with construction production rates. Production rates, as typically defined in tons-per-hour, are used to express the rate at which HMA is produced and placed. By applying the provided testing times or state-specific estimates to an assumed 10-hr production day, a reasonable estimate for sampling and testing frequency can be calculated.

Similar to time-based sampling, using the longest testing duration and maximum production rates will ensure a minimum testing frequency that is robust to worst-case conditions. If the average production rate is used, it may not be possible for the sampling and testing process to match a maximum production level. Obviously, more data are desired to

TABLE 4.5 Comparison of time and quantity for sampling

Advantages	Disadvantages
Time	
• Evaluates an isolated period where process conditions can change between periods. • Separates production into time-based sequences where process levels can be understood from a daily perspective. • May be easier to account for work using days, rather than tonnage from multiple days.	• Amount of material in each sublot and lot can vary creating an unequal evaluation of material quantities. • Smaller contractors producing a smaller tonnage will require more testing than larger contractors producing a larger tonnage. • If an insufficient number of samples are collected prior to a work stoppage, it may be difficult to collect the remaining samples in a random manner.
Quantity	
• Same amount of material is used with each acceptance decision. • Testing frequency easily related to sublot and lot size. • Lot and sublot sizes can be managed to ensure samples are collected correctly. • Small and large contractors sample at the same rate.	• Material from multiple days may have different characteristics and confound the statistical properties of the lot. • Combining days with unequal means and variances may provide incorrect statistical estimates. • Specification limits should acknowledge the possibility of added variation from combining material from different days.

develop more reliable estimates, but this creates a situation where production exceeds sampling and testing rates, resulting in an abundance of samples that may never be tested. This could create a potential problem by incorrectly reassigning samples to stratified sublots and violating randomization principles.

When calculating a quantity-based sampling frequency, the time (in hours) is multiplied by production rate (tons-per-hour) to yield the testing frequency in tons. For example, 2.5 hr for volumetric tests multiplied by a 300 ton-per-hour production rate would yield a testing frequency of 750 tons (or 800 tons). In some states, the sampling and testing procedure for hot-mix samples, including all the standard tests on the sample, may require 3 hr. This time requirement will result in an increased sampling frequency of every 900 tons or 1,000 tons.

Table 4.6 provides average production within a day, along with the minimum and maximum tonnage, from the projects in this study. Production from several of these projects were based on an approximate 10-hr work day, and many of the minimum production days were the result of intermittent paving, caused by weather or plant delays. In some cases, production data from a project, or from several days within a project, could not be obtained.

Sampling frequency for hot-mix samples ranged from 444 to 1,654 tons. The estimates for Wisconsin projects were based on a 10-hr workday, when in fact the contractor worked 13- to 15-hr workdays. Sampling frequency for density was estimated in a range of 89 to 331 tons. States may elect to convert the density sampling tonnage to another measurement unit, such as length (meters) or area (meters^2).

Although these estimates may seem too high or too low, they provide a method with which to design a sampling plan. It is recommended that each state evaluate the production rates of several contractors within a state when designing a sampling plan. Testing levels should not dictate production rates. Imposing a production constraint on contractors to satisfy a predetermined sampling and testing frequency will cause an immediate increase in operating and overall project costs.

4.4 CONTRACTOR QC TESTING LEVELS

This section describes a process for determining QC testing levels. Testing levels were defined as the frequency of sampling and testing necessary to provide satisfactory quality control and product acceptance. The experimental design established in the work plan was not specifically designed to adjust levels of QC inputs in an effort to understand resulting outputs. It was not feasible to adjust a contractor's QC operations for this study, so emphasis was placed on gathering QC testing data from the field projects.

TABLE 4.6 Determination of quantity-based sampling frequency

		Daily Total Tonnage			Maximum	Plant Tests		Density (cores)	
Name of Project	Number of Paving Days	Min	Avg	Max	Production tons per hour	Estimated test time, hours	Sampling frequency, tons	Estimated test time, hours	Sampling frequency, tons
Arizona									
Phoenix	48	558	1,975	3,282	328	2.5	821	0.5	164
Topock	10	518	1,527	2,443	244	2.5	611	0.5	122
Wick.	9	2,728	3,464	3,788	379	2.5	947	0.5	189
Witt.	10	1,527	2,275	3,024	302	2.5	756	0.5	151
Florida									
Lake City	47	106	1,477	3,265	327	2.5	816	0.5	163
Panama[a]	12	350	1,667	2,900	290	2.5	725	0.5	145
Kentucky									
Lexington	25	182	1,776	3,869	387	2.5	967	0.5	193
Florence	32	154	1,783	3,714	371	2.5	929	0.5	186
Minnesota									
Winona	19	202	1,463	3,651	365	2.5	913	0.5	183
Ohio									
Dayton	13	566	1,143	2,019	202	2.5	505	0.5	101
Findlay	19	683	1,315	1,774	177	2.5	444	0.5	89
Wisconsin[b]									
Plainville	20	359	3,499	6,617	662	2.5	1,654	0.5	331
Baldwin	37	155	2,088	5,043	504	2.5	1,261	0.5	252

Days under 100 tons excluded from analysis.
Data from Minnesota I-494 and Wisconsin I-94 unavailable.
[a]Estimated by number of days to construct a 4,000-ton lot.
[b]Nuclear density gauge used on these projects. Core times used for sample calculation only.

The procedure development identified several interrelated, yet discrete, variables that a contractor can evaluate when determining QC testing levels. Testable hypotheses were developed for several of these variables. Using conclusions of the hypothesis tests, methods were developed for determining levels of aggregate gradation (raw materials), plant mixing, and compaction.

4.4.1 Aggregate Gradation (Raw Materials)

Aggregate gradation samples were collected from three Arizona drum-mix projects at two common sampling points: (1) coldfeed belts and (2) behind the paver. Sufficient data for both coldfeed and hot-mix aggregate gradation tests were not available from the other 13 projects. Coldfeed belt samples and mat samples were collected on a daily basis during production from the Arizona projects. Stockpile data were also collected on two of these projects. Stockpiles were built before and during HMA production, and aggregate gradation tests occurred as the stockpiles were being built.

The hypothesis test developed in the previous chapter was used to determine if the mean and standard deviation were equivalent between blended coldfeed samples (input) and mat samples (output). The conclusion of this test pertains only to data specific to those projects.

Hypothesis #1

H_O: Mean of input and output aggregate gradation are equal ($\mu_{input} = \mu_{output}$).
H_A: Mean of input and output aggregate gradation are not equal ($\mu_{input} \neq \mu_{output}$).

Hypothesis #2

H_O: Variation of input and output aggregate gradation are equal ($\sigma^2_{input} = \sigma^2_{output}$).
H_A: Variation of input and output aggregate gradation are not equal ($\sigma^2_{input} \neq \sigma^2_{output}$).

A statistical comparison of means was made using the t-test procedure, given in the AASHTO *Implementation Manual for Quality Assurance*, to determine if there was a mean difference between the two sampling locations. The *F*-test procedure was used to determine if the variations between sampling locations were equal or not. Small *p*-values would suggest rejection of the null hypothesis, and the means and/or standard deviations would be concluded to be different. The mean and standard deviation for the three projects, along with the *p*-values are shown in Table 4.7.

For the hypothesis test of the mean difference between coldfeeds and mat samples, results of the *t*-test suggest rejection of the null hypothesis. There was strong evidence to conclude that coldfeed aggregate gradation tests have different means than mat samples when evaluated across the entire project.

The data show an increase in the mean between coldfeeds and mat samples for nearly all sieve sizes. Shifts in the mean are caused by one or more of four variation components: (1) materials, (2) production, (3) sampling, and (4) testing. Production may cause shifts in the mean through aggregate breakdown from the coldfeeds through the drum mixer. Sampling the material at different locations can introduce bias because a representative sample is not collected. Testing with different laboratories can also introduce a systematic bias between test results. On the Phoenix and Benson projects, the

TABLE 4.7 Summary statistics for aggregate gradation samples

Project	Sampling Location	Mean of Percent Passing, %				
		n	19mm	9.5mm	2.36mm	75µm
Phoenix I-10	Coldfeed Blend	233	83.5	53.8	21.6	2.9
	Hot-Mix	212	84.8	56.0	23.3	4.8
	p-value		**0.000**	**0.000**	**0.000**	**0.000**
Benson I-10	Coldfeed Blend	100	88.8	42.3	20.0	3.3
	Hot-Mix	76	90.9	45.9	22.2	3.2
	p-value		**0.000**	**0.000**	**0.000**	0.162
Wicken. USH-60	Coldfeed Blend	36	88.1	52.0	22.0	3.0
	Hot-Mix	36	85.0	52.6	23.3	3.7
	p-value		**0.000**	0.525	**0.016**	**0.000**
		Standard Deviation of Percent Passing, %				
		n	19mm	9.5mm	2.36mm	75µm
Phoenix I-10	Coldfeed Blend	233	2.64	3.57	3.06	0.84
	Hot-Mix	212	3.80	5.66	2.36	0.51
	p-value		**0.000**	**0.000**	**0.000**	**0.000**
Benson I-10	Coldfeed Blend	100	4.39	4.95	1.99	0.63
	Hot-Mix	76	3.20	5.06	1.82	0.54
	p-value		**0.005**	0.835	0.403	**0.011**
Wicken. USH-60	Coldfeed Blend	36	2.09	1.86	1.54	0.23
	Hot-Mix	36	3.91	5.28	2.65	0.59
	p-value		**0.000**	**0.000**	**0.002**	**0.000**

contractor testing laboratory was used for the coldfeed tests while Arizona DOT was used for hot-mix tests. The difference in the 19 mm sieve can be explained by testing bias between laboratories, which was 1.5 percent (refer to Appendix G of the research team's final report for bias values). However, the smaller sieves (9.5 mm, 2.36 mm, and 75 µm) had large differences between coldfeed and hot-mix samples that cannot be explained by bias between laboratories. On the Wickenburg project, the contractor laboratory was used for both coldfeed and hot-mix testing, and a mean difference in sieves was found for three of the four sieves.

Different mean levels cause concern when using coldfeeds to control mean levels of the final hot-mix aggregate gradation. Contractors must be aware that bias may be introduced between the two sampling locations on other projects. If bias can be attributed to production, contractors may typically add a small percentage of baghouse dust during the mix design to simulate the generation of dust from aggregates during mixing. Contractors should continue to collect data to quantify changes in coldfeeds and hot-mix samples for percentage passing the sieves for a given aggregate type. Sampling at different locations has a definite possibility of introducing bias. Different test procedures for coldfeed and hot-mix aggregates are a potential source of bias.

The hypothesis test for the variation between tests concluded a rejection of the null hypothesis. The variation between the two sampling locations was different for many of the sieve sizes on these projects and no clear pattern was observed. On the Wickenburg project, variation in hot-mix samples was greater than coldfeed samples. In general, the data indicate that input variation from coldfeeds will be transferred to output variation in the hot-mix samples to some degree.

The large project data sets in the prior analysis resulted in greater ability to detect differences. A *t*-test and *F*-test were conducted for smaller data sets at the start of production to understand if an initial difference exists from the mean and standard deviation. Table 4.8 provides the summary statistics from the start of production.

There was moderate evidence to conclude that coldfeed and hot-mix tests will have different means with the smaller sample sizes. There was no evidence that the standard deviation between the tests was different, but it is apparent that variation from coldfeeds is passed directly to the hot-mix samples. This analysis suggests that detection of a difference in mean level and standard deviation is more difficult to determine with small sample sizes. Any difference becomes more pronounced as additional data are collected throughout the project.

An informal analysis of the stockpile data found that variation in stockpiles was transferred to the coldfeed belts and drum mixer. Opportunities to control initial variation begins with the stockpiles, and control of stockpile aggregate gradation requires QC testing and visual inspection during crushing, screening, belt transfer, trucking operations, and related material handling operations. Bin loading practices, feed rates, and belt transfer operations can generate additional variation, and visual inspection and periodic testing are necessary to control their effects on the hot-mix aggregate gradation. Accumulated tests from each stockpile can be used to develop estimates for the mean and standard deviation in order to understand potential effects on the mix design and to anticipate

TABLE 4.8 Summary statistics from comparisons at start of production

Project	Sampling Location	n	19mm	9.5mm	2.36mm	75µm
Mean of Percent Passing, %						
Phoenix I-10	Coldfeed Blend	4	84.8	54.0	22.3	2.6
	Hot-Mix	4	87.0	54.5	21.3	4.4
	p-value		**0.044**	0.845	0.551	**0.015**
Benson I-10	Coldfeed Blend	4	84.3	39.3	20.3	4.0
	Hot-Mix	4	89.8	46.8	22.5	3.4
	p-value		**0.035**	**0.087**	0.189	0.418
Wicken. USH-60	Coldfeed Blend	4	88.0	51.8	19.8	2.8
	Hot-Mix	4	87.0	54.2	23.4	4.1
	p-value		0.594	0.204	**0.035**	**0.048**
Standard Deviation of Percent Passing, %						
		n	19mm	9.5mm	2.36mm	75µm
Phoenix I-10	Coldfeed Blend	4	0.50	4.08	2.63	0.95
	Hot-Mix	4	1.41	2.64	1.74	0.34
	p-value		0.121	0.494	0.497	0.128
Benson I-10	Coldfeed Blend	4	3.20	5.38	1.89	1.00
	Hot-Mix	4	2.50	4.99	2.38	0.49
	p-value		0.694	0.905	0.715	0.281
Wicken. USH-60	Coldfeed Blend	4	1.15	1.50	1.71	0.43
	Hot-Mix	4	3.32	2.86	2.10	0.92
	p-value		0.116	0.316	0.747	0.241

variation sources prior to blending. Testing is also needed when large amounts of new aggregates are added to replenish stockpiles during production.

A contractor can use coldfeed samples to control aggregate gradation. However, the dissimilar levels of the mean and variation found between the two sampling locations create concern when used throughout the entire project. A bias was found for samples during the start of production, and this bias carried through the entire project. The advantage of coldfeed tests is that samples can be tested at a greater frequency than hot-mix samples because of reduced testing time.

A quantity-based testing frequency is recommended for controlling aggregate gradation to minimize aggregate particle size variation in the final mixture. The sampling frequency should be calculated using the methodology described earlier. It is also recommended that sample sizes of n equal to 4 or greater be collected to increase the power of the test. Agencies should recommend in the specifications that this procedure be performed by the contractor at the start of production if the contractor chooses to use coldfeeds to control final hot-mix aggregate gradation. Mandated testing levels for quality control should not be specified; only recommendations should be made to disseminate this finding.

4.4.2 Plant Mixing

The primary purpose of QC testing of the plant-produced mix is to ensure that specifications are met by achieving JMF targets with minimal variation. If variation is to be controlled, contractors need to discern between random variation and variation that can be assigned to known project factors. For example, is production variation between days significantly different or not? How this information can be translated to practice is key to effective quality control.

4.4.2.1 ANOVA—Random Effects Model

A formal analysis of variance (ANOVA) was used to measure sources of variation found in the plant mixing data. By conducting a formal ANOVA, it can be understood where sources of variation are found. This information can then be used to design testing levels that manage those sources of variation.

The Random Effects Model, also known as the Model II ANOVA, was used to analyze variation in the collected data. The Random Effects Model is appropriate where there is a single classification of data and no specific treatments are applied to any of the data *(25)*. The objectives of the Random Effects Model are different from those of the Model I ANOVA. The Model I ANOVA makes a deliberate comparison among specific treatments. Because there are no specific treatments in HMA construction, the Random Effects Model was selected for the analysis.

Figure 4.5 shows the hierarchical structure of the Random Effects Model with testing data from one laboratory. Two primary stages of sampling material to meet specification requirements are lots and sublots. Lots are typically 1 day's production or a specified tonnage, and sublots are collectively used to create lots. Sublots can be either one of several tests within a day or equally sized tonnage within a predetermined lot tonnage. The sublot sample can be split into two, three, or four specimens for individual testing. Tests that have multiple specimens are laboratory bulk-specific gravity (G_{mb}) and laboratory air voids. Currently, there are two specimen tests with the Superpave™ Gyratory Compactor that produce two G_{mb} values. In Marshall testing, there can be three or four specimens to determine G_{mb}. When using either Superpave™ or Marshall testing, the individual G_{mb} specimens are averaged to produce a single G_{mb} value. Likewise, two laboratory air voids specimens are averaged to produce a single air void value. Core sample testing produces only one G_{mb} value, while nuclear gauge testing for density can obtain multiple G_{mb} values for each sample (i.e., individual test site). Variation of multiple specimen tests within each sample provides a direct measure of testing variation.

An example of this hierarchical structure for the Arizona I-10 (Phoenix) project is shown in Figure 4.6. The contractor collected split-samples with the agency on April 21 through 23, 1997, and designated them as Lots 5 through 7, respectively. During construction of Lot 6 on April 22, 1997, four random samples (designated as sublots) were taken from behind the paver during nighttime paving with the first sample taken at 10:06 p.m. and the last taken at 2:06 a.m. The material from Sublot 2 is divided into three 5,000-gm samples for Superpave™ gyratory compaction. This example shows the complete structure for Sublot 2 within Lot 6, but the same concept applies for all other lots.

The hierarchical structure of this figure allows an allocation of variation to the lot, sublot, and specimen testing

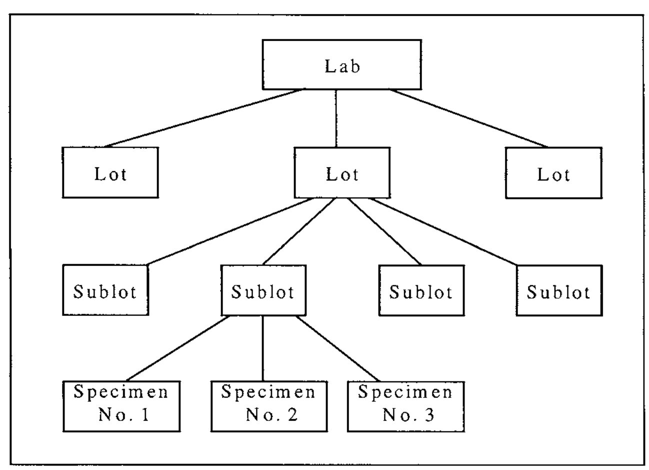


Figure 4.5. Hierarchical structure of testing data for one laboratory.

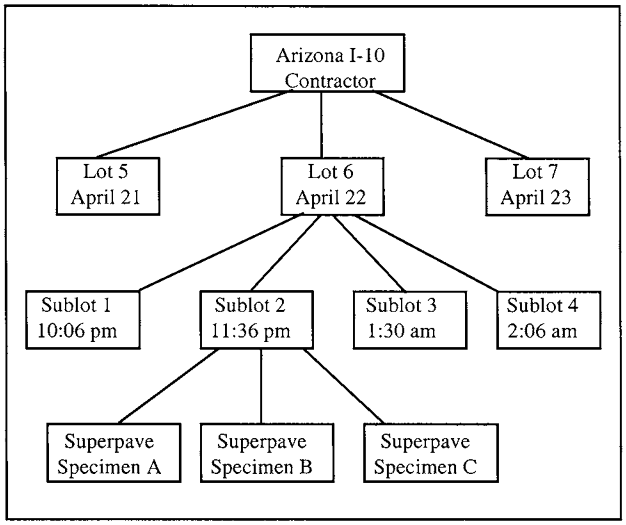


Figure 4.6. Arizona I-10 data applied to hierarchical testing structure.

components. These variation components are defined using Equation 4.1.

$$Y_{ijk} = \mu + L_i + S_{ij} + M_{ijk} \qquad (i = 1, \ldots, I; \quad j = 1, \ldots, J; \; k = 1, \ldots, K) \tag{4.1}$$

where

$L_i \sim N(0, \sigma^2_L)$ Lot;
$S_{ij} \sim N(0, \sigma^2_S)$ Sublot; and
$M_{ijk} \sim N(0, \sigma^2_M)$ Specimen Testing.

Y_{ijk} is the observed kth subsample test of the jth sublot from the ith lot. The term μ is the mean of all tests within a project for a particular mix property (e.g., aggregate gradation and asphalt content) or density. The term L_i represents the variation between the ith lots. L_i is a random variable with mean zero and variance σ^2_L. The term S_{ij} is used to express the sublot for the jth test within each ith lot. This term can be thought of as material variation measured on a lot basis to determine compliance with specifications. The distribution has mean zero and variance σ^2_S. The term M_{ijk} is the specimen test for each kth Marshall or Superpave™ specimen from the jth sublot and ith lot. The distribution of M_{ijk} has mean zero and variance σ^2_M. This term is removed from the model where no multiple specimen testing is used (e.g., aggregate gradation and density cores). The terms L_i, S_{ij}, and M_{ijk} are assumed independent. Table 4.9 provides the ANOVA table for each variance component, degrees of freedom, and expression for the Expected Mean Square (EMS). The lot, sublot, and testing variances are then estimated using the EMS.

Many HMA construction tests only include the variance components for lot and sublot, such as aggregate gradation and asphalt content. Specimen testing is only found with G_{mb} and air voids tests, and some maximum specific gravity (G_{mm}) tests. Tests not having specimen testing eliminate this variance component from the model. Most of the analysis in the study was conducted with only the lot and sublot components.

TABLE 4.9 ANOVA table for one laboratory

Source	Degrees of Freedom	Expected Mean Square
Lot	I-1	$\sigma^2_M + K\sigma^2_S + JK\sigma^2_L$
Sublot	I(J-1)	$\sigma^2_M + K\sigma^2_S$
Specimen Testing	IJ(K-1)	σ^2_M
Total	IJK – 1	-----

Of extreme importance in quality control is whether variation between lots is equal or different (days or tonnage). From a practical viewpoint, this is of great interest when designing testing levels for quality control. Because the goal is to meet specification requirements by achieving design targets with minimal variation, it must be understood where the greatest amounts of variation can be found during the plant-mixing process so that testing levels can be allocated to those areas. Many lots are created rather arbitrarily with a select number of tests (i.e., n equal to 4 or n equal to 5), so the data are partitioned into days to provide a practical way of interpreting the data. A fundamental question is whether variations between multiple days are similar or different from each other. If between-day variation is equal, it would indicate that the plant-mixing process may be a random process, having discrete assignable causes of variation not primarily attributed to days. If, however, between-day variation is not equal, this would suggest that variation can be attributed to a difference in characteristics found between days of production.

A hypothesis test was used to determine if variation *between* days is significant.

<u>Hypothesis #3</u>

H_O: Day-to-day variation is zero ($\sigma^2_{Day} = 0$).
H_A: Day-to-day variation is not zero ($\sigma^2_{Day} \neq 0$).

Two standard statistics were calculated and used to determine significance: (1) *F*-value and (2) *p*-value. The *F*-value was calculated from the ratio of variances, and then plotted on the *F*-distribution to determine a probability level of significance, or *p*-value. Those *p*-values equal to or less than 0.10 would conclude rejection of the null hypothesis. Equation 4.2 shows how the *F*-value is calculated when there is no specimen testing.

$$F_{Lot} = \frac{MS(Lot)}{MS(Sublot)} \tag{4.2}$$

4.4.2.2 Results of ANOVA—Mix Properties

Table 4.10 provides a summary of mix properties analyzed with the Random Effects Model. Lots were designated as days, and many of the daily sample sizes were 2, 3, or 4 sam-

TABLE 4.10 Summary analysis of mix properties between days

Project	Lab	n	Aggregate Gradation									Asph.	Gmb	Gmm	Voids	VMA	VFA
			12.5mm	9.5mm	4.75mm	2.36mm	1.18mm	600um	300um	150um	75um						
AZ Phoenix I-10	Contractor	24		X		X					X	X	X		X	X	x
	Agency	24		x							X	x	X	X	X	X	X
	Asph. Inst.	24											X			X	
AZ Benson I-10	Contractor	16						x									
	Agency	136			x	x	X	X	X	X	X	X	X	X	X	x	X
AZ Topock I-40	Contractor	52		X		X					X		X	X	X	X	X
	Agency	52	x	X		X					X		X	x	X	X	X
AZ Wickenburg USH-60	Contractor	36		x								x		x			
	Agency	36												X			
AZ Wittmann USH-60	Contractor	40						x				x	X		X	X	X
	Agency	40						x			x	X	X	X	X	X	X
Florida I-10	Contractor	24	x	x													
	Agency	24															
	Asph. Inst.	24	x	X													
Florida USH-321	Contractor	34	x	x		X	x	x			x	x	x		x	x	
	Agency	27		x	x	X	X			X		x	X		X	X	x
Kentucky I-64	Contractor	24											X	X	X		
	Agency	24	x			x	x	x					X	x	X	x	X
	Asph. Inst.	24			x	X	X	X	X	X	X	x	X		X	x	X
Kentucky STH-18	Contractor	30					x										
	Agency	21												X		x	
Minnesota I-90	Contractor	36		x							x	x		X			
	Agency	26									X			x			
Minnesota I-494	Contractor	35			X	X					X		X	x		X	
	Agency	34			X	X					x		X		x	X	x
	Asph. Inst.	24			x	x					X		x			x	
Ohio I-75 Dayton	Contractor	41					x	x			X	X					
	Agency	24	x		x		x		x		x	x					
Ohio I-75 Findlay	Contractor	29		x	x												
	Agency	17				x		x									
	Asph. Inst.	17						x	x					x			
Wisconsin STH-13	Contractor	39						x	X	X	X		x			x	
	Agency	24						x							x		
Wisconsin USH-12	Contractor	74	x	x	X	X	X	X	X			x	X		X	X	x
	Agency	24	x	x	x			x				x					
	Asph. Inst.	24															
Wisconsin I-94 19.0mm	Contractor	44				x	x	X	x				X		x	X	
	Agency	25	x		x	x	x	x							x		
Wisconsin I-94 12.5mm	Contractor	45						x	x				X		X	X	x
	Agency	25						x	x	x	x			x	x		x

X denotes Highly significant (p-value < 0.01). Test unavailable for property
x denotes Significant (p-value < 0.10).

ples per day. Both contractor and agency data were evaluated. The laboratory used for acceptance had the greater number of tests on all projects. On six projects, there were data from the three-way split-sampling between the contractor, agency, and Asphalt Institute.

Based on the data from these projects, the hypothesis test concluded that there is significant variation between days. Successive days of production could have no between-day variation. However, a series of production days compared within a project have between-day variation for one or more mix properties. Most projects had numerous tests with significant between-day variation. At first this finding may seem rather trivial or obscure, but when considering there are small sample sizes (three or four tests per day) having diminished sensitivity to detect differences, this variation is large.

A visual way of interpreting between-day variation is shown in Figure 4.7. This figure illustrates where significant between-day variation was found for air voids on the Wisconsin USH-12 project. An upward trend was observed on Day 4, downward trend on Day 5, and within-day variation on Day 6. Variation was relatively consistent within days. However, the variation between days was relatively high from

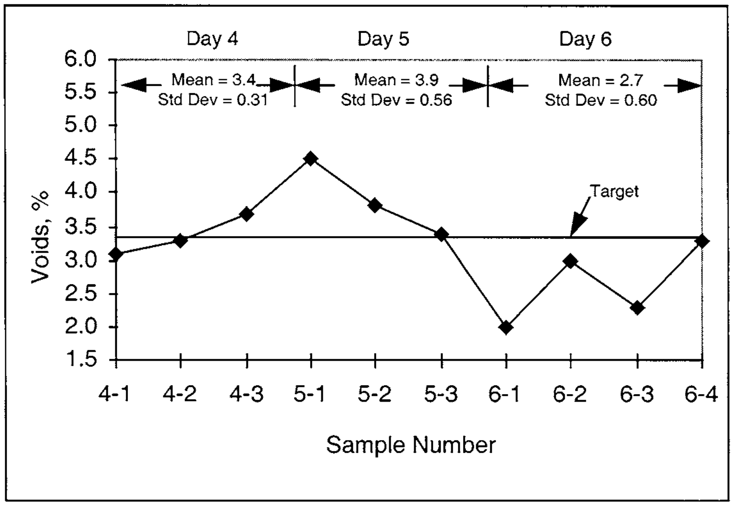


Figure 4.7. Significant between-day variation (Wisconsin USH-12 data).

daily trends and different means (F equal to 5.07 and p-value equal to 0.0435 computed with numerator and denominator degrees of freedom of 7 and 2, respectively).

Figure 4.8 provides an illustration of the air voids data on the Minnesota I-90 project where there was no significant variation between days (F equal to 0.69 and p-value equal to 0.4911 computed with numerator and denominator degrees of freedom of 7 and 2, respectively). There were smaller changes in the daily mean, as compared with the Wisconsin project. Test results were dispersed randomly at a level slightly below the target value, with an upward trend on Day 11 and random dispersion on the other days.

The analysis provides insight for contractor QC testing levels. Contractors can expect significant variation in the daily production of the plant-mixing process. A necessary component when establishing testing levels for quality control is to detect assignable causes of variation on a daily basis. Between-day variation can be attributed to several assignable causes, such as material properties, environmental conditions, personnel decisions, equipment, and construction methods. Controlling variation requires constant feedback-testing to identify discernible trends or variation and make appropriate adjustments so that target levels are met.

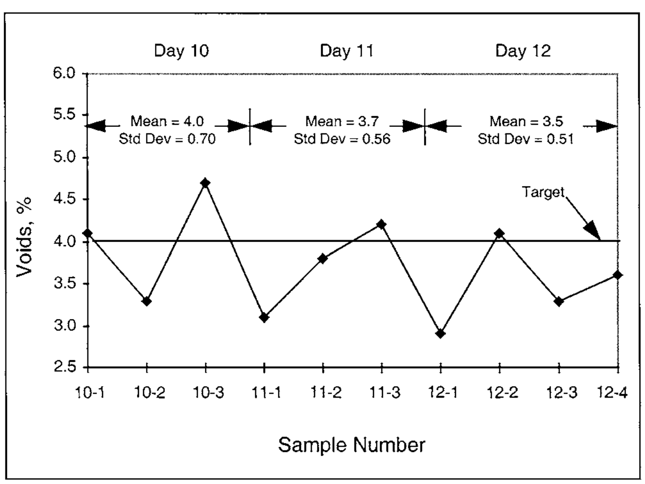


Figure 4.8. No significant between-day variation (Minnesota I-90 data).

This analysis demonstrated the ability to detect a source of variation that was not highly visible when scanning the data. The analysis was structured to meet an objective of the study, determining a rational method for contractor QC testing levels, but this same approach could be applied to other random effects of the plant mixing process. Examples include weather effects (e.g., moisture on stockpiles and removal of fines from rainfall), loading practices (e.g., operator working a different face at the stockpiles), belt speeds (e.g., oscillation, bearing wear, and proportioning), and asphalt binder metering (e.g., pump condition and pressure surge).

A rational method to determine QC testing levels for the plant mixing process is to allocate QC testing to the significant sources of variation. If few testing resources are available, the contractor should ensure that samples are collected and tested within each day of production, rather than high-frequency testing on periodic days. In other words, daily testing is recommended given the presence of significant between-day variation. Specifications should be written that recommend the contractor perform QC testing daily. It is recommended that a time-based or quantity-based sampling frequency be developed using the methodology provided earlier. It is further recommended that contractors use control charts to assist in discerning any trends or unusual behavior evident in the process.

4.4.2.3 *Testing Variation*

The ability to stay within specification limits is largely governed by the variability of the overall construction process, including the variation components of materials, production, sampling, and testing. Prior to laboratory testing, the sample has been subjected to several variation sources, including materials (e.g., aggregate size, aggregate composition, asphalt content, and asphalt grade), production (e.g., bin percentages, conveyor belt speeds, and asphalt flow rate), and sampling (e.g., truck or mat and location within sampling region). These variation components are further compounded with the introduction of testing variation when laboratory tests are performed on the sample.

A basic ANOVA model was developed to measure air voids variation from two sources: (1) Materials + Production + Sampling and (2) Testing. The production component could not be measured directly because detailed records of plant operations and process adjustments were not available (e.g., coordination issues, accuracy of data, and lack of on-site engineers to supervise data collection). If these data could have been obtained systematically, it would have been possible to isolate and measure this variation component. The sampling variation component could not be directly measured during field data collection because it was decided to collect only split samples for comparing contractor and agency test results. If the experiment had been designed to isolate the sampling component, a direct estimate of sampling variation would have been possible. Thus, the materials, production, and sampling components were combined and treated as one cumulative variation component: Materials + Production + Sampling (MPS).

The testing variation component was derived from the sample being split into two, three, or four specimens for individual air voids testing. Currently, there are a minimum of two specimen tests with the Superpave™ Gyratory Compactor, and a minimum of three or four specimens with Marshall testing. When using either Superpave™ or Marshall testing, the individual air voids tests are averaged to produce a single air voids test value. Variation of these multiple specimen tests within each sample provides a direct measure of testing variation.

MPS and Testing variation components are defined below using Equation 4.3:

$$Y_{ij} = \mu + MPS_i + T_{ij} \qquad (i = 1, \dots, I; \quad j = 1, \dots, J) \qquad (4.3)$$

where

$MPS_i \sim N(0, \sigma^2_{MPS})$ Materials + Production + Sampling; and
$T_{ij} \sim N(0, \sigma^2_T)$ Testing.

Y_{ij} is the observed jth specimen test from ith sample. The term μ is the mean of all tests within the project for air voids. The term MPS_i represents the variation between the ith samples from materials, production, and sampling, where MPS_i is a random variable with mean zero and variance σ^2_{MPS}. The term T_{ij} is used to express the jth Marshall or Superpave™ specimen test within each ith sample, having a distribution with mean zero and variance σ^2_T. The terms MPS_i and T_{ij} are assumed independent. Table 4.11 provides the ANOVA table for each variance component, degrees of freedom, and expression for the EMS. The MPS and Testing variances are then estimated using the EMS.

The estimated variance for MPS and Testing can determine the relative percentages of variation they provide to total variation. A hypothesis test was conducted to determine if testing variation is a smaller percentage of total project variation. A value of 50 percent was chosen because it is necessary to determine whether more individual samples, and testing of those samples, are necessary to measure and control materials, production, and sampling variation, or whether more testing on each individual sample is necessary to measure and control testing variation.

<u>Hypothesis #4</u>

H_O: Testing variation is less than one-half of total project variation.
H_A: Testing variation is greater than one-half of total project variation.

The laboratory with the largest project data set was selected for the analysis. Some project data were not available or were incomplete. Table 4.12 provides the percentage of variation for MPS and Testing from 13 projects.

Testing variation of air voids ranged from 6.4 percent to 38.6 percent of the total project variation. It can be concluded that testing variation is a smaller percentage of total project variation, and the null hypothesis is accepted. There appeared to be more testing variation from the Marshall projects, with the exception of two Superpave™ projects (Arizona I-10 Benson and Wisconsin I-94 19.0-mm mix). Many practitioners generally agree that the precision (test repeatability) of the Superpave™ Gyratory Compactor is much better than the Marshall hammer. A noteworthy observation was the Wisconsin STH-13 project where testing variation was 38.6 percent of total variation in air voids. The remaining variation was largely explained by changes in material passing the 2.36-mm sieve.

Based on these relative percentages of testing variation, it is recommended that more samples be collected and tested, rather than more subsample testing on each sample (for air voids). Greater variation is found with materials, production,

TABLE 4.11 ANOVA table for laboratory air voids

Source	Degrees of Freedom	Expected Mean Square
MPS	I-1	$\sigma^2_T + J\sigma^2_{MPS}$
Testing	I(J-1)	σ^2_T
Total	IJ – 1	-----

TABLE 4.12 Percentages of MPS and testing variation for air voids

Project	MPS, %	Testing, %	Test Method
Arizona			
I-10 Phoenix	87.0	13.0	Superpave
I-10 Benson	76.2	23.8	Superpave
USH-60 Wickenburg	83.4	16.6	Superpave
USH-60 Wittmann	93.6	6.4	Superpave
Florida			
I-10	92.4	7.6	Superpave
USH-231	82.2	17.8	Superpave
Kentucky			
I-64	72.6	27.4	Marshall
STH-18	78.8	21.2	Marshall
Ohio			
I-75 Dayton	75.6	24.4	Marshall
I-75 Findlay	72.1	27.9	Superpave
Wisconsin			
STH-13	61.4	38.6	Marshall
USH-12	91.6	8.4	Marshall
I-94 19.0mm	75.7	24.3	Superpave
I-94 12.5mm	83.1	16.9	Superpave

and sampling, and a greater portion of limited testing resources should be allocated to these three components of variation.

Performing a similar analysis on other tests, such as asphalt content and aggregate gradation, would require multiple-specimen testing similar to laboratory compaction of individual test specimens. This analysis was not performed because of limited study resources. It is recommended that similar analyses be conducted in future studies to determine the relative percentages of MPS and Testing variation for the other common construction tests to understand their contribution to total variation. Also, experiments can be specifically designed to isolate and measure materials, production, and sampling variation components. Using the results from these experiments, the testing program (i.e., selection of properties to measure, test methods to use, and number of samples to collect) can be made on the basis of how much variation there is in the results.

4.4.3 Density

Density data were analyzed with the Random Effects Model in a way similar to that used for the mix property data. Density lots were classified as days for QC purposes. The same hypothesis test was used to determine if variation between days is significant.

<u>Hypothesis #3</u>

H_O: Day-to-day variation is zero ($\sigma^2_{Day} = 0$).
H_A: Day-to-day variation is not zero ($\sigma^2_{Day} \neq 0$).

Table 4.13 provides a statistical summary for analysis of data from bulk specific gravity (G_{mb}) and density tests. Results from the G_{mb} test were included to provide a comparison with density tests, the density test result is two-dimensional having

TABLE 4.13 Summary analysis of density tests between days

Project	n	Gmb Significance	Density Significance
Arizona I-10 Phoenix	190	X	X
Arizona I-10 Benson	270	X	X
Arizona I-40 Topock	130	X	X
Arizona USH-60 Wickenburg	90	X	x
Arizona USH-60 Wittmann	100	X	X
Florida I-10 Lake City			
12.5-mm mix	211	x	X
9.5-mm mix	210	X	X
Florida USH-231 Panama	143	X	X
Kentucky I-64 Lexington	53		
Kentucky STH-18 Florence	30		
Minnesota I-90 Winona	143	X	X
Ohio I-75 Dayton	130	X	X
Ohio I-75 Findlay	90	X	X
Wisconsin STH-13 Plainville	175	x	x
Wisconsin USH-12 Baldwin	138	x	x
Wisconsin I-94 Milwaukee			
19.0-mm mix	634	X	X
12.5-mm mix	627	X	X

X denotes Highly significant (p-value < 0.01).
x denotes Significant (p-value < 0.10).

both G_{mb} and the maximum specific gravity (G_{mm}) in the test calculation.

This analysis concludes that nearly all states participating in the study had significant between-day variation for density. Typical project density data found mean levels changing by 1 to 2 percent between days, with the standard deviation fluctuating from 1.0 to 1.8 percent. Reviewing the Kentucky density data revealed relatively small sample sizes ranging from 1 to 5 cores per day, providing an insufficient sample size for evaluation of density data. The results of the *F*-test for these projects should be viewed with caution.

The data analysis provides insight for contractor QC testing. Contractors can expect the compaction process to have significant variation between days. Successive days of production could operate at a consistent mean level with a certain degree of variation, but a series of production days within a project will have days operating at different mean levels, producing significant between-day variation. Factors contributing to significant between-day variation are changes in mix properties, compaction methods, and environmental conditions. Day-to-day variation caused by changes in the mean level can be expected when weather conditions change. A day with constant external factors and little change in the compaction process can expect to yield less variation. A necessary component of establishing testing levels for contractor density testing is to perform testing daily and allocate available testing resources to this variation source.

A method for determining daily density testing frequency is found by applying principles of the average method. Contractors typically do not use QLA-type specification methods to control density, such as PWL. The PWL measure is multidimensional and combines the mean, standard deviation, and number of tests into one measure of quality. Rather, singledimensional estimates for both the mean and standard deviation, having high reliability, are needed to monitor and control density.

It is recommended that Equation 4.4 be used as a starting point to determine the number of daily density tests for quality control. There are four components to this equation that contractors can specify: (1) confidence limits of the average, (2) Z-statistic for the probability level, (3) estimated standard deviation, and (4) number of tests. As this equation implies, fewer tests are required when the confidence limits are increased, when the probability level is reduced, or when the standard deviation is small. Reasonable estimates for the standard deviation can be found from data collected in this study, at the start of paving on an individual project, or from a prior Wisconsin density study that obtained reliable estimates for the population standard deviation *(26)*. Equation 4.4 is defined as:

$$\text{C.L.} = z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \tag{4.4}$$

where

σ^2 = production variance;
C.L. = Confidence Limits for the average;
$z_{\alpha/2}$ = standardized normal statistic; and
n = number of tests in the moving average.

A practical application of how Equation 4.4 can be used to set QC testing levels is shown in Table 4.14. As a starting basis, the standard deviation of density found within the mat is typically 1.2 percent to 1.8 percent for 1 day's production, so an average value of 1.5 percent may be chosen. A reasonable probability level would be 95 percent (Z-statistic equal to 1.96). This table provides the relationship between number of tests, confidence limits for a 95-percent probability level, and a known σ_{Density} equal to 1.5 percent.

This table shows a decrease in the confidence limits as the number of tests is increased. The greatest rate of change occurs between 5 and 10 tests, where the interval is reduced from ±1.31 percent to ±0.93 percent, respectively. If 20 tests are taken, the interval drops to ±0.66 percent. However, smaller confidence limits require more tests.

Nuclear density gauges can produce more tests than cores because they require less testing time, and nuclear density test results are in real time, thereby allowing proactive modifications in rolling operations to achieve target density. Contractors have already included a nuclear density testing technician in their construction operations, and 10 to 20 nuclear density tests within a day will not increase construction costs. Core testing costs are substantially more than nuclear density testing. The Wisconsin density study found that core testing cost is about 5 times greater than nuclear testing cost for estimating an equivalent mean confidence interval *(26)*.

A rational way of obtaining unbiased density samples with a nuclear density gauge is defining the rolling zone as a stratified sublot. Density test sites would then be randomly chosen within each rolling zone. Unbiased estimates for the mean and standard deviation for a given day of production would then be obtained by combining the individual rolling zone tests into a single data set.

TABLE 4.14 Number of daily density QC Tests for 95-percent confidence limits and σ_{Density} = 1.5 percent

Number of Daily Tests	Confidence Limit (+/-) for Mean Density, measured as a % TMD
5	1.31
6	1.20
7	1.11
8	1.04
9	0.98
10	0.93
11	0.89
12	0.85
13	0.82
14	0.79
15	0.76
16	0.74
17	0.71
18	0.69
19	0.67
20	0.66

4.5 ACCEPTANCE TESTING LEVELS

A fundamental issue of this research was developing a rational method to determine acceptance testing levels. Earlier, the rational method was defined as the steps that should be taken when developing the testing specification. The minimum level of testing was defined as the minimum testing resources to allocate for a given project. Satisfactorily constructing an HMA overlay was defined as meeting the specifications with specified test properties, test methods, and compliance measures.

Three fundamental measures for acceptance testing are (1) mix properties, (2) density, and (3) smoothness. AASHTO currently recommends that agencies determine their own sublot and lot sizes for mixture properties (aggregate gradation, asphalt content, volumetric properties) and density *(1)*. Recommended testing levels for smoothness are 0.1 lane-mile sections for sublots and the entire project for lots *(1)*.

The work plan identified several interrelated, yet discrete, variables that an agency and contractor can evaluate when determining acceptance testing levels. Procedures were developed to evaluate specification limits, sample size, and the actual pay factor. A testable hypothesis was developed for distribution of samples within a lot. The following sections describe processes that should be performed to determine minimum testing levels for acceptance.

4.5.1 Specification Limits

Specification limits quantitatively define the allowable limits for typical HMA production processes and are considered the determinate between acceptable and unacceptable quality. If the limits span a narrow region about the target value, an appreciable amount of testing would be out-of-tolerance and more exhaustive testing would be necessary to control the process and avoid exceeding the limits. Wider specification limits make it easier to achieve design targets and allow greater variation in test observations.

An analysis was conducted to evaluate specification limits using variability found in actual production data. The standard deviations measured from the field projects using the QLA compliance measure were used to develop an expected production range containing 90 percent of the tests. A 90-percent range was chosen because many states have chosen an AQL value of PWL equal to 90. The production range was developed from allowable deviations in the mean, as shown by Equation 4.5:

$$\text{P.R.} = \overline{X} \pm z_{\alpha/2}(\sigma) \tag{4.5}$$

where

P.R. = Production Range;
$\overline{X}$ = mean;
$z_{\alpha/2}$ = standardized normal statistic; and
σ = standard deviation for production.

Laboratory air voids are illustrated for this analysis; however, the same method could be applied to other test properties, such as aggregate gradations and asphalt content. A comparison was made only using data from Arizona because the type of compliance measure used by each state can affect the contractor response to the specification limits. The Arizona specification, which used the QLA method to measure PWL, provided a pure measure of whether the contractor could meet the limits. The other five states specified compliance measures other than QLA. The QLA-type specification evaluated the distribution of tests between limits, as opposed to the other compliance measures, which used different statistical properties.

Table 4.15 provides the mean, standard deviation, and total number of tests from the projects, along with the target design value and specification limits. The 90-percent confidence limits (production range) were then compared with the range of the actual project specification limits. The coefficient of variation (*cv*) was calculated for these projects. The *cv* ranged from 13.2 to 22.8, indicating somewhat stable values for this statistic.

The comparisons within Arizona found all contractor production ranges could meet specification limits. However, the production range was off target by 0.5 percent on two projects, resulting in some material to be measured outside the specification limits.

This analysis exemplifies the need for agencies and contractors to consider development of specification limits accounting for the variation found in the production process. Specification limits should be evaluated on a state-by-state basis using variabilities from several contractors. A recommendation in *NCHRP Synthesis of Highway Practice 232* indicates that a specification writer should choose "typical" variability on which to establish the specification limits, perhaps slightly below the median variability value *(23)*. An appropriate specification range should be chosen that allows for slight variations in the process mean and variability to allow contractors to achieve PWL equal to 90 (100-percent payment). Wide specification limits will permit a greater range of test values and tolerate a correspondingly larger variability from construction.

A reasonable basis to determine if specification limits are correct is through collection of actual performance data. Ideally, specification limits should be based on both pavement performance and the ability of a contractor to meet the specification limits. Because it has been difficult to accurately translate the relationship between construction vari-

TABLE 4.15 Comparison of production range and specification limits

		Production						Specification				Production
		Mean	Std Dev	C.V.[a]	90% Production Range			Target	Specification Limits, %			greater than
Project	N	%	%	%	Lower	Upper	Range	%	Lower	Upper	Range	Spec Limits
Arizona												
I-10 Phoenix	212	4.6	1.05	22.8	2.9	6.3	3.5	5.1	3.1	6.6	3.5	No
I-10 Benson	136	5.1	1.05	20.6	3.4	6.8	3.5	5.0	3.0	6.5	3.5	No
I-40 Topock	52	5.3	0.84	15.7	4.0	6.7	2.8	5.0	3.0	6.5	3.5	No
USH-60 Wick.	36	4.8	0.64	13.2	3.8	5.9	2.1	5.0	3.0	6.5	3.5	No
USH-60 Witt.	40	4.4	0.92	20.6	2.9	5.9	3.0	4.9	2.9	6.4	3.5	No

[a]Coefficient of Variation is the standard deviation divided by the mean.

ability and performance, a reasonable alternative is to set specification limits using construction variability. As performance data are collected, they should be compared with the level of contractor variability used to construct the project and, possibly, the limits should be revised.

4.5.2 Distribution of Samples within a Lot

An important consideration during the development of an acceptance specification is the distribution of samples within a lot. It was learned earlier that between-day variation is significant for both plant mix properties and density, so this effect could affect the distribution. Either time-based or quantity-based sampling could be used to create the lot. Typical time-based lots are 1 production day, while quantity-based lots are some arbitrary tonnage. In either lot configuration, the acceptance plan must recognize how samples are distributed so that necessary statistical assumptions are satisfied. A key assumption of Statistical Quality Assurance (SQA) specifications is that the data are normally distributed.

An experiment was conducted to test the normality assumption on actual projects. Hot-mix sampling frequencies were increased from normal levels to 10 daily samples to create sufficient sample sizes to conduct the normality test. Maximum specified hot-mix sampling frequencies were typically four per day. Current density sampling rates were 10 per day, so additional sampling and testing were not necessary.

Data for the normality test were collected from three projects in the study: (1) Arizona I-10 (Benson), (2) Florida USH-231, and (3) Ohio I-75 (Findlay). The agency or contractor was requested to collect 10 hot-mix samples, within 1 day of plant production, using standard random-sampling procedures. The agency tested the hot-mix samples on the Arizona and Florida projects, while the Asphalt Institute tested the 10 samples on the Ohio project. Arizona and Ohio both specified 10 core samples per day, so additional core sampling was unnecessary. Florida specified 5 cores per 1,500 lane-meters; in many cases, more than 3,000 meters were paved within 1 day, allowing for 10 or more core samples to be collected.

A hypothesis test was conducted to determine if test data are normally distributed within days as follows:

Hypothesis #5

H_O: Test data are normally distributed within days.
H_A: Test data are not normally distributed within days.

Two statistical procedures were used to determine normality: the Anderson-Darling Normality Test and Ryan-Joiner Normality Test. A probability level of significance (*p*-value) was chosen at 0.10 for both tests because of the relatively small sample sizes. Results of the normality tests for the three projects are provided in Table 4.16. The power of the test to discriminate between normality and non-normality was diminished with a sample size of n equal to 10. It is recommended that, in future, normality tests should be done with larger sample sizes so as to increase the power of the test.

It was concluded that both hot-mix and density test data within 1 day's production are normally distributed, based on data from these three projects. There was evidence of non-normality with the asphalt content and maximum specific gravity (G_{mm}) tests on the Florida project. A review of the Florida data found a somewhat skewed distribution, as shown by the histograms in Figure 4.9. The inverse relationship of asphalt content and maximum specific gravity was apparent, because higher specific-gravity aggregates were displaced with lower specific-gravity asphalt binder, thereby reducing the overall specific gravity of the sample.

The limitation of the above procedure was that it only hypothesized a normal distribution within days of production. Hypothesis tests earlier concluded that mix properties and density have significant variation between days. Successive days of production could have no between-day variation, but a series of production days within a project have between-day variation. Given this fact, the question was whether between-day samples are normally distributed. This is a fundamental concern for quantity-based lots, where an arbitrary number of samples from different days are used to create a lot.

TABLE 4.16 Results of normality tests

Project	Aggregate Gradation									Asph.	Gmb	Gmm	Voids	VMA	VFA	Density[a]
	12.5mm	9.5mm	4.75mm	2.36mm	1.18mm	600um	300um	150um	75um							
Arizona I-10 Benson	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Florida USH-231	X	X	X	X	X	x	X	x	X	x	X		X	X	X	X
Ohio I-75 (Findlay)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

X denotes normality for both Anderson-Darling and Ryan-Joiner tests (p-value > 0.10).
x denotes normality for either Anderson-Darling or Ryan-Joiner tests (p-value > 0.10).
[a]Normality tests conducted for several days on each project.

Asphalt Content

Midpoint	Count	
5.10	1	*
5.15	1	*
5.20	0	
5.25	0	
5.30	1	*
5.35	3	***
5.40	3	***
5.45	0	
5.50	1	*

Max. Sp. Gr. (Gmm)

Midpoint	Count	
2.512	1	*
2.516	3	***
2.520	3	***
2.524	1	*
2.528	0	
2.532	0	
2.536	1	*
2.540	1	*

Figure 4.9. Histograms for asphalt content and G_{mm} (Florida USH 231 data).

A hypothesis test was conducted to determine if test data (i.e., aggregate gradation, asphalt content, volumetrics, and density) are normally distributed between days.

<u>Hypothesis #6</u>

H_O: Test data are normally distributed between days.
H_A: Test data are not normally distributed between days.

Only 10 hot-mix samples were available for a single day on the three projects because of extremely high levels of testing imposed on the respective laboratories. To overcome this obstacle, data from adjoining days, each having three or four samples, were combined to create 7- or 8-sample lots. A normality test was then conducted on these 7- and 8-sample lots.

The analysis found that about 5 percent of the reconfigured lots using adjoining days did not meet the conditions of a normal distribution. Given this small percentage, it was concluded that mix property data are normally distributed between days. This determination was important because many states are using quantity-based lots that combine material between days using discretional sizes. There will be lots, however, constructed across days that will not meet the normality condition. The power of the test to discriminate between normality and non-normality using sample sizes of n equal to 8 is diminished.

Combining some adjoining daily lots caused an increase in the standard deviation. Figure 4.10 provides an example where adjoining 4-sample lots were combined to create an 8-sample lot. The histograms and statistics indicate that the lots were somewhat different, and the distribution of combined samples did not resemble the typical "bell-shape" of a normal distribution. Results of the normality test for this 8-sample lot concluded with marginal evidence that the lot was normally distributed. Also, the standard deviation was much larger for the 8-sample lot than both single-day lots.

A hypothesis test was conducted to determine if density lots were normally distributed when tests between days were combined into a single lot. The Arizona and Ohio projects were used for the analysis because each state already sam-

Lot 4
Samples 1-4
September 24

Mean = 4.4
Std Dev = 0.41

Midpoint	Count	
4.0	2	**
4.4	1	*
4.8	1	*
5.2	0	
5.6	0	
6.0	0	
6.4	0	
6.8	0	

Lot 5
Samples 1-4
September 25

Mean = 5.8
Std Dev = 0.68

Midpoint	Count	
4.0	0	
4.4	0	
4.8	0	
5.2	1	*
5.6	1	*
6.0	1	*
6.4	0	
6.8	1	*

Lots 4 and 5
Combined

Mean = 5.1
Std Dev = 0.93

Midpoint	Count	
4.0	2	**
4.4	1	*
4.8	1	*
5.2	1	*
5.6	1	*
6.0	1	*
6.4	0	
6.8	1	*

Figure 4.10. Histograms for air voids (Arizona I-10 Benson data).

pled 10 cores per day. Between-day lots were created with the last 5 samples from a given day and the first 5 samples from the next successive day. This configuration allowed a normality test to be conducted having equivalent power to the within-day lots. The analysis found that about 8 percent of the lots configured using samples from adjoining days did not meet the conditions of a normal distribution. Given that 92 percent of the lots met the normality assumption, it was concluded that density test data were normally distributed when combined between days. Similar to the earlier tests, the power of the normality test was diminished with n equal to 10 sample sizes.

Figure 4.11 provides an example where 5 samples from adjoining production days are combined into one 10-sample lot. Histograms of the data from individual days and the combined lot indicate that days are somewhat different. Results of the normality test concluded that this single 10-sample lot was normally distributed. However, combining material from different days increased the overall standard deviation. This characteristic was found on many lots during this analysis.

The results of the hypothesis test for both within-day lots and between-day lots emphasized the importance of checking whether samples within the lot met the distributional assumptions required by the particular acceptance plan. It is recommended that an analysis of normality be performed on state-specific data to ensure that the distribution of the sample data satisfies the statistical requirements when combining material. Tests for normality between days should include larger sample sizes to ensure the test is made with high reliability.

Examples of combining material from different days of production into a single lot were given. These examples illustrate the inherent difficulties when combining material having different characteristics. Particularly, the combined samples may not meet the assumptions of a normal distribution, and the standard deviation may increase from estimates found in within-day lots. Many states elect to use quantity-based sampling using material from different days, and it is recommended that this type of analysis be performed so that both the agency and contractor are aware of the potential consequences.

4.5.3 Pay Factors

Pay factors are a fundamental concern during the development of testing levels. Goals of developing statistically based QA specifications are to understand the effects of sample size on the pay factor and to minimize (within practical limits) the risks to both the agency and contractor. The risk of a contractor having acceptable material rejected is computed at the AQL. The agency risk of accepting what is actually unacceptable material is computed at the RQL. A simplified procedure to measure these risks was developed using the data generated from computer simulation.

First, population means and standard deviations were calculated that would create a population distribution having theoretical PWL of PWL equal to 90 for acceptance at the AQL and PWL equal to 60 for acceptance at the RQL using existing specification limits. Table 4.17 provides the population mean and standard deviation for the Arizona DOT air voids specification, where the range between lower and upper specification limits was 3.5 percent. Individual specification limits are developed for each project based on the JMF target value, however, the range remains constant between lower and upper limits. Using computer simulation, sample sizes of n equal to 4, 6, 8, and 10 were then sampled independently from the population distribution to yield 5,000 estimates for the PWL that could be expected from

Lot 1 Samples 6-10 September 27 Mean = 94.0 Std Dev = 1.49			Lot 2 Samples 1-5 September 29 Mean = 91.8 Std Dev = 0.60			Lots 1 and 2 Combined Mean = 92.8 Std Dev = 1.56		
Midpoint	Count		Midpoint	Count		Midpoint	Count	
91.0	0		91.0	1	*	91.0	1	*
91.5	1	*	91.5	2	**	91.5	3	***
92.0	0		92.0	0		92.0	0	
92.5	0		92.5	2	**	92.5	2	**
93.0	0		93.0	0		93.0	0	
93.5	0		93.5	0		93.5	0	
94.0	2	**	94.0	0		94.0	2	**
94.5	0		94.5	0		94.5	0	
95.0	1	*	95.0	0		95.0	1	*
95.5	1	*	95.5	0		95.5	1	*

Figure 4.11. Histograms for density (Arizona I-10 Benson data).

TABLE 4.17 Population statistics for computing AQL and RQL

Quality Level	Population Mean[a]	Population Standard Deviation[a]
AQL (PWL=90)	4.75	1.0640
	4.20	0.9122
	3.60	0.4682
RQL (PWL=60)	4.75	2.0790
	4.20	2.0030
	3.60	1.6646

[a]Mean and standard deviation computed from LSL equal to 3.0 and USL equal to 6.5.

this distribution. PWL estimates were computed using the procedures found in the AASHTO *Quality Assurance Guide Specification (2)* and FHWA Demonstration Project 89 Software Manual *(14)*.

Using the 5,000 PWL estimates for each sample size, the mean and standard deviation of each pay factor were computed. Pay factors were calculated using the three pay factor equations presented earlier. The frequency of PWL and pay factors at intervals of PWL equal to 10 and pay factors equal to 10, respectively, were counted and an expected probability was calculated. For example, the frequency of PWL between PWL equal to 40.00 and PWL equal to 49.99 were counted, between PWL equal to 50.00 and PWL equal to 59.99, between PWL equal to 60.00 and PWL equal to 69.99 and so forth. This type of analysis allowed an efficient way of comparing the probabilities of PWL values that would be expected for each sample size. Appendix C of this report provides the computer program with summary output for this procedure.

4.5.3.1 Acceptance at AQL

Figure 4.12 shows the plot of sample size against the risks or probability of rejecting the contractor's work for the moderate pay factor with pay factor less than 85 (PWL less than 60). Plots are provided for the three combinations of the population mean and standard deviation that will produce a theoretical PWL equal to 90. This plot suggests that the risk varies for different combinations of the mean and standard deviation, and the risk decreases as the sample size increases. For example, a contractor producing any combination of the population mean and standard deviation at n equal to 4 has an approximate 2.3-percent chance of having the work rejected, while an approximate 0.3-percent chance exists at n equal to 8. The largest rate reduction in risk occurred between sample sizes of n equal to 4 and n equal to 6, and the rate of increase was much less for sample sizes from n equal to 6 to 8.

The reduction in risk is explained by less variation in estimates of the sample mean and standard deviation with larger

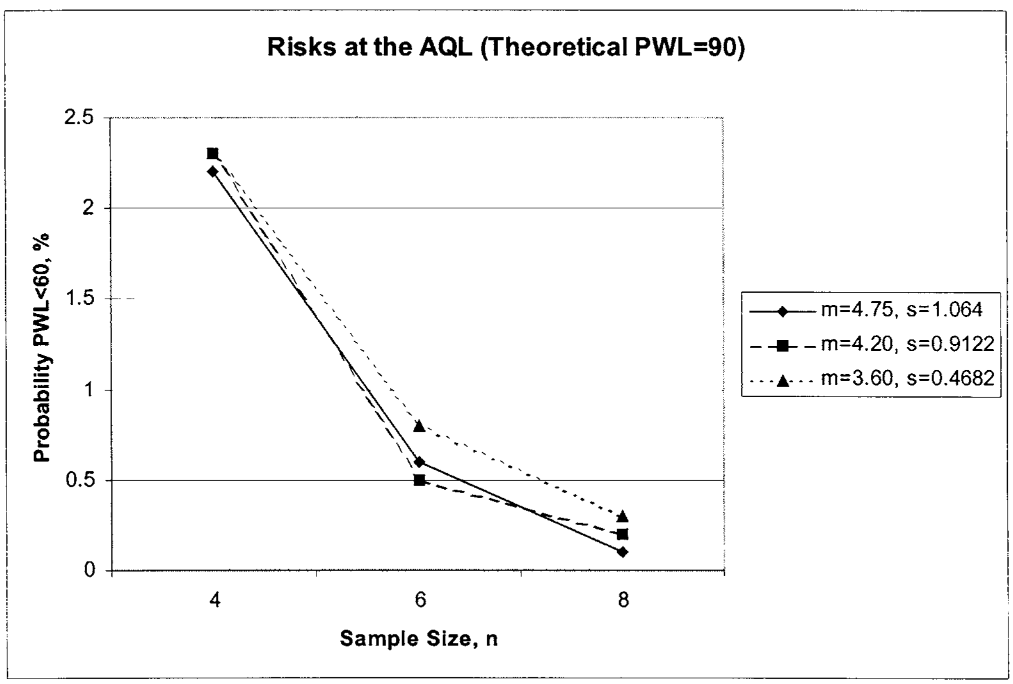


Figure 4.12. Contractor risk for moderate pay factor at the AQL.

sample sizes. When small sample sizes are selected from a population distribution, the estimate for the sample mean and standard deviation will have greater variation. Wide dispersion of these estimates produces greatly varying estimates for the PWL and pay factor. Thus, there are increased chances, or risks, for estimating percentages of material outside specification limits and rejecting what is acceptable work with smaller sample sizes. This type of analysis provides real-world implications when the contractor's process capability is equivalent to PWL equal to 90, and the resulting pay factors that will be produced. The agency and contractor must decide which level of risk they are willing to tolerate.

4.5.3.2 Acceptance at RQL

Of great importance to the agency is how the pay factor will perform at the RQL. The agency wants to reject work below the RQL and make appropriate pay adjustments for work that is accepted slightly above the RQL. However, there are instances where the agency will make full or bonus payment for defective work that is PWL less than 60. Figure 4.13 shows the plot of sample size against the probability of accepting the contractors work at a pay factor greater than or equal to 100 (PWL greater than or equal to 90) when the true population is PWL equal to 60. Plots are provided for the three combinations of the population mean and standard deviation that will produce a theoretical PWL equal to 60. Similar to the AQL analysis, the plot suggests that larger lot sizes will result in a reduction in risk and greater ability to discriminate correct payment at the RQL and that the population mean and standard deviation have little effect on risk levels.

4.5.4 Decision Analysis

The conclusions provided by the hypotheses tests and the analysis of both specification limits and pay factors offer several issues to consider during development of testing levels. Table 4.18 summarizes the hypotheses test conclusions as they pertain to acceptance testing.

It is desired to know what sample size, and corresponding lot size, to specify for acceptance. Clearly, it is to the advantage of both the agency and contractor to reduce the amount of risk with larger sample sizes. The effects of risk are reduced with a greater leniency in the pay factor given that the potential effect on life-cycle cost effects are reduced. The agency and contractor may want to consider choosing a larger sample size when operating under the severe pay equation.

This type of decision is important to consider when defining an appropriate pay factor for the severity of the project. The lenient pay equation may be more appropriate when the effect of construction quality has a lessened effect on the life-cycle costs. For example, a temporary pavement or a highway with low AADT could be classified with lenient pay factors. Severe pay equations would be assigned to projects having higher life-cycle costs, such as an interstate highway having higher traffic levels AADT. Thus, when determining testing levels, small or large sample sizes could be specified

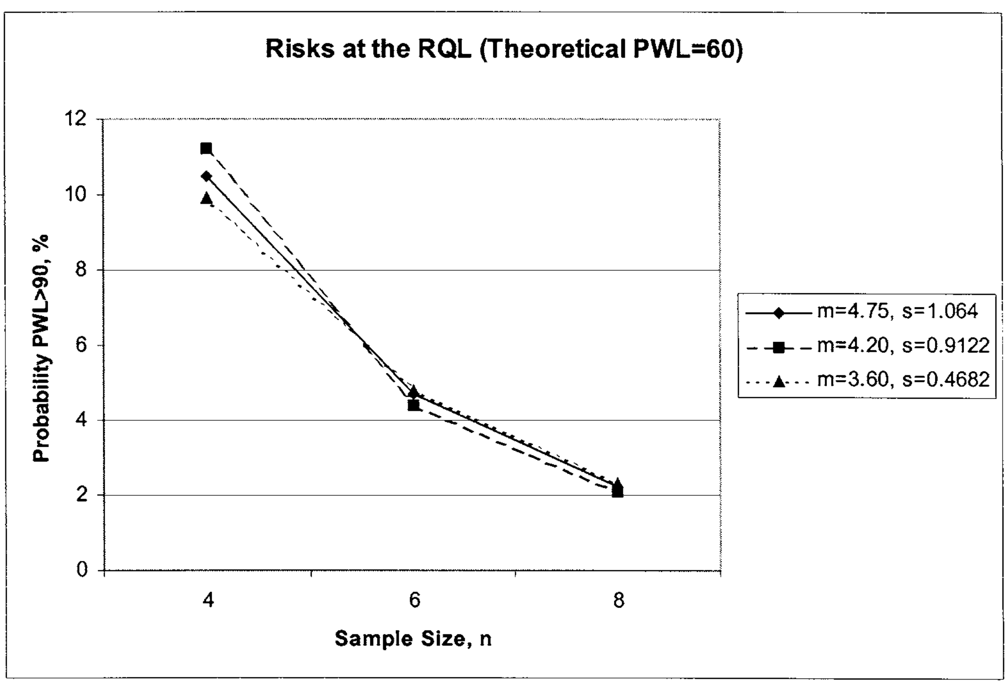


Figure 4.13. Agency risks for moderate pay factor at the RQL.

TABLE 4.18 Summary of hypotheses tests for acceptance testing

Hypothesis Number	Conclusion
5	Using data in this study, strong evidence that both hot-mix and density test data, within 1 day's production, are normally distributed.
6	Mix property data are normally distributed between days. Combining some adjoining daily lots caused an increase in the standard deviation. Density data are normally distributed between days. Combining adjoining n equal to 5 samples from different days to create a sample size of n equal to 10 produced an increase in the standard deviation.

for less critical projects, while larger sample sizes are needed for more critical projects. It is recommended that the risk between small and large sample sizes be evaluated during this decision-making process.

The method for determining sampling frequencies provided earlier may require more samples per lot for the critical projects. A rational acceptance specification would be established using the prior sampling methodology and the sample size that tolerates an acceptable risk level, as shown in Table 4.19.

The sublot size would correspond to the sampling frequency, and the lot size would correspond to the risk level that an agency and contractor are willing to tolerate for a particular project (based on criticality). The sublot size can be adjusted to improve efficiency in field sampling, and the sublot size will have no effect on the ability to achieve unbiased estimates for the population mean and standard deviation. For example, density sublots on high criticality projects could be specified as 1 per 200 tons or 2 per 400 tons.

4.6 VERIFICATION TESTING

A fundamental decision that must be made during development of any QA program is whether to use agency or contractor tests for acceptance. If the contractor QC test results are used for acceptance decisions, then the required amount of agency testing can be reduced to the amount necessary to validate the contractor's results *(1)*. The agency typically verifies the reliability of the contractor tests through periodic testing of verification samples. For purposes of this research, verification is defined as the procedure where a comparison is made between the agency or contractor tests using either split samples or independent samples to determine whether the contractor tests are satisfactory for acceptance. A commonly specified procedure is one agency verification test for each lot of contractor QC acceptance tests.

TABLE 4.19 Acceptance specification for project criticality

Project Criticality	Sublot Size	Lot Size
	Mix Properties	
Low	1 per 1,000 tons	4 per 4,000 tons
High	1 per 750 tons	8 per 6,000 tons
	Density	
Low	1 per 400 tons	5 per 2,000 tons
High	1 per 200 tons	10 per 2,000 tons

The AASHTO *Implementation Manual for Quality Assurance* provides two procedures for making statistical comparisons between agency and contractor tests *(1)*. AASHTO procedures (Appendixes F and G in the AASHTO manual) provide means to compare test results from the contractor and agency, but in different ways. Appendix F, "Comparison of Quality Control and Acceptance Tests," provides established statistical methods to compare test results with the use of the *F*-test and *t*-test. The *F*-test is used to compare variances, while the *t*-test is used to compare means. This procedure is largely concerned with determining whether contractor QC and agency acceptance tests come from the same population. However, a drawback of this procedure is that it requires more than one agency test to determine whether the contractor tests are similar to those of the agency. With limited agency resources, this procedure lacks the ability to reliably perform a single agency verification test for several contractor tests within a lot.

Appendix G of the AASHTO *Implementation Manual for Quality Assurance*, "Procedure for Evaluating Quality Control Test Results of Aggregate Gradations, Asphalt Mixes, and Portland, Cement Concrete," describes a process for performing verification testing. A series of equations are used concurrent with "If-Then" logic statements to determine whether contractor and agency tests are similar or dissimilar. A quantitative comparison is made between a computed interval of tests using the mean and range of contractor tests and determining whether the agency test lies within this interval. If the agency sample lies between the lower and upper limits of the interval, the QC samples are considered similar to the agency acceptance sample. A weakness of this procedure is the use of the range. The range is the least stable of conventional summary statistics (e.g., mean, standard deviation, and median), particularly with small sample sizes, and its use for comparing a small group of contractor and agency tests can produce widely varying conclusions.

A fundamental issue when determining verification testing levels is whether to use split samples or independent samples for comparing contractor and agency tests. Split-sampling was used in the work plan to provide a more accurate comparison between test results. No explicit definition of split-sampling was given in the AASHTO QA publications or other QA literature. For purposes of this research, it is when a single sample is taken from a predetermined time, tonnage, and location of material, and then split into equal sizes for testing by different laboratories. Split-sampling eliminates the effects of materials, production, and sampling variation caused by sampling at different locations within a truckbox or mat and by sampling at different times during production. Use of split-sampling makes the comparison between laboratory tests more precise. Successive samples taken during production will be subjected to different degrees of materials, production, and sampling variation. These components of variation are eliminated during the split-sample comparison of laboratories.

The definition of independent-sampling for this research is when laboratories independently collect samples at a different time, tonnage, and/or location. Unlike split-sampling, independent-sampling does not eliminate the variation components of materials, production, and sampling. Independent-sampling will introduce these variation components to some degree and comparison of laboratories will be obscured. Although a rational basis exists for choosing the split-sampling method, some states specify independent-sampling. Independent-sampling allows the contractor and agency to conduct the sampling process independent of each other.

The state highway agency must decide what it wants to verify. If the intent is to obtain an independent verification of the contractor's overall process, including material, production, sampling, and testing, then independent-sampling is recommended. If the intent is to verify simply the contractor's testing process, then split-sampling is the most efficient approach. If split-samples are used, then either their sampling and splitting must be observed by a state highway agency inspector, or the agency must be willing to trust the pedigree (i.e., trust that they really were obtained from the project and under proper random sampling procedures and so forth) of the split samples provided by the contractor to the agency. Thus, a necessary component when using split-sampling is having some degree of trust, cooperation, and partnership between the agency and contractor.

There are different requirements for making comparisons between laboratories when using split samples or independent samples. Table 4.20 identifies the statistical requirements when making a comparison between split samples and independent samples. Split-sampling offers a minor advantage: The normality assumption must be met for both contractor and agency data sets with independent-sampling, while this same assumption must be met for only the single split-sample data set. Independent-sampling requires an additional degree of independence, where sampling from individual laboratories must be conducted independently. The assumption of equal variances is not required with split-sampling, while this assumption is required for independent samples, but an approximate test procedure can be applied to relax this requirement.

One of the most important attributes of this table is the number of tests needed to compare agency and contractor

TABLE 4.20 Statistical requirements for split samples and independent samples

Attribute	Split Samples	Independent Samples
Normality	• Difference between data sets (e.g., one data set of differences)	• Both contractor and agency data sets
Independence	• Random sampling	• Random sampling • Samples must be collected independently (no split-sampling allowed unless by random chance).
Variances	No requirement	Equal variances for both labs
Number of Tests	$n = \sigma^2_D \dfrac{(Z_{\alpha/2}+Z_\beta)^2}{(\mu_{D,H_A}-\mu_{D,H_O})^2}$ where, n = number of pairs; σ^2_D= variance between pairs; $Z_{\alpha/2}$ = confidence level; Z_β = probability of detection; $\mu_{D, HO}$ = mean difference at null hypothesis; and $\mu_{D, HA}$ = mean difference at alternative hypothesis.	$n = 2\sigma^2 \dfrac{(Z_{\alpha/2}+Z_\beta)^2}{(\mu_{H_A}-\mu_{H_O})^2}$ where, n = number of verification tests; σ^2 = pooled variance between tests; $Z_{\alpha/2}$ = confidence level; Z_β = probability of detection; μ_{HO} = mean at null hypothesis; and μ_{HA} = mean at alternative hypothesis.
Measure of Variation	σ^2_D= Var (X_A - X_C) = VarX_A + VarX_C - 2Cov(X_A ,X_C) where, Var = Variance Cov = Covariance X_A = Agency data set X_C = Contractor data set	$2\sigma^2$ = Var (X_A - X_C) = Var X_A + Var X_C (assuming equal variances) where, Var = Variance X_A = Agency data set X_C = Contractor data set

tests. Independent-sampling will simply require an equal or greater number of tests than split-sampling. The number of tests for either split-sampling or independent-sampling is largely determined by the variance. When there is no correlation between split-samples, the variance expression for split samples (σ^2_D) and independent samples ($2\sigma^2$) is equal (σ^2_D equal to $2\sigma^2$). The advantage of split-sampling is realized when there is a positive correlation between split samples, which makes σ^2_D less than $2\sigma^2$. If the correlation is small, there may be an advantage to independent-sampling because of more degrees of freedom in the test. However, there is a large positive correlation found with split-sample data, and σ^2_D less than $2\sigma^2$, inferring that split-sample verification will require fewer tests than independent-sampling for equivalent parameters of interest.

A statistical procedure was developed for making a verification test comparison between split-sample test results using variation found in actual field data and the true difference between the mean of contractor and agency tests. Using the mean difference for verification testing with a testing tolerance table is a common practice in many QA programs, particularly for states using contractor data for acceptance. Testing tolerance tables are straightforward and do not require elaborate statistical calculations during construction. The differences between tests are considered acceptable if they do not exceed the tolerances set forth in a developed table. Table 4.21 provides QA testing tolerances from four states in the study. Arizona and Florida do not have testing tolerances because agency acceptance testing and contractor QC testing are performed independently.

A procedure was developed that calculates the testing tolerance using the standard deviation between laboratory tests from historical data (σ_D). There are large amounts of split-sample test data available to estimate σ_D. States should begin organizing these data to provide population estimates for the standard deviation of differences between laboratories on individual projects. A primary purpose of collecting 24 split-sample tests on the field projects was to estimate σ_D. Although there is variation among σ_D within states, a central or median value could be selected. Appendix G of the contractor's final report provides the summary of σ_D values for all projects in the research.

The number of verification tests for a given lot is a function of the following: (1) variation between tests, (2) difference between means, (3) probability that the mean difference is contained within a defined acceptable region, and (4) probability that the mean difference is accepted outside the defined acceptable region. It is recommended that agencies use this interrelationship when setting split-sample verification testing levels. Equation 4.6 enables the number of verification tests to be calculated. This equation is modified from the equations given in Table 4.20 (Column 2) where *d* is substituted in the denominator. A single test, or the mean of a group of tests, is used when making a comparison. Equation 4.6 is given below:

$$n = \sigma_D^2 \frac{(Z_{\alpha/2} + Z_\beta)^2}{d^2} \tag{4.6}$$

where

n = number of tests;
σ^2_D = variance between pairs (std. dev.)^2;
$Z_{\alpha/2}$ = probability level of acceptable differences (95 percent or higher suggested, $Z_{\alpha/2} = 1.96$);
Z_β = probability level of acceptance in rejectable region (80 percent or higher suggested, $Z_\beta = 0.842$); and
d = difference between means.

Table 4.22 provides a practical application of the developed procedure relating the number of verification tests and the testing tolerances using a known standard deviation. Suppose an agency verifies 1 of the contractor's 4 acceptance tests from a given lot for asphalt content (AC). Using Equation 4.6, there would be a 95-percent acceptance region for mean differences and an 80-percent chance of detecting a true mean difference of 0.6-percent AC between tests. If the agency and contractor had a difference of 0.3-percent AC, the probability of detecting a true difference would only be about 30 percent. The ability to discriminate true differences is diminished as the tests move closer together. Thus, a strong statement can only be made for true differences when the tests are further apart.

Another option would have the agency test all 4 split-samples in the lot. Then, if the agency and contractor had a

TABLE 4.21 Testing tolerances from state specifications

State	Notes	Aggregate Gradation						AC, %	Gmb, sp. gr.	Gmm, sp. gr.	Voids, %	VMA, %
		12.5mm, %	9.5mm, %	4.75mm, %	2.36mm, %	600um, %	75um, %					
Kentucky	Same equipment	-	-	-	-	-	-	0.5	-	-	1.0	1.0
	Different equipment	-	-	-	-	-	-		-	-	1.5	1.5
Minnesota	AC extraction	6.0	6.0	5.0	4.0	-	2.0	0.5	0.030	0.019	-	-
	AC ignition oven	-	-	-	-	-	-	0.2	-	-	-	-
Ohio	AC extraction	-	-	4.0	-	-	-	0.3	-	-	-	-
Wisconsin	-	6.0	6.0	5.0	4.0	3.5	2.0	0.5	0.030	0.020	-	-

TABLE 4.22 Number of verification tests and testing tolerances (Florida I-10 data)

Test	Standard Deviation Between Tests	Testing Tolerances for Select Number of Tests			
		n=1	n=2	n=3	n=4
12.5mm, %	4.50	12.6	8.9	7.3	6.3
9.5mm, %	4.29	12.0	8.5	6.9	6.0
4.75mm, %	2.52	7.1	5.0	4.1	3.5
2.36mm, %	1.07	3.0	2.1	1.7	1.5
600μm,%	0.64	1.8	1.3	1.0	0.9
75μm, %	0.39	1.1	0.8	0.6	0.5
AC, %	0.21	0.60	0.42	0.35	0.30
Gmb, sp.gr.	0.008	0.023	0.016	0.013	0.011
Gmm, sp.gr.	0.007	0.020	0.014	0.012	0.010
Voids, %	0.41	1.1	0.8	0.7	0.6

mean difference of 0.3-percent AC, the probability of detecting a true mean difference would be 80 percent, much higher than the single-test probability (30 percent). As the number of verification tests is increased for the comparison, the true mean difference between tests is decreased.

This statistically based approach applies the concept of "power." Power is the level of probability necessary to detect true differences between the contractor and agency test results. Power is computed only from population values and generally does not apply when the standard deviation is unknown. Sampling data are not used in power computations. A common question is what is an acceptable value to have for power? Based on experience, values of 0.80, 0.90, or 0.95 are preferred so there is reasonable assurance that a true difference is detected *(33)*. An example of a power curve that illustrates the ability of the agency to detect true differences in the contractor tests results during split-sample verification is provided in Figure 4.14. In this example σ_D is equal to 0.34-percent AC.

If the agency wants to achieve a high power value, or a reasonable assurance of detecting a true difference between contractor and agency tests, the difference must be 1.0 percent *or higher*. When a difference between contractor and agency tests is at a typical tolerance limit of 0.5 percent, the ability to detect a true difference is only 30 percent. Thus, the ability of the agency to verify contractor test results is greatly diminished when considering the power of the test.

The method of increasing the number of tests can provide a more accurate comparison of contractor and agency tests. Figure 4.15 shows the power curve when the mean difference of 4 split-samples is compared between the contractor and agency tests (σ_D equal to 0.34-percent AC). The ability to detect differences is greatly increased. The agency is now able to detect a true difference of 0.5 percent between asphalt content tests at a power of 70 to 80 percent. The power is diminished for smaller differences, such as 0.1- to 0.3-percent AC.

An interactive approach using project-specific data was analyzed, where a sample standard deviation (s_D) was used to estimate the population standard deviation of differences (σ_D). It was found that a project-specific estimate for the variability of

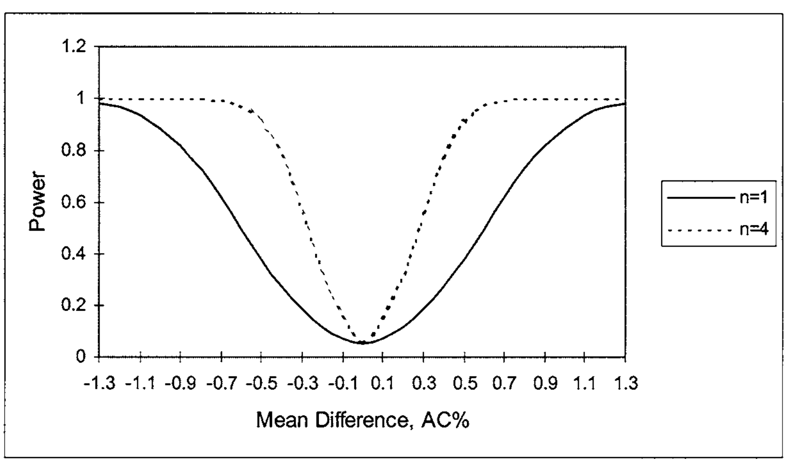
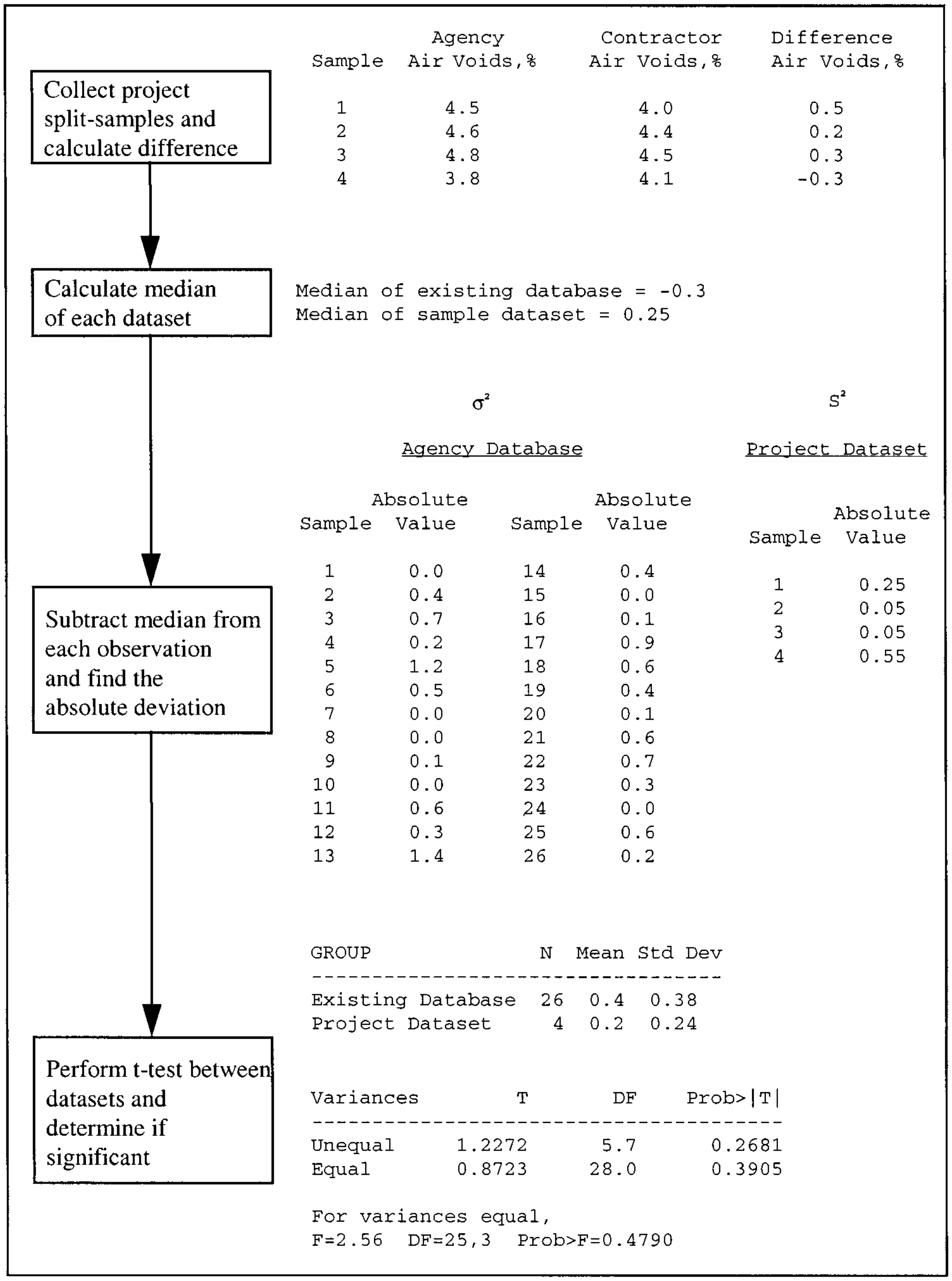


Figure 4.14. Power curve for mean difference of tests (σ_d = 0.34%).



Sample	Agency Air Voids,%	Contractor Air Voids,%	Difference Air Voids,%
1	4.5	4.0	0.5
2	4.6	4.4	0.2
3	4.8	4.5	0.3
4	3.8	4.1	-0.3

Median of existing database = -0.3
Median of sample dataset = 0.25

σ^2 Agency Database

Sample	Absolute Value	Sample	Absolute Value
1	0.0	14	0.4
2	0.4	15	0.0
3	0.7	16	0.1
4	0.2	17	0.9
5	1.2	18	0.6
6	0.5	19	0.4
7	0.0	20	0.1
8	0.0	21	0.6
9	0.1	22	0.7
10	0.0	23	0.3
11	0.6	24	0.0
12	0.3	25	0.6
13	1.4	26	0.2

s^2 Project Dataset

Sample	Absolute Value
1	0.25
2	0.05
3	0.05
4	0.55

GROUP	N	Mean	Std Dev
Existing Database	26	0.4	0.38
Project Dataset	4	0.2	0.24

Variances	T	DF	Prob>|T|
Unequal	1.2272	5.7	0.2681
Equal	0.8723	28.0	0.3905

For variances equal,
F=2.56 DF=25,3 Prob>F=0.4790

Figure 4.15. Example of variance test.

differences between laboratories (s_D) is unreliable for small sample sizes, and the normality assumption would most likely not be met. Because of these facts, it would be difficult to estimate the population of differences between laboratory tests using a small group of samples at the start of production. Developing a population estimate for σ_D would require a substantial testing effort at the beginning of production and bring into question the idea of periodic verification testing, which would essentially become continuous verification testing, to determine this estimate. Continuous verification testing to develop this estimate may even continue until the conclusion of the project. If there is uncertainty in σ_D, it is recommended that an agency calculate "n" and "d" for a range of σ_D values. (Refer to Appendix J of the contractor's final report for further analysis concerning the relationship between laboratories).

An additional reason for not adopting an interactive approach to set tolerance limits was potential problems with implementation. Knowledge of how the procedure will develop testing tolerances in practice could create the remote possibility of laboratories increasing variation between tests at project start-up to develop excessively wide tolerance limits. These wide tolerance limits would be incapable of detecting any differences between laboratory tests during continued production.

A process was developed to evaluate the project-specific s_D value during implementation. This procedure simply compares the sample s_D value with the population σ_D value. A problem successfully overcome during the analysis was ensuring that small sample sizes used to develop the s_D estimate met the assumptions of a normal distribution. Rather than using the *F*-test to compare variances with lack of assurances for a normal distribution, an alternative method known as Levene's Method was used. Levene's Method was chosen because it is quite robust to departures from normality *(19)*. The only assumptions needed are that the two samples are randomly chosen and that they are independent of each other. An example of this procedure is provided in Figure 4.15.

4.7 SUMMARY

This chapter provided important considerations for using contractor or agency data for acceptance. A sampling methodology was given that determines the sampling frequency using production rates and time to conduct tests. Methods were developed for determining testing levels for contractor QC and agency acceptance. Finally, a method was given for verifying the contractor and agency tests.

Integral components of developing QC and acceptance testing levels were conclusions drawn from the testable hypotheses. The six hypotheses were tested from areas of the testing specification that included test properties, variation, specification limits, sample size, and pay factors. Conclusions from these testable hypotheses are provided in Table 4.23.

TABLE 4.23 Conclusions from testable hypotheses

Index of Hypotheses	Testing Specification Area	Conclusion
1	Test Properties	Coldfeed aggregate gradation samples have different means than hot-mix samples.
2	Test Properties	Coldfeed aggregate gradation samples have different variation than hot-mix samples.
3	Variation	There is significant day-to-day variation for both hot-mix properties and density.
4	Variation	Testing variation is a small percentage of total project variation for air voids.
5	Distribution of samples	Both hot-mix and density test data within 1 day's production are normally distributed.
6	Distribution of samples	Both hot-mix and density test data are normally distributed when tests are combined from different days (based on data from the three projects in this study).

CHAPTER 5

PROCESS FOR DETERMINING INSPECTION LEVELS

5.1 INTRODUCTION

This chapter describes the process to determine minimum inspection levels for HMA overlay construction. The following sections are found in this chapter: (1) identifying inspection tasks, (2) determining important inspection tasks, and (3) evaluating methods to establish minimum inspection levels.

The process, shown in Figure 5.1, begins by identifying inspection tasks found during HMA production defined by process, location, activity, and task. Next, important inspection tasks identified using information gained in the literature review are combined with observed field data collected from projects in this study. Then, methods are given to set minimum inspection levels through analysis of current HMA practices and applying nationally accepted QC methods.

5.2 IDENTIFYING INSPECTION TASKS

The first step toward minimizing inspection levels is to understand all relevant inspections in HMA construction. Information previously discussed in the literature review of HMA trade publications was consulted, along with contractor input to identify inspection locations, activities, and tasks. HMA construction inspection was structured into a 4-division hierarchy: (1) process, (2) location, (3) activity, and (4) task. This hierarchical structure provided a simple, organized format for data analysis.

The construction processes have four parts: (1) raw materials, (2) plant mixing, (3) mix transport, and (4) laydown. There are 12 physical locations that fit within those four processes, and these are locations during construction where inspection activities and tasks are possible, as shown in Figure 5.2. These locations are numbered and include the following: (1) aggregate stockpiles, (2) asphalt binder, (3) aggregate blending, (4) asphalt binder delivery, (5) mix plant, (6) dust collector, (7) mix storage, (8) load-out, (9) weigh scale, (10) base condition, (11) paving, and (12) rolling. Each of these locations has specific inspection activities that can be identified, as well as specific inspection tasks within those activities. Figure 5.3 provides an example of each of the four divisions of HMA inspection with examples of an inspection process, location, activity, and task.

The data collection format was developed with the assistance of two contractors while visiting a mix plant and observing laydown operations. This was done to ensure that the data collection was comprehensive and in a usable format. The contractors were also helpful in providing practical insight into field practices, so that data collection would be complete and provide useful information.

The information gained from literature review and contractor input are summarized in Appendix B of this report. A list is included of the identified inspection tasks and a description of each as they occur by process. The descriptions are necessary to clarify the inspection task characteristics as they pertain to HMA construction. Appendix B is also provided as a reference to be used throughout the process to determine minimum inspection levels. For the descriptions to fulfill this purpose, they must be thorough and address every aspect of the inspection task. This was done using a system of analysis known as "process evaluation." Process evaluation is a systematic way of defining activities and tasks within an operation by asking basic questions: (1) where, (2) what, (3) who, (4) when, (5) how, and (6) why *(13)*. The operation in this case is HMA construction, and inspection activities and tasks are described in the table as they answer these questions. By answering these questions, a complete description of the inspection tasks was possible.

5.3 DETERMINING IMPORTANT INSPECTION TASKS

It is necessary to find those inspection tasks believed to be most important in producing quality HMA, using the inspection task descriptions as a starting point. Important inspections are determined in this section using collected field data and information from the literature.

5.3.1 Collection of Field Data

To provide a comprehensive view of current practices, it was necessary to include both observed agency and contractor field data of both mix-plant and laydown inspection. A member of the research team was present at 13 projects to collect and document field inspections and practices for both the agency and contractor. This information was synthesized

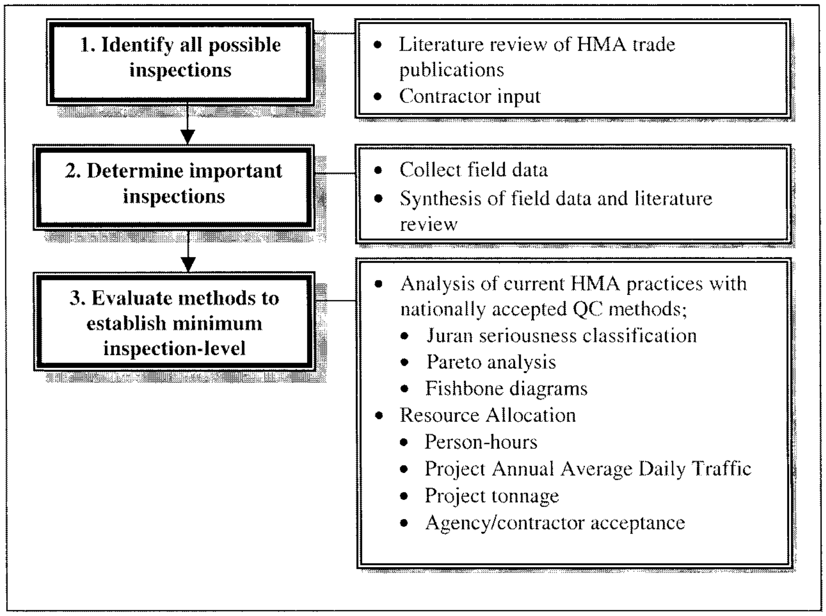


Figure 5.1. Process for determining minimum inspection levels.

by reviewing the contractor's QC plan, interviews with agency and contractor personnel at each project, and observed practices. Tables were developed to summarize agency and contractor inspection separately so that separate recommendations could be made.

5.3.2 Synthesis of Field Data and Literature Review

The information described in the literature review and collected during field data collection provides a summary of those inspection tasks thought to be important to producing an HMA quality overlay by agencies and contractors. Virtually every aspect of production can have an effect on the quality of the final product. However, there are limited resources to conduct the inspection tasks, so it is necessary to understand where the maximum benefit can be obtained. HMA trade publications and observed field practices provide valuable insight on where resources should be focused.

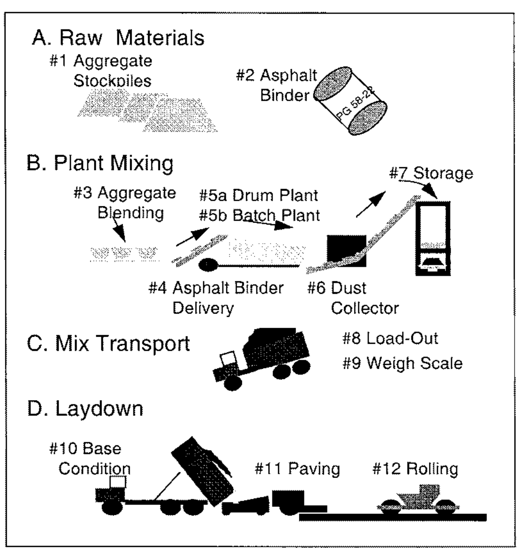


Figure 5.2. Inspection of HMA construction processes.

Inspection task importance is based on whether it is thought that the task is necessary for a quality pavement and if it was frequently found among the literature review and the collected field data. It is thought that if these tasks are performed during construction, agencies and contractors will have the ability to monitor and assess the quality of the mix or compacted pavement, thereby allowing corrective action, if necessary.

Each of the HMA trade publications was studied to identify inspection tasks. If an inspection task was found in a particular source, it was recorded in a table. Inspection tasks were also recorded in the table if they were observed during field data collection. This was done to provide an overview of each data source in common terms. Table 5.1 shows an example of how inspection tasks were recorded.

If a handbook discussed the importance of maintaining separation between different aggregate stockpiles, an "x" was marked in columns 5 and 6 of Table 5.1. Contractors were also observed inspecting their stockpiles for appropriate separation during field data collection. This is recorded in column 9. A compiled list of the inspection tasks based on these sources can be found in Appendix L of the research team's final report.

Table 5.1 also provides an example of the method used to determine importance. It was assumed for analysis purposes that inspection tasks are important if found in at least three

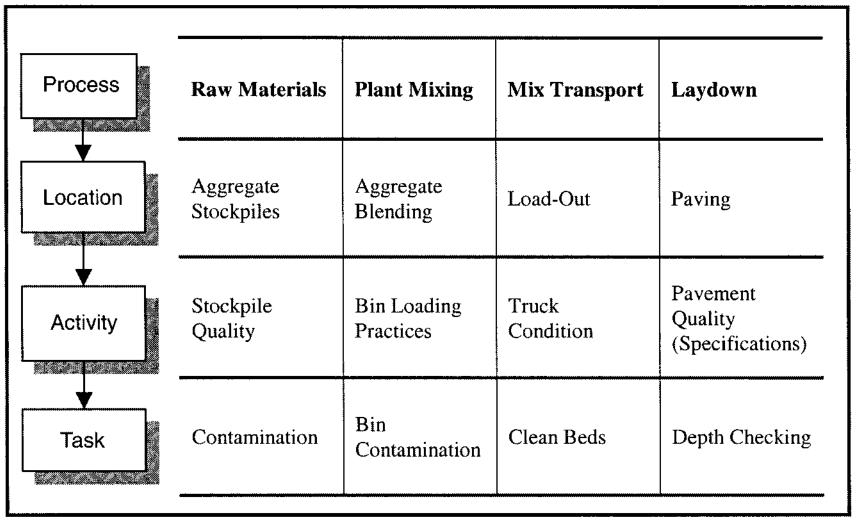


Figure 5.3. Four divisions of HMA construction inspection.

of the four possible source areas (AASHTO et al., NAPA, observed agency field data, and observed contractor field data). Tasks are also important if given a NAPA impact rating of five or greater. For example, "Stockpile Separation" is an important inspection task because it was found in three of the four source areas. "Fines Return Location" is an important inspection task because it was given a NAPA impact rating of eight (which is greater than the cutoff value of five). Each of the inspection tasks was evaluated using this method. Table 5.2 provides a listing of those inspection tasks that are considered important. This list will be used as a reference when making recommendations for minimum inspection levels.

5.4 EVALUATING METHODS TO ESTABLISH MINIMUM INSPECTION LEVELS

The important inspection tasks list provides an organized knowledge base for use in determining minimum inspection levels. It is now possible to evaluate established QC methods and apply them to this list and HMA field practices. Current practices are evaluated relative to project characteristics. This information can be used to show agencies and contractors where there is opportunity for increased efficiency. Nationally accepted QC methods were used: (1) Juran seriousness classification, (2) Pareto analysis, and (3) fishbone diagrams. A fourth method used was resource allocation using data collected from the projects in this study.

5.4.1 Juran Seriousness Classification

Juran seriousness classification is a method of identifying those inspection tasks that have the greatest effect on final pavement quality using the expertise of those closely involved in HMA construction. The seriousness classification developed by J.M. Juran provides a method of emphasizing inspection

TABLE 5.1 Method for determining important inspection tasks

Hierarchical Inspection Divisions					NAPA		Observed Field Data		
Process	Location	Activity	Task	AASHTO, FAA, FHWA, NAPA, U.S. Army Corps of Engineers Handbooks	Handbooks	Impact Rating	Agency	Contractor	Total [a]
Raw Materials	Aggregate Stockpiles	Maintain Stockpile Quality	**Pile Separation**	x	x	-	-	X	3
Plant Mixing	Dust Collector	Equipment Conditions	**Fines Return Location**	x	x	8	-	-	2 [b]

[a] Sum of columns [(5)+(6)+(8)+(9)].
[b] These inspection tasks have a NAPA impact rating greater than or equal to 5.

TABLE 5.2 Important inspection tasks for achieving HMA construction quality

Process	Location	Activity	Task
Raw Materials	Aggregate Stockpiles	Aggregate Quality	Particle Size/Shape Gradation Specific Gravity Absorption Moisture Content
		Stockpile Construction	Pile Separation
		Maintain Stockpile Quality	Segregation Contamination Consistency
	Asphalt Binder	Check Quantity	Verify Delivery Invoice
Plant Mixing	Aggregate Blending	Loading Practices	Bin Contamination Bin Levels Segregation
		Proportioning	Feed Rates
		Equipment Conditions	Bin and Belt Condition Belt /Weigh Scale Calibration
	Asphalt Binder Delivery	Binder Proportioning	Check Flow
		Equipment Condition	Pump and Pipeline Calibration Pump and Pipeline Condition
	Mix Plant	Plant Operations	Burner Efficiency Mix Discharge RAP operations Plant Efficiency
		Mix Quality	Mix Temperature
	Dust Collector	Equipment Conditions	Proper Housings/Cover Fines Return Location
	Storage	Storage Operations	Storage Duration Loading Method (Segregation)
		Mix Quality	Mix Temperature Moisture Content
Mix Transport	Load-Out	Loading Practices	Multiple Drops (Segregation) Use Tarps
		Truck Conditions	Clean Beds Proper Release Agents
		Mix Quality	Mix Temperature
	Weigh Scale	None Identified as Important	
Laydown	Base Condition	Surface Condition	Dry Surface Clean and Uniform Surface
		Tack Coat	Tack Coat Temperature Uniform Tack Coat
	Paving	Mix Delivery	Check Truck Ticket Calculate Yield Mix Temperature
		Equipment Conditions	Proper Working Order (Maintenance)
		Pavement Specifications	Establish Grade Reference Grade Control Depth Checking Slope Accuracy Width
		Paver Operations	Constant Paver Rate Paving Inconsistencies Crown Joint Construction
		Pavement Quality	Crown Joint Construction
	Rolling	Rolling Operations	Mat Temperature Establish Rolling Pattern Maintain Rolling Pattern Consistent Crown Joint Construction
		Pavement Quality	Density Smoothness

tasks so that operators, inspectors, and supervisors understand where to allocate available inspection resources. It may be possible that some of the inspection tasks commonly performed do not affect pavement quality as greatly as other less-frequently performed inspection tasks. Using this method provides a basis to reduce or eliminate some inspection tasks, allowing an allocation of resources to the more critical areas.

The Juran seriousness classification uses a system developed by Bell Telephone Laboratories, entitled the Bell System. The Bell System uses three basic steps: (1) determine the number of classes to be used (e.g., the Bell System has four classes), (2) define each of those classes, and (3) classify each defect within a class *(11)*. The specific definitions of the Bell System's four classes are shown in Table 5.3. The Bell System has been useful for other industries to structure seriousness classifications specific to those industries, such as machinery fabrication or food processing companies. An interdepartmental committee should develop seriousness classifications, according to the Bell System. This concept could be applied to HMA construction through joint agency/contractor technical committees responsible for evaluating and modifying the QA program.

When developing the seriousness classification for HMA construction, the Bell System's four classes were reduced to three more distinct classes by removing the Moderately Serious classification. This was done by the research team in an attempt to make the classification simpler and more appropriate to HMA construction. Table 5.4 includes definitions of these three classes that may be used for seriousness classification of HMA overlay inspection.

Table 5.5 shows an example of how seriousness classification was used to classify inspection tasks under the particular inspection activity, "stockpile quality." They are all "Class A—Very Serious" inspection tasks. They are assigned to this class because, according to HMA literature, at least one of the definitions given for "Class A—Very Serious" is fulfilled. If aggregate stockpile quality inspection tasks are not completed, defects in the mix may occur that "Will surely lead to problems with the mix quality and lead to pavement failure." These tables provide a tool for agencies and contractors to evaluate inspection tasks and focus inspection resources on the tasks that are Very Serious and Serious and place less emphasis on those inspection tasks classified as Not Serious.

Appendix M of the research team's final report presents by process, location, and activity, how the list of inspection tasks for HMA construction could be assigned according to the seriousness classes. According to the suggested classification, most of the inspection tasks are classified as Very Serious. These classifications may vary among agencies and contractors, so this list should be used as a starting point for individual seriousness classifications. The seriousness classification system should give direction for changes and improvements in current practices based on what are thought to be Very Serious or Serious inspection tasks.

The table of important inspections can be used by agencies and contractors when assigning tasks to the seriousness classification. Helpful insight may be found through comparison of which inspection tasks are currently thought to be important by the HMA industry and those that an agency or contractor classified as Very Serious or Serious. For example, Table L.4, in Appendix L of the research team's final report, lists "tack coat temperature" and "uniform tack coat" as important inspection tasks at the "base condition" location. Table M.4, in Appendix M of the research team's final report, classifies "tack coat temperature" as Very Serious and "uniform tack coat" as Serious. When comparing the two tables, an

TABLE 5.3 Seriousness classification of defects *(11)*

Class and Description	Cause operating failure	Cause intermittent operating trouble difficult to locate in the field	Cause substandard performance	Involve increased maintenance or decreased life	Cause increased installation effort by customer	Appearance, finish, or workmanship defects
A Very Serious	Will Surely (Not readily corrected in the field)	Will Surely	-	-	-	-
B Serious	Will Surely (Readily corrected in the field)	-	Will Surely	Will Surely	Major Increase	-
C Moderately Serious	May Possibly	-	Likely To	Likely To	Minor Increase	Major
D Not Serious	Will Not	-	Will Not	Will Not	-	Minor

TABLE 5.4 Seriousness classification for HMA inspection tasks

Class and Description	Definition
A Very Serious	Will surely lead to problems with mix quality and lead to pavement failure; Will surely lead to pavement not meeting specifications; Will surely lead to production downtime; Will surely lead to problems difficult to correct in the field during laydown; and Will surely lead to problems difficult to correct after paving is completed.
B Serious	Will probably lead to problems with the mix and lead to pavement failure; Will surely cause a problem during mix plant production that is easily corrected during laydown construction; Will surely cause a problem that is easily corrected after paving is completed; Will surely lead to a problem less serious than downtime such as a substandard performance; Will likely cause increased maintenance or decreased life of the final pavement; and Will lead to pavement construction not meeting specifications.
C Not Serious	Will not affect production or the life of the pavement; Will lead to minor defects in pavement appearance; and Will lead to management problems regarding material and money, but will not directly affect pavement quality.

agency or contractor may conclude that both of the tasks are thought to be important to the final quality of the pavement, although more emphasis should be placed on completing the inspection task, “tack coat temperature,” based on the seriousness classification.

5.4.2 Pareto Analysis

The seriousness classification can now be used to modify current inspection practices to meet the specific needs of the agency and contractor. To do this, it is necessary to identify field practices for individual projects that allow opportunities for increased efficiency. Pareto analysis is a visual tool used for finding areas in production with the greatest potential for improvement. As a general rule, a process can be broken down into a “vital few” areas with the bulk of improvement potential *(27)*. An injection molding process can be used as an example. The defects that may be encountered on a molded part could be black spots, scratches, or flow lines. Over one-half of the defects found while sampling the product were due to only one of the defects, black spots *(27)*. This information provided a way to allocate a limited amount of resources on correcting the primary defect, black spots.

Pareto analysis was applied to the four HMA processes to provide information on where effort is currently expended. Appendix N of the research team’s final report summarizes the completed inspections for each of the projects for both agency and contractor personnel. Four of the projects used agency data for acceptance, and the nine remaining projects used contractor data for acceptance. Inspection data were not collected for three of the projects. The Pareto analysis was performed separately for agency and contractor acceptance data projects so that inspection task emphasis could be compared among the four processes, as well as between agency acceptance and contractor acceptance projects. The percentages of completed tasks are similar for projects, regardless of whether agency or contractor acceptance data were used.

Table 5.6 shows a portion of Table N.4 found in Appendix N of the research team’s final report. This table presents the inspection tasks completed by the agency during the laydown process. This table consists of the four projects that used agency acceptance data.

The tables in Appendix N reveal that during the laydown process, 16 inspection tasks were completed. When the completed inspection tasks are summed for all four processes (i.e., raw materials, plant mixing, mix transport, and laydown), the total number of completed inspection tasks is 29. Therefore, the percentage of inspection tasks completed during the laydown process is 55 percent (16 divided by 29). This is nearly more than one-half of the total completed inspection tasks, which is due in part to a greater number of possible inspection tasks to complete during this process. This 55 percent is shown in Figure 5.4 along with the percentages for the other three processes for both agency and contractor acceptance projects.

TABLE 5.5 Seriousness classification for aggregate stockpile quality

			Class		
Process	Location	Activity	A – Very Serious	B – Serious	C – Not Serious
Raw Materials	Aggregate Stockpiles	Maintain Stockpile Quality	Segregation Contamination Discoloration Loading Rotation	None Defined	None Defined

TABLE 5.6 Inspection tasks completed by agency on agency acceptance data projects (laydown process)

		Location																		
		Base Condition					Paving										Rolling			
		Task																		
Project	Contractor /Agency	Surface Temperature	Clean and Uniform Surface	Dry Surface	Uniform Tack Coat	Tack Coat Temperature	Mix Temperature	Calculate Yield	Proper Working Order	Correct Grade	Pavement Depth	Slope Accuracy	Width	Segregation	Paver Rate	Paving Inconsistencies	Establish Rolling Pattern	Maintain Rolling Pattern	Joint Construction	Density
Arizona I-10 (Phoenix)	Contractor	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**1**	P	**1**
	Agency	C	C	-	-	-	1	1	-	-	-	-	**1**	C	C	-	-	C	-	-
Arizona I-10 (Benson)	Contractor	-	-	-	-	1	-	-	-	-	-	-	-	P	-	-	-	1	P	P
	Agency	-	**1**	C	1	-	-	C	-	1	C	**1**	C	C	-	-	-	1	**1**	-
Arizona USH-60 (Wittman)	Contractor	-	**1**	-	-	-	-	-	P	1	-	-	1	-	-	-	1	-	1	**1**
	Agency	-	**1**	-	-	-	C	**1**	-	-	-	**1**	-	-	-	-	1	-	1	-
Florida I-10	Contractor	-	1	1	-	-	P	-	-	-	-	-	-	-	-	-	1	1	-	1
	Agency	C	C	1	-	-	**1**	-	-	-	-	1	-	-	**1**	-	1	**1**	-	-

[a] Denoted with "C"if identified during agency interview as important, but no agency completion observed.

[b] Denoted with "P" if identified in QC plan or during contractor interview as important, but no contractor completion observed.

c Bolded when both identified and observed by agency or contractor.

Figure 5.4 illustrates that agencies are performing the greatest percentage of inspection tasks during laydown. This may be due to laydown being a more visual part of construction. By having an inspector at the laydown process, the agency can check that the pavement width, grade, depth, and slope are as specified. An inspector can also monitor difficulties with mix delivery during laydown such as segregation and mix temperature. If problems such as these occur with the mix, the inspector can discuss them with the contractor.

The percentages of completed inspection tasks are significantly less for the remaining three processes. In addition to there being fewer tasks to complete during plant operations, a possible reason for a lower percentage of plant-mixing inspection tasks is that agencies effectively use test results to monitor the plant-mixing processes. Field observation indicated that many agency project staff viewed plant operations as being more of a contractor responsibility, which would explain the lower percentage of inspection tasks.

Pareto analysis was also performed for contractor inspection data. Figure 5.5 illustrates that the greatest percentage of completed inspection tasks for contractors occurs during plant mixing, followed closely by laydown. This seems logical, given that the contractors have control of the mix during the plant operations, as compared with agencies that have more QA activities during laydown. A significant percentage of the total completed inspection tasks were also performed during laydown. This indicates a desire to ensure agency specifications are met.

Pareto analysis provides agencies and contractors with information to make potential changes or improvements in current practice. It also identifies those processes that are consuming the most resources. The seriousness classification of the inspection tasks can be applied to determine if sufficient resources are being allocated to important tasks. From the Pareto analysis, agencies can see that most inspection tasks occur during laydown. This information can be used by agencies to identify opportunities for reallocation of inspection resources. Likewise, contractors can use the information to evaluate their resources expended on inspection tasks during both plant mixing and laydown.

5.4.3 Fishbone Diagrams

Fishbone diagrams, or cause-and-effect diagrams, are a helpful tool to show the influence of external factors on the completion of inspection tasks. These diagrams are tools that enable the user (agency or contractor) to systematically apply a graphical representation leading to the root cause of a particular quality concern *(27)*. The diagrams provide emphasis in the locations most likely to provide solutions to problems at the root cause level, rather than symptoms. Fishbone diagrams

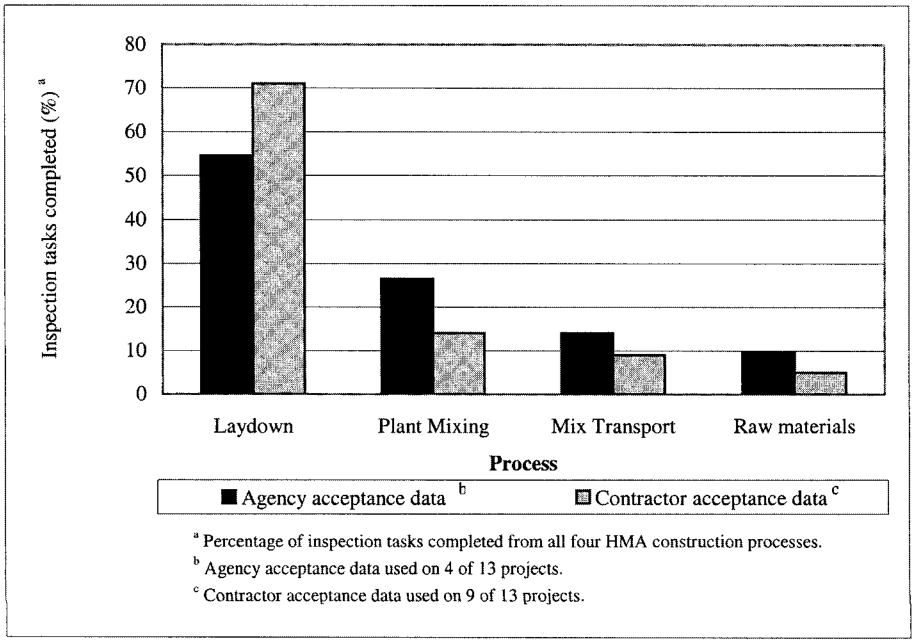


Figure 5.4. Pareto diagram of inspection tasks completed by agency.

also provide focus and a clear measure of the knowledge currently available on a particular problem *(27)*. If changes or improvements are to be made, understanding the factors affecting the inspection task completion is required.

Four different factors—personnel, method, machine, and environment—were chosen as the main factors that can affect quality. They were chosen because they best answer the questions: "What causes the problem?" and "Why do we think that's a problem?" *(27)*.

The four factors were defined for HMA inspection as follows. Personnel affects the inspection task if lack of monitoring by a staff member, or human error, contributes to a task being completed improperly. Methods are a factor if the common practices of a contractor are causing a problem or

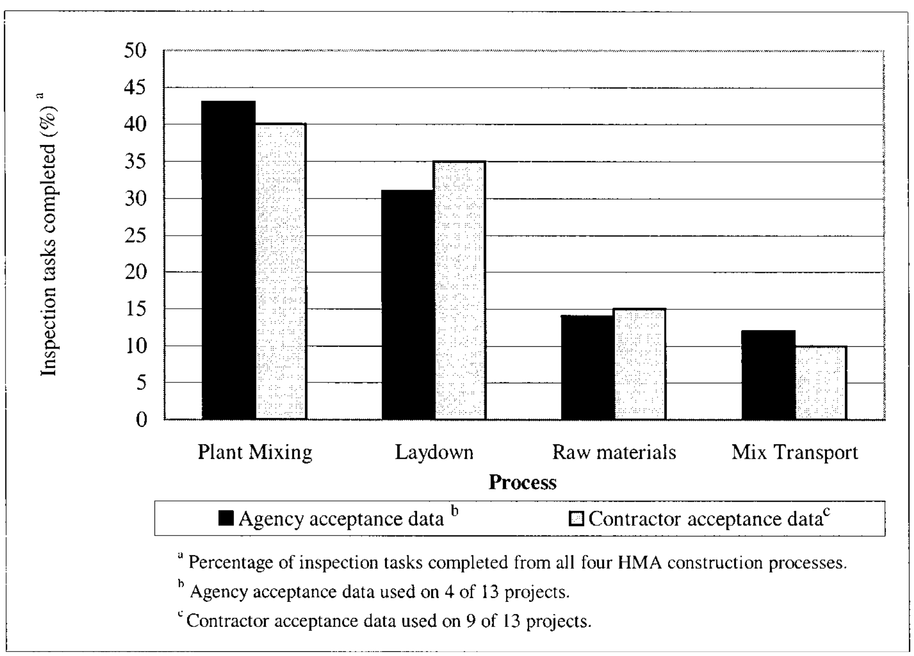


Figure 5.5. Pareto diagram of inspection tasks completed by contractor.

defect in the material. Machines are considered a factor if a problem can be attributed to malfunctioning equipment or equipment that is not in proper working condition. Environmental factors are those factors that are produced by the environment, but independent of equipment working correctly or people conducting the inspection correctly.

By categorizing inspection tasks by these four factors, direction is provided for agencies and contractors on the issues they may encounter when making any changes or improvements in the field. For example, if it is known that a particular HMA process is driven by personnel factors, the completed inspection tasks for that process should be addressed with improved training and/or management of personnel.

Fishbone diagrams illustrate how the four factors affect an inspection task using a specific diagramming technique. A large box contains the inspection task to be analyzed, and smaller boxes contain any of the four factors affecting an inspection task. Figure 5.6 provides an example of a fishbone diagram for aggregate stockpile separation. There are two factors that may affect the mixing of aggregates among different stockpiles, including personnel (loader operator) and method (stockpile construction). The loader operator has the ability to remedy these problems by preventing the mixing of aggregates between piles. Mixing of aggregate between piles can be minimized with simple methods, such as constructing the piles with appropriate spacing or using barrier walls.

Figure 5.7 provides an example of a fishbone diagram for paving equipment condition and illustrates how method and equipment factors affect that inspection task. Problems could be caused by equipment malfunctioning during production causing downtime or problems with the pavement. Contractor methods are a factor because equipment conditions depend on monitoring and scheduled maintenance. If the contractor does not plan, schedule, and execute maintenance of equipment, it is likely that this will eventually lead to equipment malfunctions, reduced productivity, and perhaps substandard quality.

It is recommended that fishbone diagrams be completed for each inspection task so that the effects of the four factors can be visualized. Appendix O of the research team's final report summarizes how each of the observed inspections could be arranged according to these factors, using the fishbone diagram. Table 5.7 shows some of the information found in Appendix O. Both personnel and method affect the inspection task "aggregate stockpile separation." Therefore, pile separation is found under both the personnel and method factors. Also shown is the inspection task "equipment condition" for the "paving" location. Because both methods and equipment affect it, it is listed under both of these factors in the table.

It is useful to apply the four factors to the inspection tasks performed by the agencies and contractors to provide knowledge of those factors affecting the greatest number of inspection tasks. The data provided in Appendix O are graphically presented in Figure 5.8.

The calculations for Figure 5.8 are similar to those used in the Pareto analysis, except that the data are not separated by who performs the acceptance testing. For example, when the number of inspection tasks completed by the agency are added together for all four processes (i.e., raw materials, plant mixing, mix transport, and laydown) for the 13 projects, the total number of completed inspection tasks is **113.** Personnel factors affect **68** of the inspection tasks that were completed during laydown. The 68 tasks include inspection tasks completed at more than one of the 13 projects. For example, if the inspection task, "clean and uniform surface," was observed at three different projects, it would be counted as a personnel inspection task three times. The percentage shown in Figure 5.8 is found by dividing the 68 personnel inspection tasks by the total number of completed inspection tasks, 113, which equals **60** percent. The same calculations are used to find all of the percentages shown in Figure 5.8. The percentages do not add up to 100 percent because some of the inspection tasks were classified under more than one factor.

Most agency inspection tasks were personnel-related—the other three factors were not as significant. This means that when evaluating the effectiveness of current inspection practices, agencies should begin first by evaluating their personnel. There is little that can be done with the other three factors

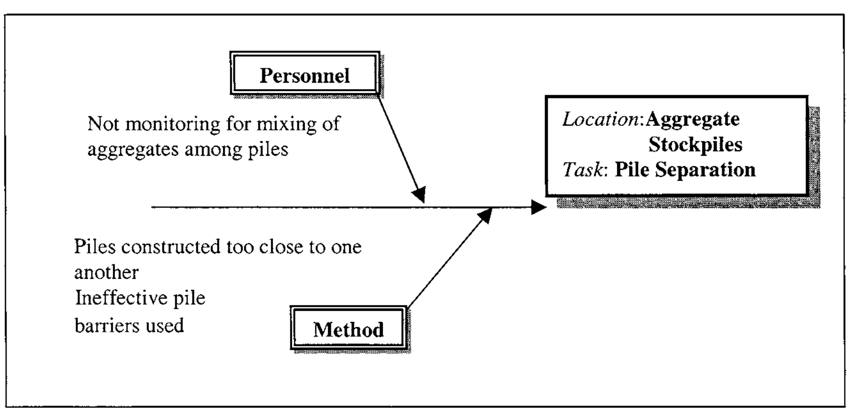


Figure 5.6. Fishbone diagram for aggregate stockpile separation.

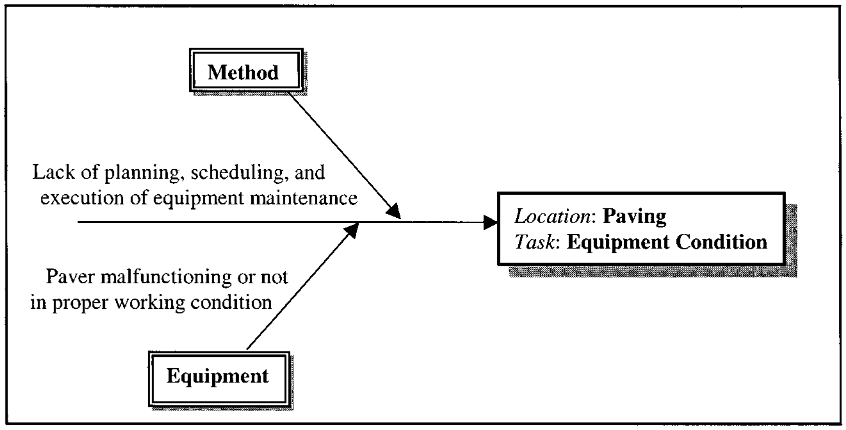


Figure 5.7. Fishbone diagram for paving equipment conditions.

because of the small number of inspection tasks performed. Inspectors must be aware of quality issues and have the knowledge, skill, and attitude to proficiently conduct inspection tasks considered important by the agency. Inspectors must be aware of the important inspections such as those identified as Important or Very Serious. It is important for agencies to ensure that personnel have the proper training, instructions, and procedures to conduct each inspection task.

Figure 5.9 illustrates that the largest percentage of inspection tasks performed by the contractor is also personnel-related. Method and equipment inspection tasks were also frequent during plant mixing and laydown, which is where most inspection tasks take place.

In addition to personnel, contractors must also address methods and equipment. Contractors must ensure that their methods of operations coincide with their own list of important inspections, based on seriousness classification. Examples of inspection tasks affected by contractor methods are plant calibration, equipment condition, and establishing a rolling pattern. Inspection tasks affected by contractor method require planning, scheduling, and execution of tasks. The effect of these factors ultimately has an influence on the quality of HMA overlay. Contractors must acknowledge personnel, method, and equipment factors when considering changes and improvements to current inspection practice.

5.4.4 Resource Allocation

Resource allocation is a method to determine minimum inspection levels considering limited resources. With an idea of where the largest number of inspection tasks are completed, field practices can now be evaluated in terms of person-hours and project characteristics. There could be nearly continuous inspection tasks performed on a project if it were not for time, staffing, and project-specific constraints. A summary is provided of observed agency and contractor inspection task completion based on (1) person-hours, (2) AADT, (3) tonnage, and (4) whether agency or contractor data are used for acceptance.

5.4.4.1 Person-Hours

Observed inspection levels are analyzed first in terms of person-hours. Person-hours were recorded during field data collection using crew balance charts, a graphical means to show the percentage of time spent performing specific activities. They were prepared for each day that an inspector was observed. The number of person-hours for inspection on a project is then calculated by summing the number of hours worked on a given day for each project staff member. For example, if two inspectors were on a project 1 day and they each worked for 8 hr, this created a total of 16 person-hours. Inspection time was analyzed as a percentage of person-hours per day to normalize the data from different projects. There were differences in the data, including production, different numbers of inspectors, and variable number of hours worked per day.

Inspection time was divided into four categories: (1) inspection-direct, (2) inspection-indirect, (3) non-inspection productive, and (4) non-productive. The four categories and

TABLE 5.7 Factors affecting aggregate stockpile separation

Process	Location	Factor			
		Personnel	Method	Equipment	Environment
		Task			
Raw Materials	Aggregate Stockpiles	Pile Separation	Pile Separation	-	-
Laydown	Paving	-	Equipment Condition	Equipment Condition	-

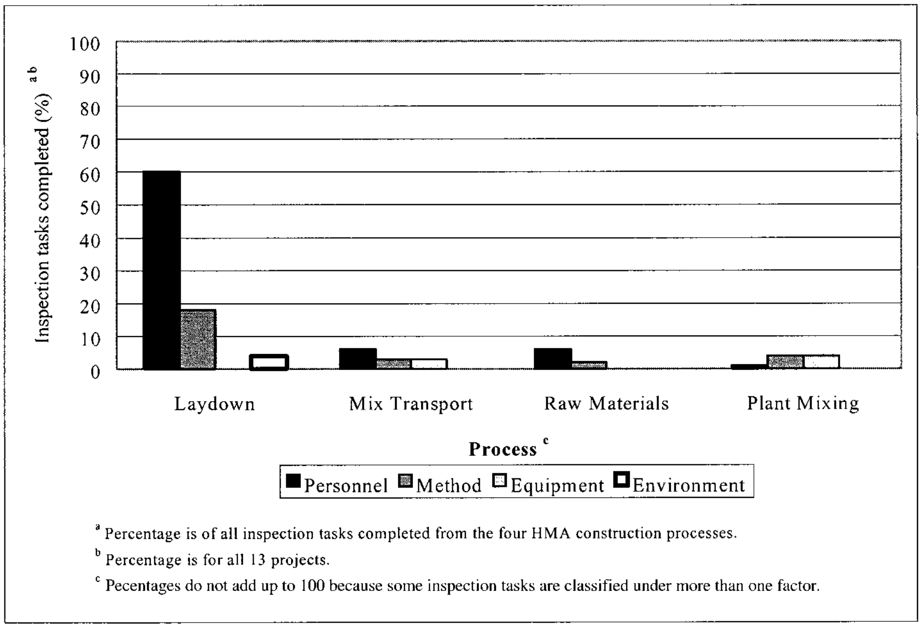


Figure 5.8. Agency-completed inspection tasks per factor.

the activities for each are shown in Table 5.8. Agency inspection-direct time included activities such as plant observation, sampling/testing, travelling to the laboratory, and contractor sampling/testing observation. Base condition, paving, and rolling inspection are also included under this category. Contractor inspection-direct time activities were sampling/testing, plant and laydown observation, and density testing/monitoring. These are all activities that relate directly to the inspection of the HMA overlay.

Those inspections not directly related to the overlay, including documentation, traffic control, and other construction inspection, were classified as Inspection-Indirect time. Time spent in meetings, in the field office, or on other projects was considered productive, but not classified as direct-

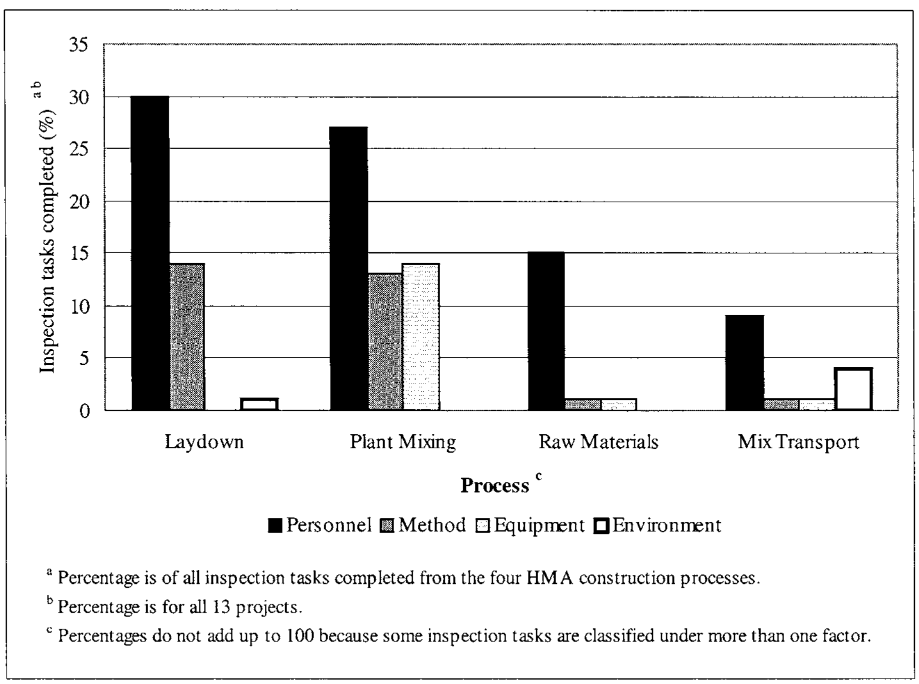


Figure 5.9. Contractor-completed inspection tasks per factor.

TABLE 5.8 Inspection category descriptions

Time Category	Agency	Contractor
Inspection-Direct Plant	Travel to Lab/Materials Transport Plant Observation Sampling/Testing Observe QC Sampling/Testing	Sampling/Testing (Mix and Gradations) Observation
Laydown	Base Condition (milling, inspection in front of paver) Paving (HMA/paving observations, ticket collection, mix temperature, yield calculation) Rolling (density testing/observation, rolling observation) Sampling/Testing (sampling at mat, core layout)	Sampling (Mix) Material Transport Density Testing/Core Sampling Observation
Inspection-Indirect	Documentation Traffic Control Other Construction	Documentation
Non-Inspection Productive	Meetings Field Office Other Projects	Other Projects
Non-Inspection Non-Productive	Rain Delay Idle Off-Site Travel On-Site Waiting for Trucks (between ticket collection and sampling)	Rain Delay Waiting for Trucks (between sampling and density testing)

inspection. A fourth category was created for non-productive activities such as time spent idle, off site (not on other projects), waiting between trucks during ticket collection or sampling, on-site travel, or delays due to rain.

These categories were chosen so that agencies and contractors could understand where inspectors spend time in the field. Percentages indicating productive inspection time were used to provide agencies and contractors with a better understanding of how and where inspectors are expending person-hours. Areas for improvement are clearer when productive time is separated from non-productive time and non-productive activities are identified. These categories provide a way to see clearly productive and non-productive time for field personnel.

The breakdown of observed inspection time for agencies is shown in Figure 5.10. A significant percentage of non-productive person-hours is illustrated in this figure. This is time spent idle, traveling on site, or waiting between trucks, that may be minimized with the knowledge that it is occurring. An effort should be made to increase the percentage of time spent performing productive activities and to find ways to minimize those that are non-productive.

Figure 5.11 provides a breakdown of observed contractor inspection time percentages. Much of the contractor inspection person-hours were allocated to productive activities, with a small percentage of time spent waiting for trucks or rain delays. This shows that there is opportunity for improvement. However, there is a much greater minimization of non-productive activities by the contractor than by the agency.

5.4.4.2 Project AADT

Project AADT was investigated because it is an important project characteristic defining the service-level of the pavement. A relationship was found between time allocation to inspection-direct time (see Table 5.8) and project AADT with the collected field data. Figure 5.12 is a scatter plot of the observed agency allocation to inspection-direct time during laydown and project AADT. The total number of data points equals 13. Three ranges of AADT were developed by the research team: low (<15,000), medium (16,000 to 39,000), and high (>40,000). These ranges were chosen using engineering judgment, because they provided an equal grouping among the data and a basis from which to define pavement service-level. Agency laydown inspection time is shown, but agency plant inspection time is not. This is because no trend was observed between plant inspection and project AADT.

The plot suggests that as project AADT increases, inspection-direct time also increases. This may indicate that inspections related directly to the HMA overlay during laydown are currently thought to be more important for projects having higher AADT. The increase in agency laydown inspection is

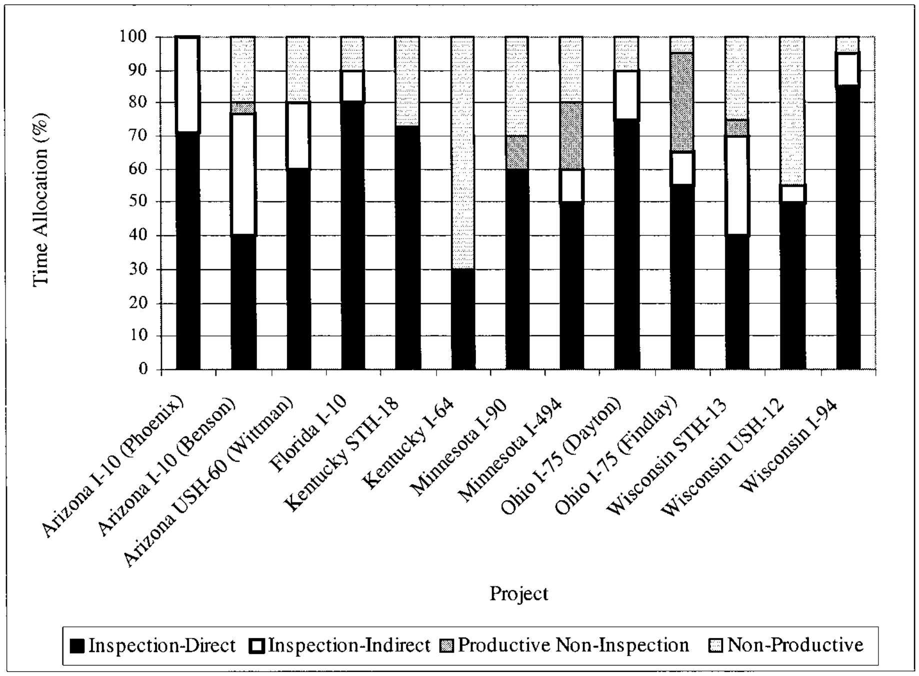


Figure 5.10. Agency inspection time percentages by category.

logical, because a higher AADT is generally a higher profile, or a more important, project, so there is possibly a greater concern for the quality than with a lower AADT project. Higher AADT projects presumably have higher visibility and risk if unacceptable quality is achieved. Because agencies currently conduct more inspection during laydown than at the plant, it is logical that an inspection increase would appear during laydown more so than at the plant.

Figure 5.13 shows contractor inspection-direct time versus project AADT for both the plant and laydown on higher AADT projects. This figure indicates that an increasing percentage of inspection-direct time occurs independent of proj-

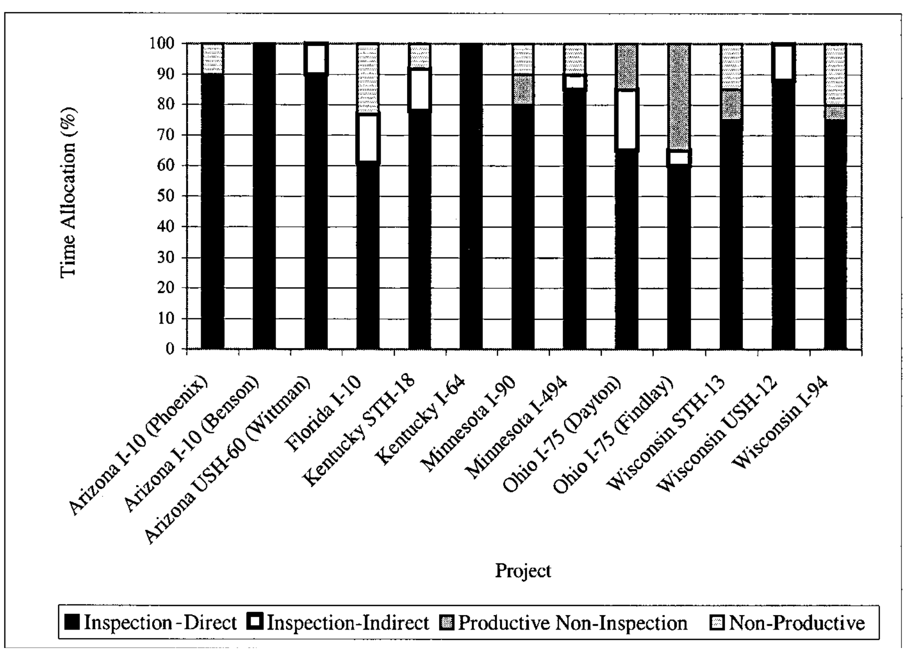


Figure 5.11. Contractor inspection time percentages by category.

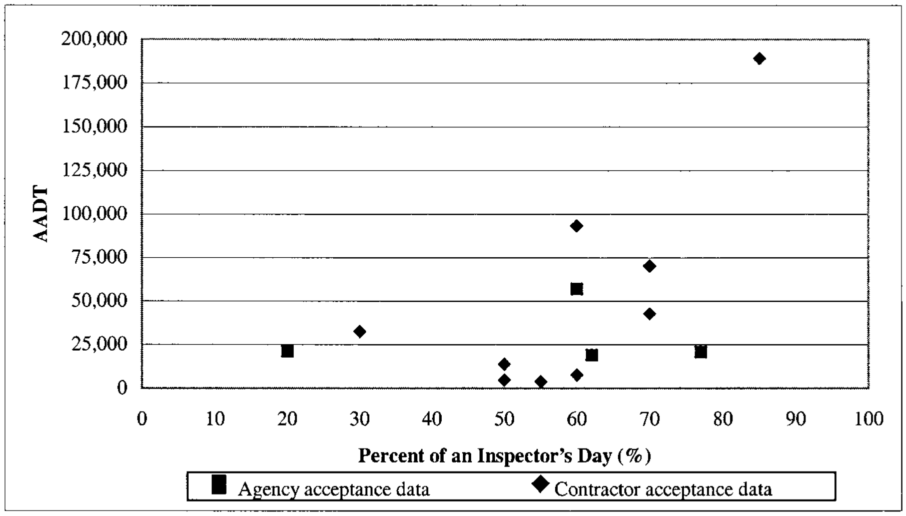


Figure 5.12. Agency laydown inspection-direct time versus project AADT.

ect AADT. Perhaps there are other factors causing the increase in inspection-direct time; however, project AADT does not appear to be a factor.

5.4.4.3 Project Tonnage

In addition to project AADT, project tonnage is a characteristic that can indicate how inspection time will be expended on a project. Figure 5.14 shows the relationship between contractor non-productive time and project size. Three discrete ranges of tonnage volumes were created: low (less than or equal to 40,000), medium (41,000 to 74,000), and high (greater than or equal to 75,000). As project tonnage increases, the contractor non-productive time shows an increase as well. This may be the result of more time spent waiting between trucks for sampling or rollers during density testing, as well as greater travelling distances. Many of the observed data points are 0 percent, because contractors are allocating most of their inspection time in productive ways. Very few of the contractor employees observed spent a significant amount of time idle or performing non-productive activities. No graphs are shown for inspection-direct and inspection-indirect time versus project size, because no clear relationship between them was observed. There is also no graph shown for agency inspection time and project size for the same reason.

5.4.4.4 Agency and Contractor Acceptance Data

Along with the previously mentioned project inspection-time indicators, whether to use agency or contractor data

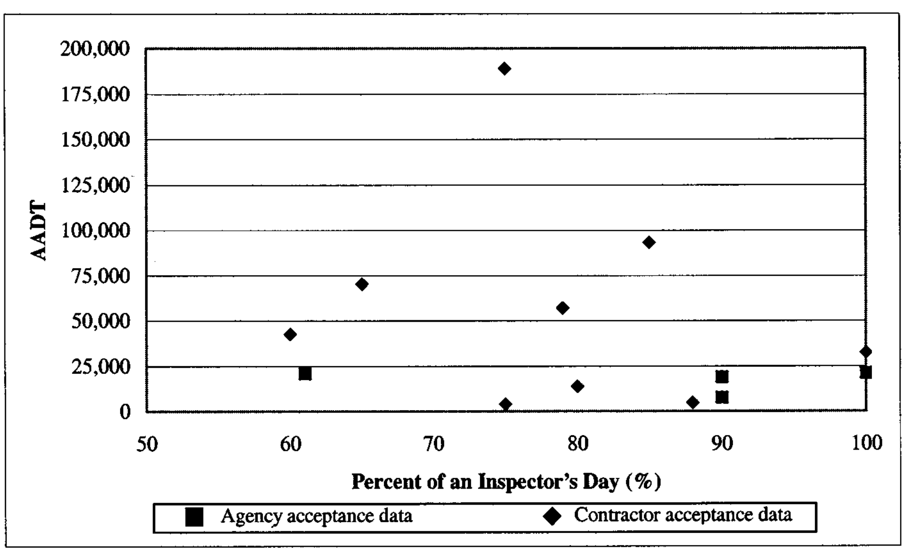


Figure 5.13. Contractor inspection-direct time versus project AADT.

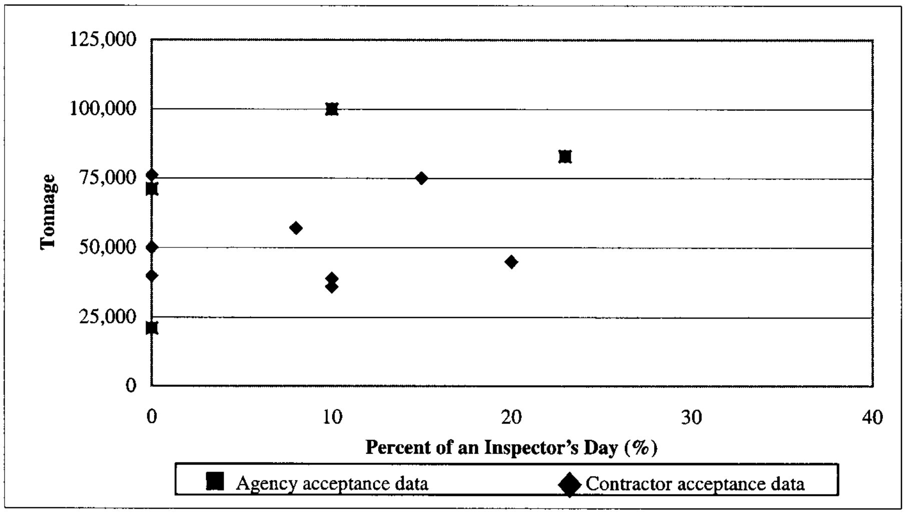


Figure 5.14. Contractor non-productive time versus project size.

for acceptance is also significant for allocating inspection resources. Table 5.9 shows how inspection time was allocated in total person-hours for plant and laydown when either agency or contractor data were used for acceptance. Not all 13 projects are shown under each category for the contractor because of data unavailability.

Agency inspectors spend more time at the plant inspecting if contractor QC data are used for acceptance. The additional plant inspection time is directly attributed to the increase in amount of time that the plant inspector spends observing contractor sampling and testing procedures. The explanation for this is that contractor sampling and testing procedures are more critical to an agency using contractor data; therefore, more time would be spent inspecting their procedures.

Contractors should monitor the production process through quality control, even if the agency is conducting independent acceptance testing. Well-managed HMA contractors are conducting the maximum number of tests possible in a day, so they have no reason to change their testing levels. Therefore, their time allocated to the mix plant does not change.

The observed data used to form the previous conclusions can be found in the tables in Appendix P of the research team's final report. These tables summarize the observed agency and contractor inspection levels based on person-hours, project AADT, and project tonnage volumes. The observed inspection tasks are summarized according to the inspection locations previously described. It is important to view the specific number of people, inspection-time allocation, and

TABLE 5.9 Range of inspection time based on agency or contractor acceptance testing

	Contractor Inspection Time (person-hours per day)		Agency Inspection Time (person-hours per day)	
Project Plant		Laydown	Plant	Laydown
Agency Acceptance Testing				
Arizona I-10 (Phoenix)	9	9	2	27
Arizona I-10 (Benson)	-	10	0	20
Arizona USH-60 (Wittman)	-	10	0	19
Florida I-10	27	9	1	33
Range	**9-27**	**9-10**	**0-2**	**19-33**
Contractor Acceptance Testing				
Kentucky STH-18	30	12	1	10
Minnesota I-90	18	10	4	18
Minnesota I-494	20	-	2	24
Ohio I-75 (Dayton)	10	10	8	8
Ohio I-75 (Findlay)	10	10	1	8
Wisconsin I-94	11	11	0	27
Wisconsin STH-13	-	14	1	28
Wisconsin USH-12	26	0	0	28
Range	**10-30**	**0-14**	**0-8**	**8-28**
Kentucky I-64 data were not included due to incomplete set of data. Not all 13 projects are shown under each category for the contractor because data were not collected on all projects.				

inspection tasks completed for each project type. Specific recommendations are made using these data.

5.5 SUMMARY

A process to determine minimum inspection levels was described by identifying all possible inspection tasks, determining which inspection tasks are important, and using four methods to determine an appropriate inspection level. Twelve physical locations were identified where detailed inspection tasks could be conducted. Within each location, all possible tasks were enumerated. A procedure was developed to identify important inspection tasks. A synthesis of field data and review of both HMA trade publications and QC literature were used to make this determination. Then, four methods were researched that could assist in determining minimum inspection levels: (1) Juran seriousness classification, (2) Pareto analysis, (3) fishbone diagrams, and (4) resource allocation analysis. The first three methods used established QC techniques to interpret the collected field data and then make recommendations of those inspection tasks thought to influence pavement quality. The fourth method, resource allocation, investigated the use of project attributes, including available staff, project AADT, and tonnage to optimize limited inspection resources for the construction project.

CHAPTER **6**

RECOMMENDATIONS

6.1 INTRODUCTION

Recommendations have been provided in the prior sections for developing rational methods to determine testing and inspection levels for HMA overlay construction. These recommendations are the result of data collected and analyzed from the HMA overlay projects selected for the study. The following two sections provide detailed recommendations specific to testing and inspection.

6.2 TESTING LEVELS

Table 6.1 provides summary recommendations for testing levels based on the research in this study. Then, more detailed recommendations are given for contractor quality control, acceptance, and agency verification of contractor acceptance tests.

6.2.1 Quality Control

6.2.1.1 Aggregate Gradation (Raw Materials) Testing Levels

A statistical comparison of means was made using the *t*-test procedure, given in the AASHTO *Implementation Manual for Quality Assurance*, to determine if there was a mean difference between the two sampling locations. The *F*-test procedure was used to determine if the variation between sampling locations was equal or not. As would be intuitively expected, the data indicated that input variation from coldfeeds is transferred to output variation at the plant discharge.

There was strong evidence that coldfeed aggregate gradation tests will have different means than hot-mix samples when evaluated across the entire project. Data show an increase in the mean between coldfeeds and hot-mix samples for nearly all sieve sizes on three projects studied. Shifts in the mean are caused by one or more of four variation components: (1) materials, (2) production, (3) sampling, and (4) testing. Production may cause shifts in the mean through aggregate breakdown from the coldfeeds through the plant. Sampling the material at different locations can introduce bias by not collecting a representative sample. Different testing procedures for coldfeed and hot-mix samples can introduce a systematic bias between test results.

Different mean levels cause concern when using coldfeeds to control mean levels of the final hot-mix aggregate gradation. Contractors must be aware that bias may be introduced when sampling from two different locations. If bias can be attributed to production, contractors may typically add a small percentage of baghouse dust during the mix design to simulate the generation of dust from aggregates during mixing. Sampling at different locations has a definite possibility of introducing bias. Different test procedures for coldfeed and hot-mix aggregates are a potential source of bias. Contractors should continue to collect data to quantify changes in coldfeeds and hot-mix samples for percentage passing the sieves for a given aggregate type.

A quantity-based testing frequency is recommended for controlling aggregate gradation to minimize aggregate particle size variation within the final mixture. The sampling frequency should be calculated using the methodology described earlier. It is also recommended that sample sizes of n equal to 4 or larger be collected to determine if a difference exists. Agencies should recommend in the specifications that this procedure be performed by the contractor at the start of production if the contractor chooses to use coldfeeds to control final hot-mix aggregate gradation. Mandated testing levels for quality control should only be specified when contractor data are used for acceptance.

6.2.1.2 Plant Mix Testing Levels

A formal ANOVA was used to measure sources of variation found in the plant-mixing data. The data analysis concluded that between-day variation is significant. Contractors can expect significant variation in the daily production of the plant-mixing process. A necessary component when establishing testing levels for quality control is to detect assignable causes of variation on a daily basis. Between-day variation can be attributed to several assignable causes, such as material properties, environmental conditions, personnel decisions, equipment, and construction methods. Controlling variation requires constant feedback-testing to identify discernible trends or variation and make appropriate adjustments so that target levels are met.

A rational method of determining QC testing levels for the plant-mixing process is to allocate QC testing to the significant

TABLE 6.1 Summary recommendations for testing levels

Implementation Issue	Recommendation
Agency or Contractor Data for Acceptance	• Agency has worked in partnership with the contractor and has reasonable assurance of contractor's honesty and integrity. • From a resource perspective, the testing alternative that minimizes the cumulative number of contractor and agency tests on a given project is by using contractor data for acceptance. Several considerations were given when selecting agency or contractor tests for acceptance. • This decision will directly affect the resource level requirements for the agency on any project. • If contractor data are used for acceptance, the agency can allocate limited resources across a larger number of projects as well as specify specific items to be included in the Quality Control plan.
Sampling	• Choose a location that provides the best opportunity to collect a representative sample of the work. Consider the four sampling attributes (material characteristics, resources, interaction of construction operations and sampling, and safety) and other state-specific attributes during this decision. • Hot-mix acceptance samples should be collected at the laydown operation to provide the most representative sample. • Perform all sampling in strict accordance with randomization principles to obtain unbiased statistical estimates. • Agencies should either collect their own samples or witness the contractor collect and split the sample and have the agency field representative immediately take possession of their portion of the split-sample. • Review the relative advantages and disadvantages of time-based and quantity-based sampling. • Determine sampling frequency using estimated times to complete each test. Time-based: sampling rate longest testing time units = samples-per-hour Quantity-based: sampling rate maximum production rate units = samples-per-tons (convert to area or length as desired) • Collect samples with the presumption they will be tested to comply with randomization principles. If not all samples are tested, remove them in accordance with randomization principles.
Quality Control Testing Levels	• Perform an independent-sample t-test and F-test between coldfeed and plant-produced aggregate gradation to determine if both locations provide a similar mean and standard deviation for quality control. • <u>Stockpiles</u> - Use quantity-based sampling and testing when building stockpiles to ensure aggregate particle sizes are consistent and variation is minimized. Use production rates to determine a quantity-based sampling frequency. Use control charts to provide visual interpretation of particle size when building stockpiles. • <u>Plant Mixing</u> - Develop a sampling and testing frequency using the sampling methodology. Daily testing is recommended since variation changes between days. If there are minimal testing resources available, the contractor should ensure that samples are collected and tested within each day of production, as opposed to high-frequency testing on periodic days. Use control charts to provide a visual interpretation of the process. • <u>Density</u> - Develop a sampling and testing frequency using the sampling methodology. Significant variation was found between days. If there are minimal testing resources available, the contractor should ensure that samples are collected and tested within each day of production, as opposed to high-frequency testing on periodic days. Use Equation 4.4 as a starting point to determine the number of daily density tests for quality control. Define the rolling zone as a sublot and then randomly collect samples within each rolling zone to obtain unbiased statistical estimates of the compaction process. • Apply the Analysis of Variance models to the plant mixing and compaction processes to determine if variation is random or if variation can be assigned to days of production. Work with consultants to interpret the data if no available expertise exists within the organization. • It is recommended that more samples be collected and tested, rather than more subsample testing on each sample. For example, a larger percentage of project variation for air voids is found with materials, production, and sampling than testing.

TABLE 6.1 *(Continued)*

Acceptance Testing Levels	• Evaluate project data to determine if specification limits can be achieved by the contractor. Until a clear relationship is found between construction variability and performance, a reasonable alternative is to set specification limits using construction variability. • An appropriate specification range should be chosen that allows for slight variations in the process mean and variability to allow contractors to achieve PWL equal to 90 (100-percent payment). • It is recommended that an analysis of normality be performed on state-specific data to ensure that the distribution of the sample data satisfies the statistical requirements when combining material. Tests for normality should include large sample sizes (10 or more) to ensure the test is made with high reliability. • Combining material from different days of production into a single lot that have different statistical characteristics can produce lots that are not normally distributed. Also the standard deviation may increase with material from different days. An analysis is recommended to understand these characteristics when determining sample size for the lot. • Goals of developing statistically based QA specifications should be to understand the effects of sample size on the pay factor and to minimize, within practical limits, the risks to both the agency and contractor. • An analysis of risks associated with sample size should be performed using computer simulation with either the program provided in Appendix C of this report or the FHWA Demonstration Project 89 Software *(28).* Perform an analysis at both the AQL and RQL. • A rational acceptance specification would be established using the prior sampling methodology and the sample size that tolerates an acceptable risk level. The sublot size would correspond to the risk level that an agency and contractor are willing to tolerate for a particular project (based on a project's criticality).
Verification Testing Levels	• Split-sample testing is recommended over independent-sample testing because the number of samples required for a comparison is less, and the variation from materials, production, and sampling is removed from the comparison. If an agency and contractor choose to perform independent-sampling, the appropriate number of samples necessary to enable a credible comparison could be approximately doubled, provided all other variables are equal. Trust, cooperation, and a partnership are an absolute requirement between the agency and contractor when performing split-sample verification testing. • The number of verification tests for a given lot is a function of the following: (1) variation between tests, (2) difference between means, (3) probability that the mean difference is contained within a defined acceptable region, and (4) probability that the mean difference is accepted outside the defined acceptable region. • Create a testing tolerance table using Equation 4.6. • Collect and evaluate several split-sample test results at the beginning of production to evaluate the project-specific s_D value with the population σ_D value.

sources of variation. If there is a minimum amount of testing resources available, the contractor should ensure that samples are collected and tested within each day of production, as opposed to high-frequency testing on periodic days. In other words, daily testing is recommended given the presence of significant between-day variation. As a starting point, it is recommended that a minimum of n equal to 3 complete hot-mix tests be made per day for quality control. This minimum testing level is derived from the time one technician can comfortably complete three tests within 1 production day. The number of daily tests may be increased to n equal to 4 if the same technician can perform the tests without increasing costs. More tests may require an additional technician, possibly require more laboratory space and equipment, and exceed the current lag times for test results.

Specifications should be written that recommend the contractor perform QC testing on a daily basis. It is recommended that a time-based or quantity-based sampling frequency be developed using the methodology provided earlier. It is further recommended that contractors use control charts to assist in discerning any trends or unusual behavior in the process. Control charts are well documented and a description is provided in the AASHTO *Implementation Manual for Quality Assurance*.

6.2.1.3 Density Testing Levels

A formal ANOVA was used to measure sources of variation found in the density data. The data analysis concluded that between-day variation is significant. It is apparent that within-day variation exists, but variation between days also contributes to overall project variation. Contractors can expect the compaction process to have significant variation between days. Successive days of production could operate at a consistent mean level with a certain degree of variation, but a series of production days within a project will have days operating at different mean levels, producing significant between-day variation. Factors contributing to significant between-day variation are changes in mix properties, compaction methods, and environmental conditions. Day-to-day variation caused by changes in the mean level can be expected when weather conditions constantly change while the compaction process is kept constant. A day with constant external factors and little change in the compaction process can expect to yield less variation. A necessary component of establishing testing levels for contractor density testing is to perform testing on a daily basis and apply limited testing resources to this significant variation source.

It is recommended that contractor density testing be performed on a daily basis to understand and control the compaction process. Without daily testing, between-day variation can go undetected, thus several tests are needed within a day to control production levels so that targets are met and variation is minimized. If minimal testing resources are available, the contractor should ensure that density tests are collected and tested within each day of production.

A rational method of determining daily density testing levels is to apply principles of the average method. It is recommended that Equation 4.4 be used as a starting point to determine the number of daily density tests for quality control. There will be an increased confidence in the daily average as the number of tests is increased, while variation remains constant.

A rational way of obtaining unbiased density samples with a nuclear density gauge is defining the rolling zone as a stratified sublot. Density test sites would then be randomly chosen within each rolling zone. Unbiased estimates for the mean and standard deviation for a given day of production would then be obtained by combining the individual rolling zone tests into a single data set.

6.2.2 Quality Acceptance

There are several interrelated, yet discrete, variables that an agency and contractor can evaluate when determining minimum acceptance testing levels. These variables include (1) specification limits, (2) distribution of samples within a lot, and (3) pay factors. Within each of these variables, sample size and corresponding lot sizes should be analyzed.

6.2.2.1 Specification Limits

An analysis was conducted to determine if the specification limits are achievable during production. The standard deviations measured from the field projects were used to develop an expected production range containing 90 percent of the tests. A 90-percent range was chosen because many states have chosen an AQL value of PWL equal to 90. The results concluded that specification limits are achievable during production.

It is recommended that specification limits be evaluated using state-specific production data and experience during QA program development. Because it has been difficult to translate the relationship between construction variability and performance accurately, a reasonable alternative is to set specification limits using known construction variability. As performance data are collected, they can then be compared with the level of contractor variability used to construct the project.

6.2.2.2 Distribution of Samples Within a Lot

Either time-based or quantity-based sampling could be used to create the lot. Typical time-based lots are 1 production day, while quantity-based lots are an arbitrary tonnage. In either lot configuration, the acceptance plan must recognize how samples are distributed so that necessary statistical assumptions are satisfied. An experiment was conducted to test the normality assumption on actual projects. It was concluded with strong evidence that both hot-mix and density test data within 1 day's production, and samples combined from adjoining days, are normally distributed.

It is recommended that an analysis of normality be performed on state-specific data to ensure that the distribution of the sample data satisfies the statistical requirements when combining material. Tests for normality between days should include large sample sizes (10 or more) to ensure the test is made with high reliability.

Combining material from different days of production into a single lot that has different statistical characteristics can produce lots that are not normally distributed. The standard deviation may increase from estimates found in within-day lots. Many states elect to use quantity-based sampling using material from different days, and it is recommended that this type of analysis be performed so that both the agency and contractor are aware of the potential consequences.

6.2.2.3 Pay Factors

The relationship of pay factors with sample size was investigated to demonstrate how the number of samples could be specified for a lot. Three pay factor equations currently specified by the states were used in the analysis. Goals of developing statistically based QA specifications should be to understand the effects of sample size on the pay factor and to minimize (within practical limits) the risks to both the agency and contractor. An analysis of risks associated with sample size should be performed using computer simulation with either the program shown in Appendix C or the FHWA Demonstration Project 89 Software *(28)*. This should include performing the analysis at both the AQL and RQL values.

6.2.2.4 Decision Analysis

The severity of the pay factor will assist in developing the testing level. A lenient pay scale may be more appropriate where the impact of construction quality has a lessened impact on the life-cycle costs. Severe pay scales should be assigned to projects having higher life-cycle costs, such as an interstate highway having higher traffic levels (AADT). Consider when determining testing levels that either small or large sample sizes could be specified for less critical projects, while larger sample sizes are needed for more critical projects.

A rational acceptance specification would be established using the prior sampling methodology and the sample size that tolerates an acceptable risk level. The sublot size would correspond to the risk level that an agency and contractor are willing to tolerate for a particular project (based on a project's criticality).

6.2.3 Verification

A fundamental issue when determining verification testing levels is whether split samples or independent samples should be used to compare contractor and agency tests. Independent-sampling will require approximately twice as many tests as split-sampling for making a comparison, provided all other variables are equal and there is no covariance between laboratories.

A statistical procedure was developed for making a verification test comparison between split-sample test results using pre-determined testing tolerances. It is recommended that testing tolerances be developed using the standard deviation between laboratory tests from historical data (σ_d). There are large amounts of split-sample test data available to estimate σ_d, and states should begin organizing these data to provide population estimates for the standard deviation of differences between laboratories on individual projects.

The number of verification tests for a given lot is a function of the following: (1) variation between tests, (2) difference between means, (3) probability that the mean difference is contained within a defined acceptable region, and (4) probability that the mean difference is accepted outside the defined acceptable region. Use Equation 4.6 to set the number of verification tests based on these input variables. A single test, or the mean of a group of tests, can be used when making a comparison.

An implementation procedure was developed for the verification process. This procedure evaluates the split-sample variability between laboratories at project start-up. It is recommended that this procedure be performed on all projects at the start of production to confirm that similar variabilities are found between sample project data and historical data.

6.3 INSPECTION LEVELS

Recommendations are provided for minimum inspection levels based on the research summarized in this report. The recommendations shown in Table 6.2 were made from information gained using the four methods selected for analysis of inspection data.

The following recommended inspection levels are provided for both the agency and contractor. Agency recommendations are based on four limiting factors: (1) project start-up inspection requirements, (2) available project staff, (3) AADT, and (4) whether agency or contractor data are used for acceptance. Contractor recommendations are based on the availability of two technicians, one at the plant and one at laydown, with limited time allotted for inspection in addition to testing functions.

6.3.1 Agency Recommendations

Recommendations for agency inspection levels are based on important inspection tasks from the literature, seriousness classifications, and observed trends. Inspection levels also depend on project-specific characteristics and limited agency resources. Specific recommendations are given below for project start-up, available project staff, and project AADT.

6.3.1.1 Project Start-Up

Project start-up requires additional inspection to ensure that a project begins satisfactorily. Table 6.3 lists suggested inspection tasks to be completed by an agency inspector at the plant and laydown during project start-up and the estimated time to complete each task. A total of 3 hr for one inspector is suggested for the completion of agency project start-up inspection. From a practical point of view, it is recommended that one-half-day be allocated. This one-half-day would be in addition to any normal production inspection.

6.3.1.2 Available Project Staff

Many times an agency will be required to make project staffing decisions based on the available staff for a particu-

TABLE 6.2 Summary of recommendations for inspection

Method	Purpose	Recommendation
Juran Seriousness Classification	• Provide resources for inspection tasks classified as "Very Serious" and "Serious"	• Agency and contractor personnel together should classify tasks according to seriousness classification using experience and expertise • Make improvements by comparing current practices with suggested inspection task emphasis
Pareto Analysis	• Provide information on HMA processes with bulk of improvement potential	• Agency: Focus on laydown process • Contractor: Focus on plant-mixing and laydown processes
Fishbone Diagrams	• Improve inspection task completion based on four factors: 1. Personnel 2. Method 3. Equipment 4. Environment	• Agency: Focus on personnel factors • Contractor: Focus on personnel, method and equipment factors
Resource Allocation	• Use observed practices to find trends between inspection time and project characteristics	Considerations: • Person-hours • AADT • Tonnage • Agency/Contractor acceptance data

TABLE 6.3 Recommended project start-up inspection for the agency

Process	Location	Activity	Task	Estimated Time (Min)
Raw Materials	Aggregate Stockpiles	Stockpile Construction	Clean foundation Dry foundation Pile Separation	2 2 2
	Asphalt Binder	Equipment Conditions	Clean Tank Dry Tank Proper Working Condition	2 2 2
Plant Mixing	Aggregate Blending	Equipment Conditions	Bin and Belt Condition Belt/Weigh Scale Calibration Scalping Screen Condition/Use	5 5 5
	Asphalt Binder Delivery	Equipment Conditions	Pump and Pipeline Calibration Pump and Pipeline Conditions	5 5
	Mix Plant	Equipment Conditions	Proper Working Condition Calibration	5 5
	Dust Collector	Equipment Conditions	Proper Housing/Cover Fines Return Location Proper Working Condition	5 5 5
Mix Transport	Weigh Scale	Equipment Conditions	Calibration Proper Working Condition	5 5
Laydown	Paving	Equipment Conditions	Proper Working Condition	5
	Rolling	Rolling Operations Equipment Conditions	Establish Rolling Pattern Proper Working Condition	120 5
			Total Time for Project Start-Up Inspection	202 [a]

[a] 202 minutes is approximately 3 hr, which can be rounded up to 4 hr or one-half-day.

lar day. The following assumptions were made regarding the available time per agency inspector for a particular day: (1) 10-hr work day, (2) 45-min productive work hour, and (3) 20 percent of the work day allocated to other productive activities (indirect-inspection or productive non-inspection). The 20 percent of the work day for other productive activities is an average from the observed project data. The balance of the productive time is allocated for inspection-direct time to directly complete the recommended inspection tasks.

To determine the inspection level possible with an available number of inspectors, it is necessary to estimate the time required to complete each agency inspection task. Table 6.4 provides estimates of the time, in 2- to 10-min increments, necessary for agency inspection tasks. These increments are reasonable estimates for completing the particular inspection task based on field observations.

Table 6.5 provides the suggested inspection level when project staffing is limited to one inspector. This table provides the number of tasks an inspector can perform based on estimated times. The more important inspections from the seriousness classification and important inspection table were chosen to be completed first when only one inspector is available. The required productive time for given inspection tasks is equal to, or less than, the number of productive hours available in a 10-hr work day for one inspector. For example, there are 8 hr of productive time available for one inspector.

Suggested inspection tasks for two inspectors are provided in Table 6.6. One inspector should complete each inspection

TABLE 6.4 Estimated task times for agency inspections

Process	Location	Activity	Task	Estimated time (Min)
Raw Materials	Aggregate Stockpiles	Stockpile Construction	Pile Separation	2
		Stockpile Quality	Segregation	2
			Contamination	2
			Discoloration	2
			Loading Rotation	2
Plant Mixing	Aggregate Blending	Loading Practices	Bin Contamination	2
			Bin Levels	2
			Segregation	2
		Proportioning	Feeder Rates	5
	sphalt Binder Delivery	Binder Proportioning	Verify Meter Reading	5
	Mix Plant	Mix Quality	Contractor Sampling/Testing Procedures [a]	10
	Dust Collector	Dust Collector Operations	Dust Carry Out	5
	Storage	Storage Operations	Loading Method	5
			Storage Duration	5
Mix Transport	Load-Out	Loading Practices	Multiple Drops to Truck	5
		Truck Conditions	Tarp Usage	5
			Clean Beds	5
			Proper Release Agents	5
		Mix Quality	Mix Temperature	5
Laydown	Base Condition	Surface Condition	Surface Temperature	5
			Dry Surface	5
			Clean and Uniform Surface	5
		Tack Coat	Tack Coat Temperature	5
			Uniform Tack Coat	5
	Paving	Mix Delivery	Collect Truck Ticket	1
			Mix Temperature	5
			Calculate Yield	5
		Pavement Specifications	Slope	5
			Grade	5
			Width	5
			Depth	5
		Paving Operations	Inconsistencies/Segregation	5
			Constant Paver Rate	5
	Rolling	Rolling Operations	Maintain Rolling Pattern	5
			Joint Construction	5
			Consistent Crown	5
			Continuous Roller Movement	5
		Pavement Quality	Surface Texture	5
			Contractor Density Testing	5

[a] Depends on whether agency or contractor data are used for acceptance.

TABLE 6.5 Agency inspection tasks with one inspector

Process	Location	Activity	Task	Number of tasks per 10-hour work day	Estimated time for 10-hr work day (min)
Raw Materials	-	-	-	-	-
Plant Mixing	Storage	Storage Operations	Storage Duration	As needed	-
Mix Transport	Load-Out	Loading Practices	Tarp Usage	As needed	-
Laydown	Base Condition	Surface Condition	Clean and Uniform Surface	10	50
			Dry Surface	As needed	-
	Paving	Mix Delivery	Collect Truck Ticket	90	90
		Pavement Specifications	Grade	10	50
			Width	10	50
			Depth	10	50
			Joint Construction	As needed	-
		Paver Operations	Inconsistencies/ Segregation	10	50
		Time required to complete inspection tasks, (minutes)			340
		Time required to complete inspection tasks, (hours) [a]			6
		Time for other productive activities, (hours) [b]			2
		Required productive time [c] (sum of previous 2 rows), hours			8

[a] All inspection times assume a 10-hr work day with a 45-min productive work-hour.
[b] Based on an average of 20 percent of 10-hr work day found from observed data.
[c] Productive time includes time for direct inspection, indirect inspection, or productive non-inspection.

TABLE 6.6 Agency inspection tasks with two inspectors

Process	Location	Activity	Task	Number of tasks per 10-hour work day	Estimated time for 10-hr work day (min)
Raw Materials *Inspector One*	Aggregate Stockpiles	Stockpile Quality	Segregation	1	2
			Contamination	1	2
			Discoloration	1	2
			Loading Rotation	1	2
Plant Mixing *Inspector One*	Aggregate Blending	Proportioning	Feeder Rates	1	5
		Loading Practices	Bin Contamination	1	2
			Bin Levels	1	2
			Segregation	1	2
	Mix Plant	Mix Quality	Contractor Sampling/Testing Procedures	Variable	-
	Storage	Storage Operations	Storage Duration	As needed	-
Mix Transport *Inspector One*	Load-Out	Loading Practices	Tarp Usage	As needed	-
			Multiple Drops to Truck	1	5
		Truck Conditions	Clean Beds	1	5
			Proper Release Agents	1	5
Laydown	Base Condition *Inspector One*	Surface Condition	Clean and Uniform Surface	10	50
			Dry Surface	As needed	-
		Tack Coat	Uniform Tack Coat	2	10
			Collect Truck Ticket	90	90
	Paving *Inspector Two*	Mix Delivery	Calculate Yield *(Inspector One)*	4	20
		Pavement Specifications	Grade*(Inspector One)*	20	100
			Width	20	100
			Depth	20	100
			Slope	10	50
			Joint Construction	As needed	-
		Paver Operations	Inconsistencies/ Segregation	20	100
			Surface Texture	2	10
	Time required to complete inspection tasks, (minutes)				664
	Time required to complete inspection tasks, person-hours [a]				11
	Time for other productive activities, person-hours [b]				4
	Required productive time [c] (sum of previous 2 rows), person-hours				15

[a] All inspection times assume a 10-hr work day with a 45-min productive work-hour.
[b] Based on an average of 20 percent of 10-hr work day found from observed data.
[c] Productive time includes time for direct inspection, indirect inspection, or productive non-inspection.

task that falls under the processes of raw materials, plant mixing, and mix transport. This same inspector would also have time during the day to complete the inspection tasks at the laydown operation, including the base condition location, calculating yield, and monitoring the grade at the paving location. The second inspector could be responsible for ensuring the construction specifications are met, including width, depth, slope, segregation, and surface texture.

Table 6.7 provides suggested inspection tasks when there are three inspectors available. Inspector number one is responsible for inspecting the raw materials, plant mixing, and mix transport processes. Given that these three processes do not have extensive inspection requirements, this inspector can also inspect at the base condition location, collect truck tickets, and calculate yield during the laydown process. Inspector number two can measure the mix temperature during mix

TABLE 6.7 Agency inspection tasks with three inspectors

Process	Location	Activity	Task	Number of tasks per 10-hr work day	Estimated time for 10-hr work day (min)
Raw Materials *Inspector One*	Aggregate Stockpiles	Stockpile Quality	Segregation	2	5
			Contamination	2	5
			Discoloration	2	5
			Loading Rotation	2	5
Plant Mixing *Inspector One*	Aggregate Blending	Proportioning	Feeder Rates	2	10
		Loading Practices	Bin Contamination	2	4
			Bin Levels	5	4
			Segregation	5	4
	Mix Plant	Mix Quality	Contractor Sampling/Testing Procedures	Variable	-
	Storage	Storage Operations	Storage Duration	As needed	-
			Loading Method	1	5
Mix Transport *Inspector One*	Load-Out	Loading Practices	Tarp Usage	As needed	-
			Multiple Drops to Truck	2	10
		Truck Conditions	Clean Beds	2	10
			Proper Release Agents	1	5
		Mix Quality	Mix Temperature	2	10
Laydown	Base Condition *Inspector One*	Surface Condition	Clean and Uniform Surface	10	50
			Dry Surface	As needed	-
		Tack Coat	Tack Coat Temperature	2	10
			Uniform Tack Coat	10	50
	Paving	Mix Delivery *Inspector One*	Collect Truck Ticket	90	90
			Calculate Yield	10	50
			Mix Temperature *(Inspector Two)*	15	75
		Pavement Spec.'s *Inspector Three*	Grade	20	100
			Width	20	100
			Depth	20	100
			Slope	20	100
			Joint Construction	As needed	-
		Paver Operations *Inspector Two*	Inconsistencies/Segregation	20	100
			Constant Paver Rate	As needed	-
	Rolling *Inspector Two*	Roller Operations	Maintain Rolling Pattern	5	25
			Continuous Roller Movement	5	25
		Pavement Quality	Surface Texture	5	25
			Density (monitor contractor)	2	10
	Time required to complete inspection tasks, (minutes)				992
	Time required to complete inspection tasks, person-hours [a]				17
	Time for other productive activities, person-hours [b]				6
	Required productive time [c] (sum of previous 2 rows), person-hours				23

[a] All inspection times assume a 10-hr work day with a 45-min productive work-hour.
[b] Based on an average of 20 percent of 10-hr work day found from observed data.
[c] Productive time includes time for direct inspection, indirect inspection, or productive non-inspection.

delivery and conduct inspection tasks for the following activities: (1) paver operations, (2) roller operations, and (3) pavement quality. Inspector number three can complete each of the inspection tasks for the pavement specification activity concurrent with inspector number two.

6.3.1.3 Project AADT

Project AADT can be used as a guide for staffing levels. The number of staff increases with the number of inspection tasks. Low and medium AADT ranges (<15,000 and 16,000 to 39,000) require two inspectors, while high AADT (>40,000) requires three inspectors.

Tables 6.8 through 6.10 list suggested inspection tasks depending on the project AADT range. The data collected from the projects in this study suggest that inspection-direct time increases as project AADT increases. Therefore, the recommended number of inspection tasks increases with project AADT. The same assumptions were made for these recommendations regarding the available time per day per inspector, with the exception that the percentage of time for other

TABLE 6.8 Agency recommendations for low level AADT (<15,000)

Process	Location	Activity	Task	Number of tasks per 10-hr work day	Estimated time for 10-hr work day (min)
Raw Materials	Aggregate Stockpiles	Stockpile Quality	Segregation	1	2
Inspector One			Contamination	1	2
			Discoloration	1	2
			Loading Rotation	1	2
Plant Mixing	Aggregate Blending	Loading Practices	Bin Contamination	1	2
Inspector One			Bin Levels	1	2
			Segregation	1	2
		Proportioning	Feeder Rates	1	5
	Mix Plant	Mix Quality	Contractor Sampling/Testing Procedures	Variable	-
	Storage	Storage Operations	Storage Duration	As needed	-
Mix Transport *Inspector One*	Load-Out	Loading Practices	Multiple Drops to Truck	1	5
			Tarp Use	As needed	-
Laydown	Bas e Condition *Inspector One*	Surface Condition	Clean and Uniform Surface	10	50
			Dry Surface	As needed	-
	Paving *Inspector Two*	Mix Delivery	Collect Truck Ticket	90	90
			Calculate Yield	4	20
		Pavement Specifications	Slope 10		50
			Grade	10	50
			Width	10	50
			Depth	10	50
		Paving Operations	Inconsistencies/Segregation	10	50
	Rolling	Rolling Operations	Maintain Rolling Pattern	5	25
	Inspector One	Pavement Quality	Surface Texture	5	25
		Time required to complete inspection tasks, (minutes)			484
		Time required to complete inspection tasks, person-hours			8
		Number of people recommended to complete inspection tasks [a]			**2**

[a] Number of people assumes a 10-hr work day with a 45-min productive work-hour. The second person would be required for only a partial day.

TABLE 6.9 Agency recommendations for medium level AADT (16,000–39,000)

Process	Location	Activity	Task	Number of tasks per 10-hr work day	Estimated time for 10-hr work day (min)
Raw Materials	Aggregate Stockpiles	Stockpile Quality	Segregation	1	2
Inspector One			Contamination	1	2
			Discoloration	1	2
			Loading Rotation	1	2
Plant Mixing	Aggregate Blending	Loading Practices	Bin Contamination	1	2
Inspector One			Bin Levels	1	2
			Segregation	1	2
		Proportioning	Feeder Rates	1	5
	Mix Plant	Mix Quality	Contractor Sampling/Testing Procedures	Variable	-
	Storage	Storage Operations	Storage Duration	As needed	-
Mix Transport *Inspector One*	Load-Out	Loading Practices	Multiple Drops to Truck	2	10
			Tarp Use	As needed	-
		Truck Conditions	Clean Beds	2	10
			Proper Release Agents	2	10
Laydown	Base Condition *Inspector One*	Surface Condition	Clean and Uniform Surface	10	50
			Dry Surface	As needed	-
	Paving *Inspector Two*	Mix Delivery	Collect Truck Ticket	90	90
			Mix Temperature	10	50
			Calculate Yield	5	25
		Pavement Specifications	Slope	10	50
			Grade	10	50
			Width	10	50
			Depth	10	50
		Paving Operations	Inconsistencies/Segregation	10	50
	Rolling *Inspector One*	Rolling Operations	Maintain Rolling Pattern	5	25
			Joint Construction	As needed	-
			Continuous Roller Movement	5	25
		Pavement Quality	Surface Texture	10	50
			Contractor Density Testing	5	25
		Time required to complete inspection tasks, (minutes)			639
		Time required to complete inspection tasks, person-hours			11
		Number of people recommended to complete inspection tasks [a]			**2**

[a] Number of people assumes a 10-hr work day with a 45-min productive work-hour.

productive activities is based on the average time for that specific AADT class from collected field data.

6.3.2 Contractor Recommendations

Contractor recommendations for inspection are different from the agency because quality control serves a different function than quality assurance. Most practitioners recognize that the contractor should be performing quality control. There is an opportunity for certain inspection tasks to be performed by plant and density technicians both at the plant and laydown. Those suggested inspection tasks would provide contractors with a formal way of monitoring the construction process. Because the technicians are not in constant contact with each other throughout the day,

TABLE 6.10 Agency recommendations for high level AADT (>40,000)

Process	Location	Activity	Task	Number of tasks per 10-hr work day	Estimated time for 10-hr work day (min)
Raw Materials *Inspector One*	Aggregate Stockpiles	Stockpile Quality	Segregation	2	4
			Contamination	2	4
			Discoloration	2	4
			Loading Rotation	2	4
Plant Mixing *Inspector One*	Aggregate Blending	Loading Practices	Bin Contamination	2	4
			Bin Levels	2	4
			Segregation	2	4
		Proportioning	Feeder Rates	2	10
	Mix Plant	Mix Quality	Contractor Sampling/ Testing Procedures	Variable	-
	Storage	Storage Operations	Storage Duration	As needed	-
			Loading Methods	2	10
Mix Transport *Inspector One*	Load-Out	Loading Practices	Multiple Drops to Truck	2	10
			Tarp Use	As needed	-
		Truck Conditions	Clean Beds	2	10
			Proper Release Agents	2	10
Laydown	Base Condition *Inspector One*	Surface Condition	Clean, Uniform Surface	20	100
			Dry Surface	As needed	-
		Tack Coat	Uniform Tack Coat	10	50
			Tack Coat Temperature	5	25
	Paving	Mix Delivery *Inspector One*	Collect Truck Ticket	90	90
			Calculate Yield	10	50
			Mix Temperature *(Inspector Two)*	5	25
		Pavement Specifications *Inspector Three*	Slope	20	100
			Grade	20	100
			Width	20	100
			Depth	20	100
		Paving Operations *Inspector Two*	Inconsistencies/ Segregation	20	100
	Rolling	Rolling Operations *Inspector Two*	Maintain Rolling Pattern	5	25
			Joint Construction	As needed	-
			Continuous Roller Movement	5	25
		Pavement Quality *Inspector Two*	Surface Texture	10	50
			Contractor Density Testing	5	25
	Time required to complete inspection tasks, (minutes)				1043
	Time required to complete inspection tasks, person-hours				17.5
	Number of people recommended to complete inspection tasks [a]				**3**

[a] Number of people assumes a 10-hr work day with a 45-min productive work hour.

they would be able to make independent and objective observations.

Table 6.11 summarizes inspection tasks that can be completed by plant and density technicians in their respective locations. These inspection tasks and times are limited by the available time of these two technicians. Given that much of their time is spent collecting samples and performing tests, the inspection tasks listed in the table are provided with the assumption that testing responsibilities would be completed first, and then any available time may be allocated to performing additional tasks listed in the table.

The observed data on some projects show that plant technicians allocate time during the day to observe plant processes. Observations of processes may be made, particularly if there are concerns from test results. These recommendations provide the plant technician with an organized and standardized

TABLE 6.11 Contractor recommendations for daily inspection

Process	Location	Activity	Task	Estimated Time (Min)
Plant Technician				
Raw Materials	Aggregate Stockpiles	Stockpile Quality	Contamination	2
			Consistency	2
			Segregation	2
			Loading rotation	2
Plant Mixing	Aggregate Blending	Loading Practices	Bin contamination	2
			Bin levels	2
			Segregation	2
		Proportioning	Feeder rates	5
		Equipment Condition	Proper working condition	5
	Asphalt Binder Delivery	Binder Proportioning	Flow	5
		Binder Temperature	Temperature	5
		Equipment Condition	Proper working condition	5
	Mix Plant	Plant Operations	Mix discharge	5
		Equipment Condition	Proper working condition	5
	Storage	Storage Operations	Loading methods	5
			Storage duration	-
		Mix Quality	Mix temperature	5
Mix Transport	Load-Out	Loading Practices	Multiple drops	5
			Use tarps	-
		Truck Conditions	Clean	5
			Proper release agent	-
		Time Required to Complete Inspection Tasks		69
Density Technician				
Laydown	Base Condition	Surface Condition	Clean and uniform surface	5
		Tack Coat	Tack coat temperature	5
	Paving	Paving Operations	Truck to paver discharge	5
			Paver rate	-
			Paving inconsistencies	-
	Rolling	Rolling Operations	Mat Temperature	5
			Continuous Movement	As needed
			Maintain Rolling Pattern	
		Roller Conditions	Adequate Release Agent	
			Impact Spacing	
		Equipment Conditions	Proper working condition	5
		Time Required to Complete Inspection Tasks		25

list of tasks, which would allow for more systematic completion of this inspection. Density technicians may have time while waiting between tests. Non-productive time could be allocated to conducting brief visual inspection tasks related to the paving and rolling locations.

It is recommended that contractors use this table as a starting point to adjust additional plant and density technician inspection responsibilities according to contractor-specific or project-specific requirements.

6.4 SUMMARY

This chapter summarized study recommendations for both testing and inspection levels. Specific recommendations were given for quality control, quality acceptance, and verification testing levels. For inspection levels, recommendations were provided for important inspection tasks and those projects where a limited number of agency inspectors are available.

CHAPTER 7

IMPLEMENTATION

7.1 INTRODUCTION

This chapter provides guidelines and a case application of how the research results can be applied to actual projects in practice. Practical issues to consider when implementing the research results into practice are also provided.

7.2 GUIDELINES

Testing and inspection levels for any highway construction work evolve through decisions made by the agency during the life cycle of a project. The project life cycle can be viewed as consisting of four interdependent phases: (1) planning, (2) design, (3) construction, and (4) operation and maintenance, as shown in Figure 7.1. During each phase, decisions are made that influence testing and inspection levels, whether directly or indirectly. A systems approach was used that methodically applies a rational method for determining testing and inspection levels to satisfactorily construct HMA overlay projects.

7.2.1 Planning

A project begins at the planning stage. Fundamental planning components addressed for the highway facility during this stage include (1) current pavement condition, (2) forecasting future demands of the facility, and (3) performing an economic analysis of rehabilitation alternatives. This study is concerned with how these planning components can affect testing levels to satisfactorily construct HMA overlays.

7.2.1.1. Current Pavement Condition

Measuring the current pavement condition is accomplished by collecting field distress data and synthesizing the data to identify appropriate alternatives for rehabilitation or reconstruction. Many agencies have pavement management systems (PMS) to assist with data collection, evaluation, and decision-making during this process. The present serviceability index (PSI) is a common tool for quantifying information concerning the serviceability of the pavement. A primary factor used in establishing the PSI is the roughness of the surface profile. The PSI can also include standard distress criteria such as rutting, fatigue cracking, and thermal cracking. The Strategic Highway Research Program (SHRP) has adopted these criteria as consensus properties for establishing pavement performance.

Testing and inspection levels are important to this process because key distress criteria describing current pavement condition can be compared with as-built construction data. The PMS should attempt to establish relationships between design parameters, field construction data, and actual pavement performance. By establishing a certain level of testing and inspection, agencies are assured that data will be available to understand actual performance achieved from design and construction inputs. These data become vital to the PMS and provide a continuous feedback system that allows for improved designs and construction methods. They define reasonable targets and allowable variation for tests known to have a significant influence in constructing satisfactorily performing pavements.

7.2.1.2. Forecasting Demands

Forecasting traffic demands on the highway facility is another primary function of planning. Projected service levels of the highway facility are found using forecasted traffic volume and highway capacity information. Quantifying future traffic volumes can be described with Equivalent Single-Axle Loads (ESAL) or AADT. How an agency incorporates these data into the planning stage is unique to each project.

Forecasting the demands on the pavement has an impact on testing and inspection levels. HMA overlays serving a short-term life (e.g.,1 or 2 years) may not have the same level of testing and inspection as an overlay serving a longer service life (e.g., 10, 15, or 20 years).

7.2.1.3. Life-Cycle Cost Analysis

An emerging method for planning pavement rehabilitation or reconstruction using economic analysis is known as Life-Cycle Cost Analysis (LCCA). LCCA is a process for evaluating the total economic worth of a usable project segment by analyzing initial costs and discounted future costs (e.g.,

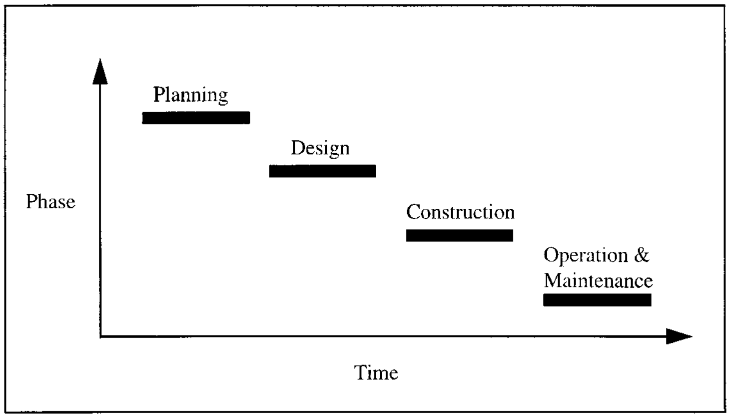


Figure 7.1. Four basic project phases.

maintenance, reconstruction, rehabilitation, restoring, and resurfacing costs) over the life of the project segment *(29)*. LCCA analysis periods should be sufficiently long to reflect long-term cost differences associated with reasonable design strategies. In general, the analysis period should be longer than the pavement design period, except in cases of long-lived pavements. As a rule of thumb, the analysis period should be long enough to incorporate at least one rehabilitation activity. The FHWA's Final LCCA policy statement recommends an analysis period of at least 35 years for all pavement projects, including new or total reconstruction projects, as well as rehabilitation, restoration, and resurfacing projects *(29)*.

LCCA can directly influence testing and inspection levels. Overlays that have a higher economic value that must meet a certain life period will require more stringent construction requirements. A means to measure and enforce these requirements is through testing and inspection and QA specifications. Results must reflect the impact of quality through pay factors. More severe pay factors may require increased testing levels to ensure more reliable estimates.

Defining different life-cycle costs for various overlay components within a project can influence testing and inspection levels. Mainline overlays, which include the driving lane or passing lane for 4-lane roadways, typically require more testing and inspection than shoulder overlays or base patching. For density testing, mainline overlays usually require higher design targets than shoulders, thereby resulting in closer scrutiny of density test results.

7.2.2 Design

The second phase of a project is design. Project scope and objectives developed during the planning phase are translated to functional designs. Fundamental requirements of the facility are specified, design issues are described in detail, and both preliminary and final drafts are prepared. During the design phase, three primary components influence testing levels: (1) contracting method, (2) pavement design, and (3) specification design.

7.2.2.1 Contracting Method

The standard and emerging contracting methods available for highway construction include Methods, QA, Design/Build, Design/Build/Operate, Warranty, and Multi-Parameter. Each of these methods has an influence on testing and inspection levels and the degree of agency involvement in the design and construction phases. This study was concerned with QA contracting, and the testing and inspection levels for agencies and contractors necessary to satisfactorily construct HMA overlays.

A potential misconception is believing that the ability to satisfactorily construct an HMA is a function of the contracting method. The potential surely exists with new contracting methods, such as warranties, but quality improvement must be an integral component of the contractor's way of doing business. As many practitioners and researchers would agree, quality in itself cannot be tested into the product—quality must be built into the product. When QA contracting is chosen over a warranty, this does not guarantee that a satisfactory HMA overlay will be constructed. What these contracting methods do stipulate is the responsible party for the work during operation and maintenance. Contractor responsibilities under QA contracting typically last 1 year, while warranty responsibilities may last 3, 5, or 8 years. Current warranty durations represent only a portion of the operation and maintenance phase and do not extend across the entire life-cycle period. The responsibilities for a design/build/operate contract may last throughout the overlay life-cycle period and encompass design through operation and maintenance phases.

Under different contracting methods, the agency and contractor share different degrees of risk in operation and maintenance. The agency must ensure that the final product meets the expectations and life-cycle period of the operation and maintenance phase, as set forth during the planning phase. With new warranty contracting methods, agency risk has been shifted to the contractor. The contractor's risk has been increased only during a specified period of operation and maintenance, while agency risk remains for the latter half of the phase.

The important question for QA contracting is how much testing and inspection is needed by both the agency and contractor during construction to ensure that a satisfactory HMA overlay has been constructed. Agencies are employing QA contracting methods with the responsibility to ensure that quality has been constructed. In some cases, the expectation of quality construction may not be met, and the agency uses a penalty to cover expected losses, usually in the form of contractor pay adjustment. On-site agency inspections document any deficient construction practices so that future disputes may be resolved quickly.

7.2.2.2 Pavement Design

A second component of the design phase that influences testing and inspection levels is pavement design. It is during this stage that structural requirements and loading criteria are specified. The design must satisfy technical engineering and economic considerations for alternate combinations of subbase, base, and surface material providing an adequate load-carrying capacity. In HMA overlays, designs for the base (i.e., milling or existing surface), lower lift, surface lift, and/or friction course require careful consideration.

7.2.2.2.1 Testing Levels. Mixture type is one pavement design factor that may directly affect testing levels. Those mixes specified for higher volume roadways, where life-cycle costs are greater, may require greater testing and inspection to ensure satisfactory construction. Wisconsin DOT specifies three primary types of mix types: (1) Low Volume (Type-LV), (2) Medium Volume (Type-MV), and (3) High Volume (Type-HV). Mean density target levels are increased for traffic volume. A higher mean may require increased testing and inspection to ensure that the target is met.

Temporary pavement designs may require different levels of testing than traditional overlay designs. Because of lower life-cycle costs, temporary pavements may not require the normal frequency of standard mix testing. In an effort to reduce construction costs, these pavements may not be required to satisfy the rigorous pavement base and mix designs normally specified on a project.

7.2.2.2.2 Inspection Levels. During the pavement design phase, it is possible to determine inspection levels with known project characteristics. Figure 7.2 provides guidelines to follow during pavement design.

An agency can begin by considering project AADT, project tonnage, and any other project-specific characteristics (Step #1). The standard specifications are also considered (Step #2), as well as the decision to use agency or contractor data for acceptance (Step #3). The choice of whether to use agency or contractor data for acceptance provides a decision point where one of two paths is followed. The paths include the same steps, but the actual implementation is different.

Step #4 determines what inspection tasks should be completed and how frequently. A staffing plan can be developed from these inputs, or the inspection tasks can be adjusted for the available staff. If, for example, contractor data are used for acceptance, more emphasis will be placed on inspecting contractor sampling and testing procedures than if agency data are used.

The guidelines are for normal production days and shown in one-half-day increments for ease of planning. It is also assumed that for field implementation, one-half-day allotments are practical from a scheduling point of view. There are certain project types that require less than 2 hr of inspection at the plant for a given day and are typically the low AADT or tonnage projects. For these projects, it is recommended that one inspector make a brief stop at the plant to conduct the required inspection tasks.

Tables 7.1 and 7.2 provide inspection-level guidelines based on three different AADT ranges combined with whether agency or contractor acceptance data are used. Table 7.1 recommends that the number of inspection tasks increase as AADT and tonnage increase. This table is for projects using agency acceptance data and, therefore, small increments of time are suggested for inspection tasks at the plant. Table 7.2 recommends longer increments of plant inspection for observation of contractor sampling and testing procedures, because contractor data are used for acceptance. It is recommended that the number of inspection tasks, staffing, and amount of inspection time increase with project AADT.

As these tables indicate, inspection levels increase as project AADT increases. Plant inspection is also emphasized more on projects where contractor test results are used for acceptance. These tables can be used by agencies as a starting point to determine the necessary staffing level for their specific project characteristics. Agencies should adjust inspection levels to accommodate their policies and project-specific situations.

The design stage would also be an appropriate time to create a QC plan for the project. The importance of QC plans can be supported with observed field data. Inspection tasks identified and completed by contractor and agency staff are summarized.

A QC plan increases the awareness among project staff, through communication and standardization. If project staff are told what is important and what to specifically watch for, they are more likely to complete that specific task. It is recommended that agencies and contractors make an effort to create QC plans. A sample QC plan from an observed project is provided in Appendix Q of the research team's final report.

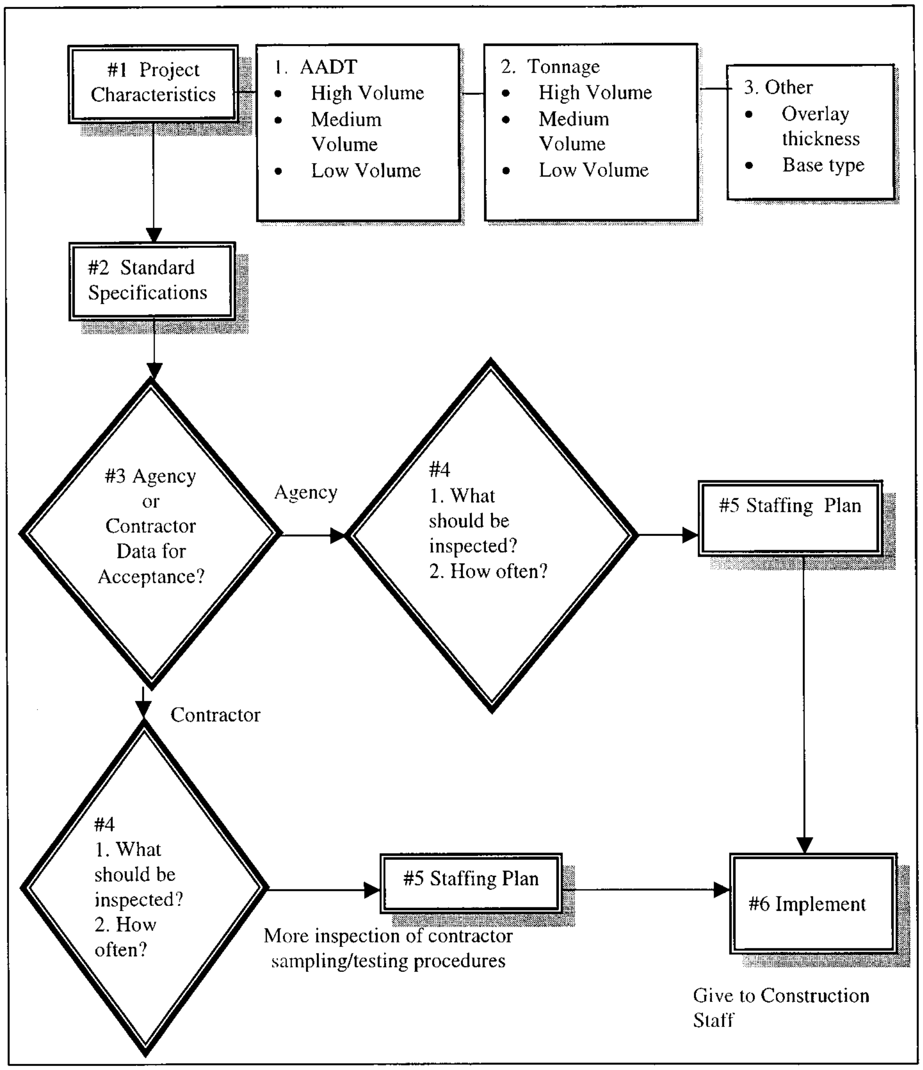


Figure 7.2. Agency guidelines for inspection-level implementation during design phase.

7.2.2.3 Specification Design

The specification component of the design phase is central to establishing testing levels in this study. First, plans and specifications are developed from design inputs, then the QA program is inserted into the project documents. It is during this development of the QA documents that both agency and industry are needed to successfully design a comprehensive QA testing program. There are four fundamental issues to evaluate for when developing the testing specification: (1) practical issues, (2) statistical issues, (3) economic issues, and (4) equity issues.

7.2.2.3.1 Practical Issues. The practical issues in HMA testing impose explicit constraints when establishing testing levels. Resources, particularly time and people, are practical constraints to any testing program. Available technology, governed largely by physical tests, has imposed time constraints on the ability to collect a large number of mix samples during a given production day. The maximum daily sample size for a complete mix property test ranged from 4 to 5 in this study, depending on available resources. The normality testing on the Arizona I-10 project, where 10 samples within a day were collected, required 2.5 days and 3 technicians to complete testing. Laboratory space can also limit the number of tests. In portable construction testing trailers, there may be space for only one or two technicians to perform testing procedures adequately.

The available staffing has imposed an additional constraint when determining appropriate testing levels. Many agencies have gone through rapid downsizing, thus limiting their ability to collect samples and complete the tests. This constraint has been difficult to overcome, and many agencies have shifted from using agency acceptance data to contractor QC data for acceptance.

TABLE 7.1 Agency inspection guidelines when agency data are used for acceptance

Process	Inspection Tasks AADT Low (<15,000)	Inspection Tasks AADT Medium (16,000-39,000)	Inspection Tasks AADT High (>40,000)
Raw Materials	Segregation Contamination Discoloration Loading Rotation	Segregation Contamination Discoloration Loading Rotation	Segregation Contamination Discoloration Loading Rotation
Plant Mixing	Bin Contamination Bin Levels Segregation Feeder Rates Storage Duration	Bin Contamination Bin Levels Segregation Feeder Rates Storage Duration Contractor Sampling/Testing Procedures	Bin Contamination Bin Levels Segregation Feeder Rates Storage Duration Loading Methods Contractor Sampling/Testing Procedures
Number of Inspectors Required at Plant	**1 for 30 minutes**	**1 for 30 minutes**	**1 for 60 minutes**
Mix Transport	Multiple Drops to Truck Tarp Use	Multiple Drops to Truck Tarp Use Clean Beds Proper Release Agents	Multiple Drops to Truck Tarp Use Clean Beds Proper Release Agents
Laydown	Clean, Uniform Surface Dry Surface Collect Truck Ticket Calculate Yield Slope Grade Width Depth Inconsistencies/ Segregation Maintain Rolling Pattern Surface Texture	Clean, Uniform Surface Dry Surface Collect Truck Ticket Mix Temperature Calculate Yield Slope Grade Width Depth Inconsistencies/ Segregation Maintain Rolling Pattern Joint Construction Continuous Roller Movement Surface Texture Contractor Density Testing	Clean, Uniform Surface Dry Surface Uniform Tack Coat Tack Coat Temperature Collect Truck Ticket Mix Temperature Calculate Yield Slope Grade Width Depth Inconsistencies/ Segregation Maintain Rolling Pattern Joint Construction Continuous Roller Movement Surface Texture Contractor Density Testing
Number of Inspectors Required at Laydown	**2 per day**	**2 per day**	**3 per day**

7.2.2.3.2 Statistical Issues. Statistical attributes of the specification are a fundamental concern when establishing testing levels. The previous analysis of specification compliance measures illustrated the different approaches for obtaining statistical estimates of the construction work. This study focused on the statistical procedures in the QA publications, namely the QLA method. States have also chosen the Absolute Average Deviation, Average, and Moving Average methods to determine specification compliance.

The practical aspects of a testing specification impose constraints on the sample size and the ability to obtain statistical estimates of the work. Because the number of samples is limited, attention must be placed on how the samples are grouped to create lots. The lot becomes a stand-alone declaration of the work, so decisions regarding the lot formulation are important.

7.2.2.3.3 Economic Issues. The economic aspects of testing costs must be considered during the development of QA testing specifications. Specifying certain levels of testing results in requirements for time and money, and careful thought must be given when establishing these levels. Particularly, the agency must be made aware of these costs, given that they are responsible for enforcing the specification and incur both agency and contractor costs. From interviews and field observations, QC testing is considered a strict cost item with little or no incentive for profit. This cost is passed directly to the agency during preparation of project estimates and final bids.

TABLE 7.2 Agency inspection guidelines when contractor data are used for acceptance

Process	Inspection Tasks AADT Low (<15,000)	 Medium (16,000-39,000)	 High (>40,000)
Raw Materials	Segregation Contamination Discoloration Loading Rotation	Segregation Contamination Discoloration Loading Rotation	Segregation Contamination Discoloration Loading Rotation
Plant Mixing	Bin Contamination Bin Levels Segregation Feeder Rates Storage Duration Contractor Sampling/Testing Procedures	Bin Contamination Bin Levels Segregation Feeder Rates Storage Duration Contractor Sampling/Testing Procedures	Bin Contamination Bin Levels Segregation Feeder Rates Storage Duration Loading Methods Contractor Sampling/Testing Procedures
Number of Inspectors Required at Plant	**1 for 60 minutes**	**1 for 60 minutes**	**1 for 4 hours**
Mix Transport	Multiple Drops to Truck Tarp Use	Multiple Drops to Truck Tarp Use Clean Beds Proper Release Agents	Multiple Drops to Truck Tarp Use Clean Beds Proper Release Agents
Laydown	Clean and Uniform Surface Dry Surface Collect Truck Ticket Calculate Yield Slope Grade Width Depth Inconsistencies/ Segregation Maintain Rolling Pattern Surface Texture	Clean and Uniform Surface Dry Surface Collect Truck Ticket Mix Temperature Calculate Yield Slope Grade Width Depth Inconsistencies/ Segregation Maintain Rolling Pattern Joint Construction Continuous Roller Movement Surface Texture Contractor Density Testing	Clean and Uniform Surface Dry Surface Uniform Tack Coat Tack Coat Temperature Collect Truck Ticket Mix Temperature Calculate Yield Slope Grade Width Depth Inconsistencies/ Segregation Maintain Rolling Pattern Joint Construction Continuous Roller Movement Surface Texture Contractor Density Testing
Number of Inspectors Required at Laydown	**2 per day**	**2 per day**	**1 for 4 hours** **2 per day**

7.2.2.3.4 Equity Issues. A primary goal when designing a QA testing specification is one that is fair and equitable to both the agency and contractor. Contractors are motivated to reduce the variance in the mean value of the tested property when bonus opportunities exist. Agencies desire a quality work product, typically described with tests as having minimal amount of variation and meeting the specified design targets. Thus, when establishing testing levels, variation and meeting target levels are goals shared by both the agency and contractor. These needs must be defined interchangeably with the statistical element of the specification. Key elements of an equitable specification include specification limits and pay factors.

Specification limits have evolved from years of construction experience, agency studies, and recommendations from AASHTO. Specification limits have a direct effect on determining the level of HMA quality and appropriate testing levels. If the specification limits are stringent, an appreciable number of tests will be outside the limits. With large specification limits, there would be a wide range of tests found within limits. Most QA specifications have developed limits considered equitable to, and achievable by, the contractor.

Pay factors are an important element of an equitable specification. In theory, pay factors should represent the difference in life-cycle costs from design and as-built construction. The pay factor translates to a reduced service life from the as-built and measured construction materials. It is likely that pay factors will remain at set levels until further data are available to warrant any changes. In addition, decisions are made whether to apply different pay factors to different classes of pavement mixes. For example, more testing may be specified for high traffic-level roadways.

7.2.3 Construction

Testing and inspection levels developed during the design phase are executed during the construction phase. Prior to construction, it is clearly defined which party is responsible for acceptance testing so that the highway can be constructed with minimal disruption. However, there is one primary issue that can influence the level of inspection, namely available staff. Another important issue during construction is documentation and feedback.

7.2.3.1 Available Staff

For a specific HMA overlay project, issues may arise that cannot be foreseen during the design phase. When considering inspection of HMA construction on a particular day, available staffing may not match the original staffing plan from the design stage. This results in revising the staffing plan, along with the inspection tasks that should be completed.

Figure 7.3 illustrates the steps an agency could follow to revise the original staffing plan. The number of available

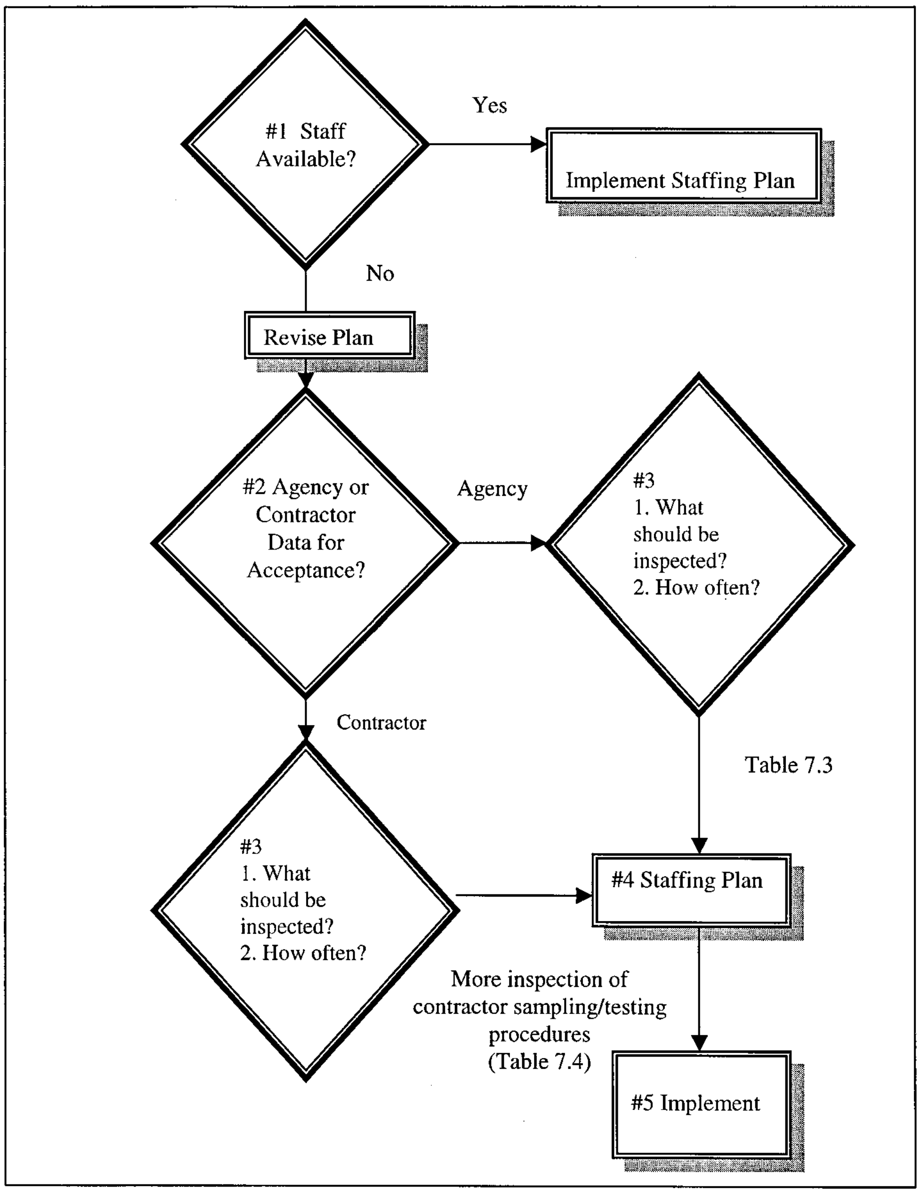


Figure 7.3. Agency inspection guidelines during construction.

staff should be considered first, then whether agency or contractor data are used for acceptance. A revised plan of what should be inspected and how often should then be determined and implemented.

Tables 7.3 and 7.4 provide guidelines that the agency may follow based on whether agency or contractor data are used for acceptance, respectively. Inspection time at the plant increases when contractor data are used for acceptance. Inspection tasks increase in number and frequency as the available project staff increases.

7.2.3.2 Documentation and Feedback

Documentation and feedback are critical during HMA construction. Feedback should be provided to understand how the testing and inspection plan are performing during actual construction. This will allow the agency and contractor to make appropriate adjustments using the considerations defined during development of the specification.

It is recommended that agencies and contractors create their checklists to standardize the data collection process. A

TABLE 7.3 Agency inspection guidelines when agency data are used for acceptance (available project staff)

	Inspection Tasks		
	Staff Available		
Process	1	2	3
Raw Materials	Contractor Sampling/ Testing Procedures Storage Duration	Segregation Contamination Discoloration Loading Rotation	Segregation Contamination Discoloration Loading Rotation
Plant Mixing	None	Bin Contamination Bin Levels Segregation Feeder Rates Contractor Sampling/Testing Procedures Storage Duration	Bin Contamination Bin Levels Segregation Feeder Rates Contractor Sampling/Testing Procedures Storage Duration Loading Method
Number of Inspectors Required at Plant	**1 as needed**	**1 for 30 minutes**	**1 for 60 minutes**
Mix Transport	Tarp Use	Tarp Use Multiple Drops to Truck Clean Beds Proper Release Agents	Tarp Use Multiple Drops to Truck Clean Beds Proper Release Agents Mix Temperature
Laydown	Clean, Uniform Surface Dry Surface Collect Truck Ticket Grade Width Depth Joint Construction Inconsistencies/ Segregation	Clean and Uniform Surface Dry Surface Tack Coat Temperature Collect Truck Ticket Calculate Yield Grade Width Depth Slope Joint Construction Inconsistencies/Segregation Surface Texture	Clean and Uniform Surface Dry Surface Tack Coat Temperature Uniform Tack Coat Collect Truck Ticket Mix Temperature Calculate Yield Grade Width Depth Slope Joint Construction Inconsistencies/Segregation Constant Paver Rate Surface Texture Density (monitor contractor) Maintain Rolling Pattern Continuous Roller Movement
Number of Inspectors Required at Laydown	**1 per day**	**2 per day**	**3 per day**

TABLE 7.4 Agency inspection guidelines when contractor data are used for acceptance (available project staff)

	Inspection Tasks		
	Staff Available		
Process	1	2	3
Raw Materials	None	Segregation Contamination Discoloration Loading Rotation	Segregation Contamination Discoloration Loading Rotation
Plant Mixing	Contractor Sampling/ Testing Procedures Storage Duration	Bin Contamination Bin Levels Segregation Feeder Rates Contractor Sampling/ Testing Procedures Storage Duration	Bin Contamination Bin Levels Segregation Feeder Rates Contractor Sampling/ Testing Procedures Storage Duration Loading Method
Number of Inspectors Required at Plant	**1 as needed**	**1 for 4 hours**	**1 for 4 hours**
Mix Transport	Tarp Use	Tarp Use Multiple Drops to Truck Clean Beds Proper Release Agents	Tarp Use Multiple Drops to Truck Clean Beds Proper Release Agents Mix Temperature
Laydown	Clean,Uniform Surface Dry Surface Collect Truck Ticket Grade Width Depth Joint Construction Inconsistencies/ Segregation	Clean, Uniform Surface Dry Surface Uniform Tack Coat Collect Truck Ticket Calculate Yield Grade Width Depth Slope Joint Construction Inconsistencies/Segregation Surface Texture	Clean, Uniform Surface Dry Surface Tack Coat Temperature Uniform Tack Coat Collect Truck Ticket Mix Temperature Calculate Yield Grade Width Depth Slope Joint Construction Inconsistencies/Segregation Constant Paver Rate Surface Texture Density (monitor contractor) Maintain Rolling Pattern Continuous Roller Movement
Number of Inspectors Required at Laydown	**1 per day**	**1 for half day** | **1 per day**	**1 for half day** | **2 per day**

sample checklist for plant inspection is provided in Appendix R of the research team's final report. Similar checklists can be modified for all inspection levels. The *Hot Mix Asphalt Construction Participant Manual* provides a sample checklist for the inspection of base conditions. Checklists are also helpful for conducting inspections of equipment conditions. Sample checklists for inspection of base conditions and equipment maintenance are also provided in Appendix R of the research team's final report *(13)*.

Daily diaries provide the agency inspector with the means to record daily events. If early pavement failure or a dispute between the agency and contractor occur, such a diary would be valuable in resolving any questions concerning a particular day of paving. For this reason, it is recommended that both the contractor and agency field personnel complete daily diaries. Appendix S of the research team's final report contains a sample daily diary.

7.2.4 Operation and Maintenance

The final phase of the project is the operation and maintenance phase (O&M). This phase could also be considered the performance period, where the project must perform as a functional facility to meet the needs and expectations of the end-user.

When developing a rational method for determining testing and inspection levels to satisfactorily construct an HMA overlay, tests and observations collected during the construction phase should be used in conjunction with pavement performance data. Through the PMS or some other type of data analysis, the construction data can be compared with in situ pavement performance. For example, in the context of design for pavement density, the mean and standard deviation of plant-produced air voids and density test results can be compared with in situ rut depths. A feedback system would then be developed to determine which tests, and testing levels, are necessary to satisfactorily construction an HMA overlay.

Many states are beginning Long-Term Pavement Performance (LTPP) monitoring programs that collect field performance data, as well as climate, maintenance, and traffic data. For those projects not under the formal LTPP program, a wealth of construction data are available from QA projects. These data can be combined with the PMS and used throughout the continuous life cycle of projects.

The Pareto analysis of inspection tasks completed and the factors that affect those inspection tasks can be considered when examining final pavement quality. It was found when evaluating observed field data that most inspection tasks are performed during the laydown process for the agency, and during plant-mixing and laydown processes for the contractor. Factors affecting the inspection tasks during these processes are primarily personnel-related for the agency and contractor. The level of these factors from inspection observation can be compared with in situ performance data to understand their significance for future changes and improvements.

7.2.5 Summary of System Concept for Determining Testing Levels

The systems approach for determining testing and inspection levels provides a holistic viewpoint of different phases of a project, and components of each phase, that may influence testing and inspection levels. The conceptual representation of the project phases can be modified to include the components within each phase. Figure 7.4 illustrates the phase components within the testing system.

7.3 CASE APPLICATION

A case application was provided that leads an agency through the steps to establish appropriate testing and inspection levels for an actual project. Data from the Minnesota I-90 project are used for this example application. Estimated values were provided when there were no actual data from this project (or state).

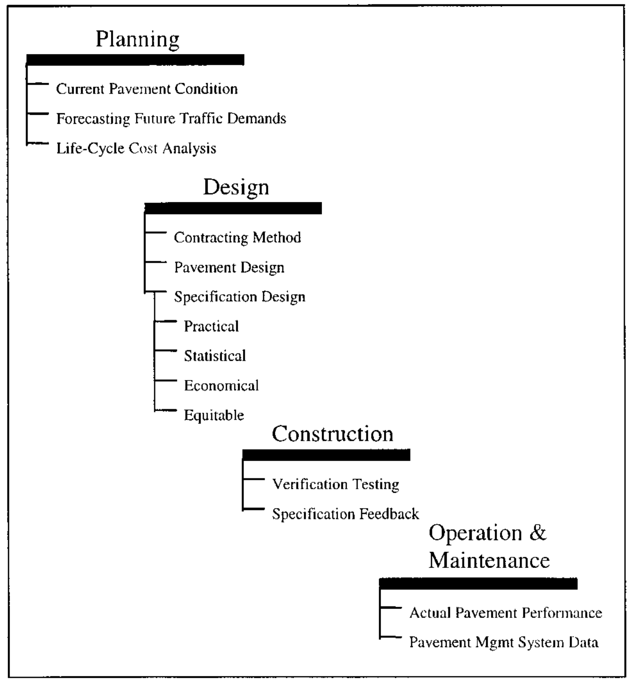


Figure 7.4. System concept for testing and inspection levels.

7.3.1 Testing Levels

7.3.1.1 Sampling Frequency

Determining the sampling frequency can be accomplished using the steps outlined in Figure 7.5

Step 1. A thorough consideration is given to the sampling attributes of hot-mix samples in Table 4.2. These attributes include material characteristics, resources, interaction of construction operations and sampling functions, and safety. It is decided to collect hot-mix samples behind the paver.

Step 2. The minimum time requirements to perform typical HMA tests are collected from meetings with agency district representatives and contractors at technical meetings. Figure 4.4 provides average time requirements and number of possible tests during a 10-hr work day. These times are based on current testing procedures and must be adjusted to the unique consideration of each individual state.

Step 3. The agency and contractor choose quantity-based sampling after considering the attributes in Table 4.5.

Step 4. Production data are collected from several contractors throughout the state to analyze current production rates and develop a rational sampling frequency. Table 7.5 provides the calculations of sampling frequency using the minimum testing time requirements and col-

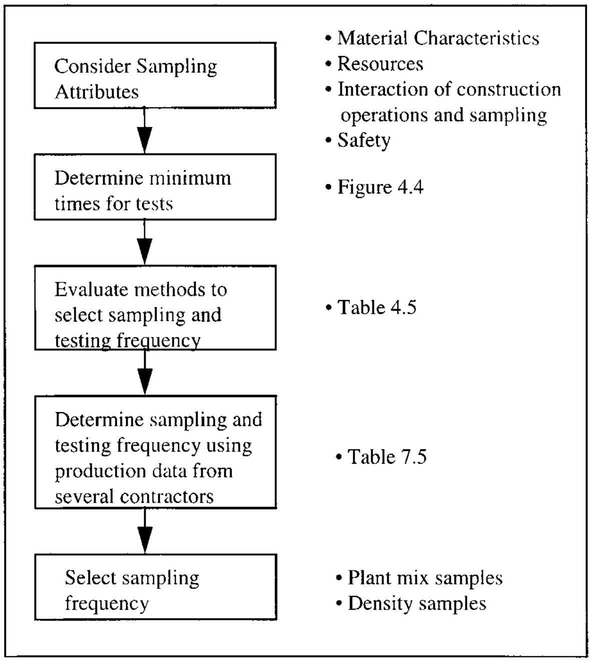


Figure 7.5. Determining sampling frequency.

lected production rates (estimated production rates were used). Sample projects A and B represent construction by mid-sized contractors, Project C represents a smaller contractor at a lower production rate, and Projects D and E represent larger contractors.

Step 5. Sampling frequencies are selected on the basis of an evaluation of state-specific production data. To select a sampling frequency that is fair to both size contractors, sampling frequencies for hot-mix samples would be 1 per 1,000 tons and density samples would be every 1 per 200 tons. The density sampling frequency may later be changed to a length or area consistent with the tonnage. It may be plausible that the sampling frequencies be slightly increased for more critical projects having higher life-cycle costs. For example, the hot-mix sampling frequency would be 1 per 800 tons, and density would be 1 per 150 tons.

7.3.1.2 Quality Control

7.3.1.2.1 Aggregate Gradation (Raw Materials). Figure 7.6 outlines the process to determine sampling location for aggregate gradation quality control.

Step 1. Testing levels of aggregate gradation (raw materials) are determined independently of acceptance testing, assuming that hot-mix aggregate gradation will be used for acceptance. The contractor chooses to perform coldfeed gradation tests to provide an indication of the final output gradation for QC purposes only.

Step 2. The contractor collects several samples at the start of production to determine if the coldfeed aggregate gradation will provide a prediction of hot-mix aggregate gradation. Table 7.6 provides gradation data from the first 3 days of production on the project.

Step 3. The independent-sample *t*-test is used to compare the mean of the two sampling locations. The *F*-test is used to compare variation. Those *p*-values greater than 0.10 would indicate no statistical difference. Figure 7.7 provides the calculations for the *t*-test and *F*-test.

Step 4. The *t*-tests concluded no significant difference between the mean and standard deviation, except for the 75-µm sieve. The contractor should run coldfeed testing with reasonable confidence for the larger sieves to provide an indication of final hot-mix aggregate gradation. The results of coldfeed tests for the 75-µm sieve will not produce a reliable prediction of the final hot-mix aggregate gradation tests since a mean difference exists in both the mean and standard deviation.

The *F*-test concluded that the 75-µm sieve will provide a different amount of variation between coldfeed and hot-mix samples. There is also marginal evidence that the 12.5-mm and 4.75-mm sieves will provide a different estimate of variation.

7.3.1.2.2 Plant Mixing. Figure 7.8 provides the steps to determine QC testing levels for plant mixing.

TABLE 7.5 Calculation of suggested sampling frequencies

Name of Project	Daily Total Tonnage Min	Avg	Max	Maximum Production tons per hour	Plant Tests Test time, hours	Sampling frequency, tons	Density (cores) Test time, hours	Sampling frequency, tons
Project A	200	1,500	3,600	360	3	1,080	0.5	180
Project B	400	1,100	2,500	250	3	750	0.5	125
Project C	250	1,300	1,750	175	3	525	0.5	88
Project D	360	3,500	4,500	450	3	1,350	0.5	225
Project E	155	2,100	4,200	420	3	1,260	0.5	210

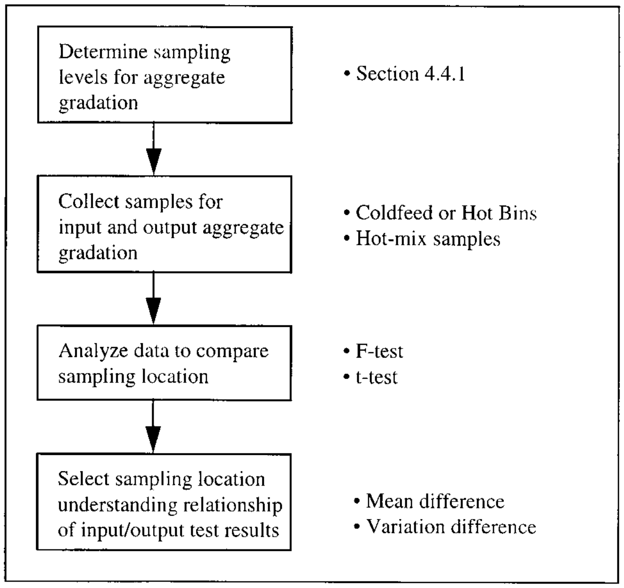


Figure 7.6. Flow diagram of steps to determine sampling location of aggregate gradation.

Step 1. Use the sampling frequency calculated earlier.

Step 2. Collect historical data from other projects for those plant-mix properties that will be used during quality control.

Step 3. Conduct an analysis of variance to determine if there is significant between-day variation in the historical data. Use the models developed in Chapter 4. Work with consultants within or outside the contractor's organization during this effort.

Apply data from other components of the project to understand where the variation is being produced. Examples include weather effects (e.g., moisture on stockpiles and removal of fines from rainfall), loading practices (e.g., operator working a different face at the stockpiles), belt speeds (e.g., oscillation, bearing wear, and proportioning), and asphalt binder metering (e.g., pump condition and pressure surge).

Step 4. If there is variation between days, allocate testing to each day of production to ensure that there is daily testing during production, rather than high-frequency testing on periodic days when the technician is available. There is most likely a full-time plant technician available for the project.

Step 5. Plot the test data on control charts to visually understand the process.

7.3.1.2.3 Density. Figure 7.9 outlines the steps to determine density testing levels for quality control.

Step 1. Use the sampling frequency calculated earlier.

Step 2. Collect historical density data from other projects.

Step 3. Conduct an analysis of variance to determine if there is significant between-day variation in the historical data. Use the developed models in Chapter 4. Work with consultants within or outside the company during this effort.

Step 4. Apply data from other components of the compaction process to understand where the variation is being produced. Examples include changes in mix properties (e.g., resulting compaction from air voids and asphalt content within a particular sublot), compaction methods (e.g., 3-pass, 5-pass, or 7-pass), and environmental conditions (e.g., cooling rates and mat thickness).

Step 5. The contractor should assemble production data from paving projects to determine current process capability. Table 7.7 provides density data from the Minnesota

TABLE 7.6 Aggregate gradation data from initial production

	Sample	Test	Aggregate Gradation Percent Passing, %				
Date	Location	Number	12.5mm	95mm	475mm	236mm	75µm
7/30	coldfeed	1	92	86	60	37	2.6
7/30	coldfeed	2	87	76	46	26	2.5
7/31	coldfeed	3	84	72	44	27	2.8
8/1	coldfeed	4	81	74	43	25	2.5
7/30	extraction	50	90	84	57	35	2.3
7/30	extraction	51	90	81	52	29	3.6
7/30	extraction	52	85	76	46	25	2.6
7/31	extraction	53	87	78	48	26	3.7
7/31	extraction	54	88	80	48	26	3.1
7/31	extraction	55	83	71	43	25	3.3
8/1	extraction	56	87	78	46	26	3.3
8/1	extraction	57	87	77	48	27	3.8
8/1	extraction	58	85	75	49	26	2.9

12.5mm Sieve

Location	n	Mean	Std Dev
Coldfeed	4	86.00	4.69
Extraction	9	86.88	2.31

Variances	T	DF	Prob>|T|
Unequal	-0.3600	3.7	0.7389
Equal	-0.4701	11.0	0.6474

Equal variances F=4.10 DF=3,8 Prob>F=0.0979

9.5mm Sieve

Location	n	Mean	Std Dev
Coldfeed	4	77.00	6.22
Extraction	9	77.78	3.73

Variances	T	DF	Prob>|T|
Unequal	-0.2322	4.0	0.8277
Equal	-0.2846	11.0	0.7813

Equal variances F=2.77 DF=3,8 Prob>F=0.2212

4.75mm Sieve

Location	n	Mean	Std Dev
Coldfeed	4	48.25	7.93
Extraction	9	48.56	4.00

Variances	T	DF	Prob>|T|
Unequal	-0.0730	3.7	0.9456
Equal	-0.0947	11.0	0.9262

Equal variances F=3.93 DF=3,8 Prob>F=0.1083

2.36mm Sieve

Location	n	Mean	Std Dev
Coldfeed	4	28.75	5.56
Extraction	9	27.22	3.15

Variances	T	DF	Prob>|T|
Unequal	0.5140	3.9	0.6353
Equal	0.6424	11.0	0.5338

Equal variances F=3.11 DF=3,8 Prob>F=0.1773

75µm Sieve

Location	n	Mean	Std Dev
Coldfeed	4	2.60	0.14
Extraction	9	3.18	0.51

Variances	T	DF	Prob>|T|
Unequal	-3.1544	10.2	0.0100
Equal	-2.1924	11.0	0.0508

Equal variances F=12.85 DF=8,3 Prob>F=0.0596

Figure 7.7. Statistical calculations for comparing sampling locations.

I-90 project. Process data indicate that the standard deviation of density ranges from 0.82 percent to 1.98 percent for 1 day's production. A median value of 1.5 percent is chosen from this table.

The starting point for determining a minimum level of daily density testing for quality control is Equation 4.4. The selected probability level is set at 95 percent ($Z_{\alpha/2} = 1.96$). Table 7.8 provides the relationship between the confidence limits for the mean and number of density tests.

The contractor could select nine density tests per day and have confidence limits of ±0.98 percent. Thus, if a contractor estimates a mean density of 92 percent, the mean can be found anywhere from 91.02 percent to 92.98 percent at a confidence level of 95 percent. For more critical

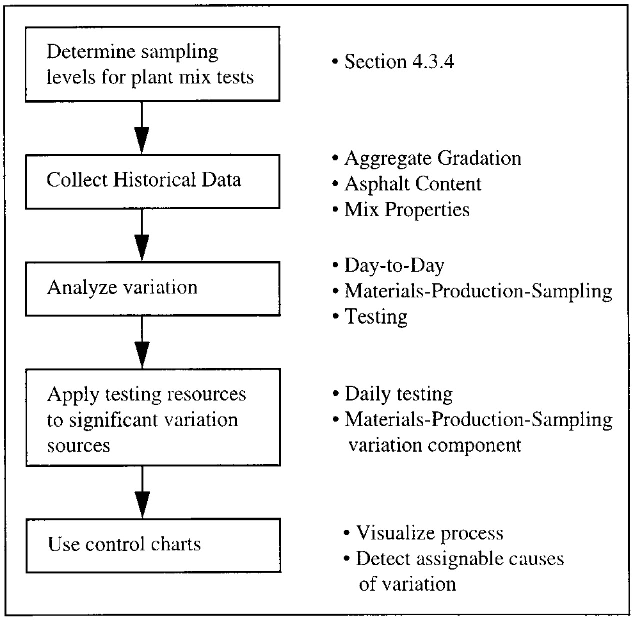


Figure 7.8. Flow diagram of steps for plant mixing.

projects, where the target density level must be met, a great number of density tests should be taken.

Step 6. A rational way of obtaining unbiased density samples is defining the rolling zone as a stratified sublot. Density test sites would then be randomly chosen within each rolling zone. Unbiased estimates for the mean and standard deviation for a given day of production would then be obtained by combining the individual rolling zone tests into a single dataset.

7.3.1.3 Acceptance

Figure 7.10 provides the necessary steps for determining acceptance levels. This case application is only applied for density. However, the same steps would be followed for plant-mix properties.

Step 1. The sampling frequencies for plant-mix samples were determined to be 1 per 1,000 tons and density samples were 1 per 200 tons. Higher frequencies may be possible at 1 per 800 tons (mix) and 1 per 150 tons (density) for more critical projects.

Step 2. Determine if the specification limits are achievable by the contractor. The Minnesota specification limits for hot-mix properties were designed for the moving average, while the average method was used for density. The range of the moving average is always less than the distribution of individual values, so this analysis is confounded when evaluating the specification limits solely with the distribution of individual values. For the purposes of the analysis,

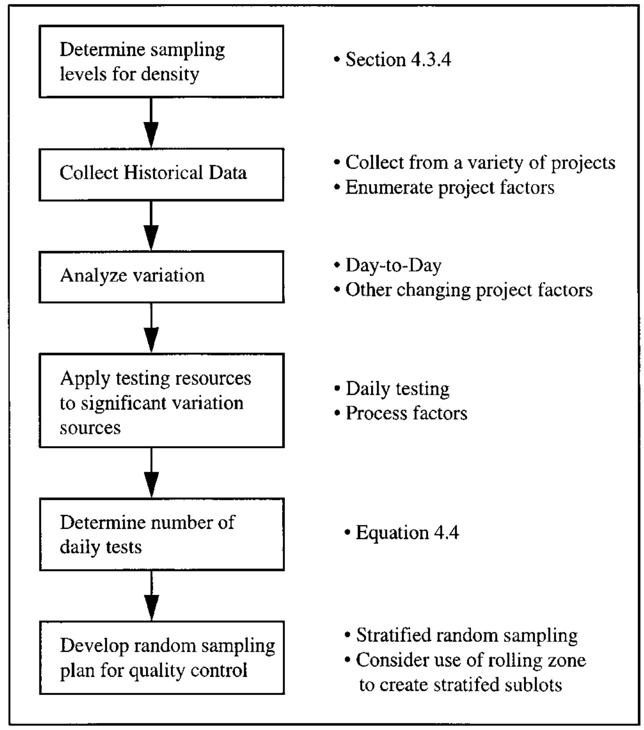


Figure 7.9. Flow diagram of steps to determine density testing levels.

it is presumed that the limits are defined for a QLA-type specification that measures the percentage of material within limits.

Construct 90-percent confidence limits for the production range and compare this range with the actual project specification limits, as shown in Table 7.9. Use Equation

TABLE 7.7 Process data for density, measured as a percent of TMD

Date	Number	Mean	Std. Dev.
709	6	89.6	1.23
711	4	92.2	1.66
715	4	89.7	1.28
717	4	89.0	1.75
718	6	91.1	1.56
721	6	90.0	1.72
722	6	89.4	1.33
723	2	91.3	1.98
725	8	90.0	1.00
730	4	91.4	0.94
731	10	92.5	1.75
801	10	92.3	1.25
802	10	92.6	1.04
806	8	92.2	1.64
807	10	91.7	1.11
808	10	91.6	0.82
809	9	91.7	1.01
811	10	91.0	1.22
812	8	90.9	1.38
815	8	90.3	1.55

TABLE 7.8 Number of daily density QC Tests for 95-percent confidence limits and $\sigma_{Density}$ = 1.5 percent

Number of Daily Tests	Limit (+/-) for Mean Density, measured as a % TMD
5	1.31
6	1.20
7	1.11
8	1.04
9	0.98
10	0.93
11	0.89
12	0.85
13	0.82
14	0.79
15	0.76
16	0.74
17	0.71
18	0.69
19	0.67
20	0.66

4.5 to calculate the limits. Use 90-percent confidence limits because it is desired, at a minimum, that the contractor achieve PWL equalling 90 and receive 100-percent payment.

The data indicate that the I-494 contractor had greater relative variability, as expressed by the standard deviation and coefficient of variation. Production exceeded specification limits for all coarse sieve sizes and volumetric tests. The I-90 contractor had lower variability. Based on a limited number of projects, it appears that the specifications can be achieved by the contractor.

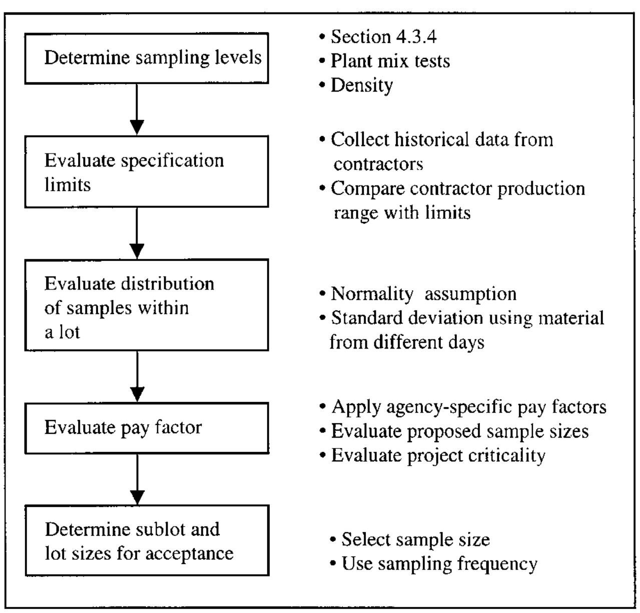


Figure 7.10. Flow diagram for determining acceptance testing.

Step 3. Evaluate the distribution of samples within a lot. Conduct an analysis of normality to confirm whether the data are normally distributed. Also, determine if the standard deviation increases when material is combined for adjoining days.

Step 4. Compare the effect of the specification limits and proposed sample sizes on the pay factor. For purposes of this study, the analysis from Section 4.5.3 (pay factors) should be followed.

Step 5. Determine the appropriate sample size for acceptance. It is in the interest of the contractor to receive larger payment at the AQL, and for the agency to have minimal risk at the RQL. It may be decided that the n equal to 10 sample size is more appropriate based on evaluation of risks from Section 4.5.3. Based on this selection, the lot size is determined. The sampling frequency for density samples was 1 per 200 tons (as determined earlier). Multiply the number of samples (n equal to 10) by the sampling frequency (1 per 200 tons) to determine the lot size (2,000 tons).

7.3.1.4 Verification

Figure 7.11 provides the steps necessary to determine verification testing levels.

Step 1. Collect split-sample test data from a range of projects. The mean and standard deviation of differences are calculated from the projects, and a median value is chosen. Table 7.10 provides split-sample data from two Minnesota projects.

Step 2. Equation 4.6 is applied using the standard deviation and number of verification tests to yield a testing tolerance at each sample size. Recommended values are as follows: $Z_{\alpha/2}$ equal to 1.96 (95-percent probability) and Z_{β} equal to 0.842 (80-percent power). Table 7.11 provides a practical application of Equation 4.6 using these input variables.

Interpretation of the table is demonstrated by the following example. Suppose the agency wants to verify one of the contractor's four acceptance tests from a given lot for asphalt content. There would be an 80-percent chance of detecting a true difference of 0.6 percent between tests. If the agency and contractor had a difference of 0.3 percent, the probability of detecting a true difference would be much less (30 percent to 40 percent). The ability to discriminate differences is lost as the tests move closer together.

TABLE 7.9 Comparison of production range and specification limits

Project	n	Production: Mean %	Std Dev %	C.V. %	90% Production Range: Lower	Upper	Range	Specification: Target %	Specification Limits, %: Lower	Upper	Range	Production greater than Spec Limits
12.5mm Sieve												
I-90	35	87.0	2.26	2.6	83.3	90.7	7.4	84.0	79.5	88.5	9.0	No
I-494	35	89.0	2.07	2.3	85.6	92.4	6.8	89.0	84.0	90.0	6.0	Yes
9.5mm Sieve												
I-90	36	77.3	3.52	4.6	71.5	83.1	11.6	76.0	71.5	80.5	9.0	Yes
I-494	35	74.7	3.68	4.9	68.6	80.8	12.1	79.0	74.0	84.0	10.0	Yes
4.75mm Sieve												
I-90	36	48.0	2.73	5.7	43.5	52.5	9.0	43.0	38.5	47.5	9.0	No
I-494	35	37.0	4.58	12.4	29.5	44.5	15.1	43.0	38.0	48.0	10.0	Yes
2.36mm Sieve												
I-90	36	27.2	1.81	6.7	24.2	30.2	6.0	27.0	23.0	32.0	9.0	No
I-494	35	26.3	3.74	14.2	20.1	32.5	12.3	29.0	25.0	33.0	8.0	Yes
75 m Sieve												
I-90	36	2.9	0.51	17.6	2.1	3.7	1.7	2.9	1.4	4.4	3.0	No
I-494	35	4.6	0.68	14.8	3.5	5.7	2.2	4.0	2.0	6.0	4.0	No
Asph. Cont.												
I-90	36	5.20	0.28	5.38	4.74	5.66	0.92	5.5	5.1	5.9	0.8	Yes
I-494	35	4.98	0.19	3.82	4.67	5.29	0.63	4.8	4.4	5.2	0.8	No
Air Voids												
I-90	36	3.5	0.73	20.9	2.3	4.7	2.4	5.0	3.7	6.3	2.6	No
I-494	35	3.6	1.09	30.3	1.8	5.4	3.6	4.0	2.7	5.3	2.6	Yes
VMA												
I-90	36	13.4	0.62	4.6	12.4	14.4	2.0	14.8	13 min	n/a	n/a	n/a
I-494	35	13.8	1.04	7.5	12.1	15.5	3.4	13.0	13 min	n/a	n/a	n/a
VFA												
I-90	36	74.1	4.78	6.5	66.2	82.0	15.7	66.3	65.0	75.0	10.0	Yes
I-494	35	74.4	6.27	8.4	64.1	84.7	20.6	n/a	65.0	75.0	10.0	Yes
Density												
I-90	44	91.3	1.94	2.1	88.1	94.5	6.4	n/a	92 min	n/a	n/a	n/a
I-94 density data unavailable												

Develop a testing tolerance table using the values for n equal to 1 in Table 7.11. A testing tolerance table would be incorporated into the specifications as shown in Table 7.12.

Step 3. A variance test procedure is applied during project start-up to determine if the project-specific split-sample variance is equal to the population split-sample variance. Figure 7.12 provides an example of applying this procedure to actual project data.

Step a. Collect four or more split-sample test results and calculate the difference of each.

Step b. Determine the median value of both the split-sample data set and the historical database.

Step c. Calculate the absolute value of all deviations from the median for both the new split-sample data set and historical database. In other words, for each value, subtract the median, and take the absolute value of the result. If, in either sample, there is an odd number of observations, delete exactly one value of “0” from the list of absolute values of deviations. Both sample sets have an even number of observations, so no observations are deleted.

Step d. Perform a test for comparing the means of the split-sample data set and historical database using the independent-sample *t*-test with variances assumed equal. This output concludes that there is no mean statistical

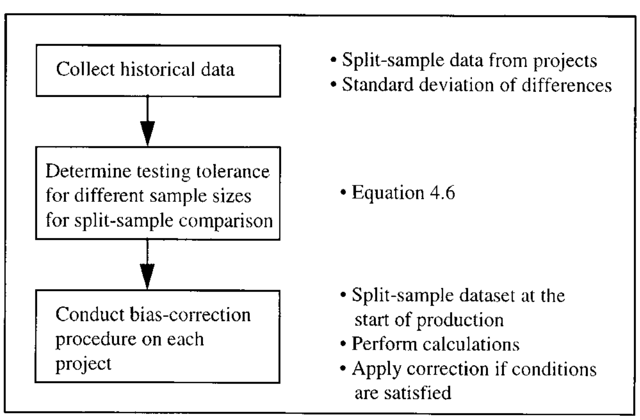


Figure 7.11. Flow diagram for determining verification testing levels.

TABLE 7.10 Split-sample project data, measured as a percent of TMD

			Standard Deviation of Difference, %									
			Aggregate Gradation									
Project	Comparison	n	19.0mm, %	9.5mm, %	2.36mm, %	75um, %	AC, %	Gmb, sp.gr.	Gmm, sp.gr.	Voids, %	VMA, %	VFA, %
(a) Mean difference between split-samples												
Minnesota I-90	QC/QA	26	n/a	1.7	0.2	-0.9	-0.01	0.004	-0.003	-0.3	-0.2	1.6
Minnesota I-494	QC/QA	34	n/a	-0.5	-0.5	-0.7	0.11	-0.003	-0.002	0.0	0.2	0.2
	QC/AI	24	n/a	0.7	0.1	-0.1	-0.24	0.000	0.000	0.0	-0.2	-0.6
	QA/AI	24	n/a	1.2	0.4	0.6	-0.32	0.002	0.002	0.0	-0.4	-0.8
(b) Standard deviation of difference between split-samples												
Minnesota I-90	QC/QA	26	n/a	4.99	1.96	0.80	0.32	0.013	0.006	0.55	0.43	3.68
Minnesota I-494	QC/QA	34	n/a	2.24	1.26	0.33	0.14	0.013	0.013	0.80	0.48	5.17
	QC/AI	24	n/a	3.51	1.65	0.19	0.17	0.012	0.011	0.67	0.42	4.42
	QA/AI	24	n/a	2.98	1.18	0.35	0.12	0.012	0.009	0.67	0.45	3.89
QC/QA = Contractor and Agency Split QC/AI = Contractor and Asphalt Institute Split QA/AI = Agency and Asphalt Institute Split												

difference between the existing database and sample data set.

7.3.2 Inspection Levels

The Minnesota I-90 project was chosen as a case study because contractor data were used for acceptance. It was selected, in part, because many agencies are beginning to specify contractor data for acceptance. Figure 7.13 shows how the determination of a minimum inspection level could be achieved for this project.

The key project characteristics for Minnesota I-90 were project AADT of 13,810 and project tonnage of 36,000. These numbers are within the low ranges for both AADT and tonnage. Contractor data were used for acceptance testing, so more observations of contractor sampling and testing procedures were conducted.

Inspection tasks and frequencies were determined earlier and are provided in the recommendations. For this case study, it was determined that 1 hr should be allocated to inspection tasks at the plant by one inspector. The remaining portion of this inspector's day should be spent at laydown, along with an additional inspector who spends the entire day at laydown. Summaries of "what to inspect" and "how often" are provided in Table 7.13, and the resulting staffing plan in Table 7.14.

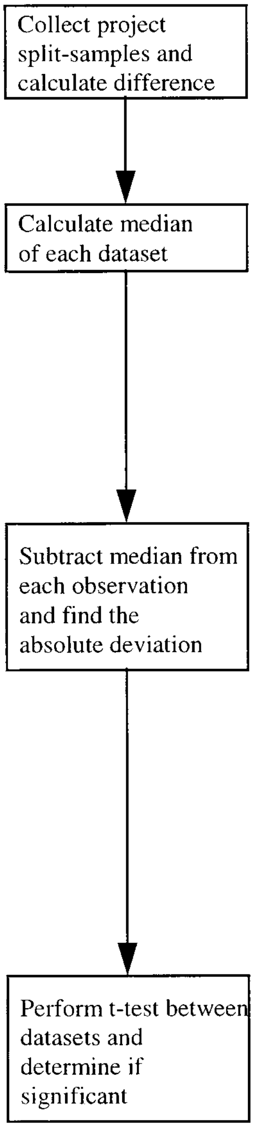
The suggested staffing plan shown is different from what was actually observed for this project. During site observation, two inspectors were observed completing inspection tasks at laydown operations throughout the day. An additional inspector was observing contractor sampling and testing procedures at the plant for an entire day, with a portion of the day spent conducting aggregate gradation tests and transporting samples to the agency laboratory for testing. This example shows where this agency might be able to minimize inspection levels. This case study is an example of how

TABLE 7.11 Number of verification tests and testing tolerances

Test	Standard Deviation between tests[a]	Testing Tolerances, +/-, for a select number of tests			
		n=1	n=2	n=3	n=4
19.0mm, %	3.0	8.4	5.9	4.9	4.2
9.5mm, %	3.0	8.4	5.9	4.9	4.2
2.36mm, %	2.0	5.6	4.0	3.2	2.8
75 m, %	0.8	2.2	1.6	1.3	1.1
AC, %	0.2	0.6	0.4	0.3	0.3
Gmb, sp.gr.	0.013	0.036	0.026	0.021	0.018
Gmm, sp.gr.	0.010	0.028	0.020	0.016	0.014
Voids, %	0.6	1.7	1.2	1.0	0.8
VMA, %	0.4	1.1	0.8	0.6	0.6
VFA, %	2.9	8.1	5.7	4.7	4.0
[a]Standard Deviation based on median of historical project data.					

TABLE 7.12 Testing tolerances for specification

Test	Testing Tolerance, +/- %
19.0mm	8.4
9.5mm	8.4
2.36mm	5.6
75 m	2.2
AC (Extraction)	0.6
Gmb	0.036 sp. gr.
Gmm	0.028 sp. gr.
Air Voids	1.7
VMA	1.1
VFA	8.1



Sample	Agency Air Voids,%	Contractor Air Voids,%	Difference Air Voids,%
1	4.5	4.0	0.5
2	4.6	4.4	0.2
3	4.8	4.5	0.3
4	3.8	4.1	-0.3

Median of existing database = -0.3
Median of sample dataset = 0.25

Agency Database				Project Dataset	
Sample	Absolute Value	Sample	Absolute Value	Sample	Absolute Value
1	0.0	14	0.4	1	0.25
2	0.4	15	0.0	2	0.05
3	0.7	16	0.1	3	0.05
4	0.2	17	0.9	4	0.55
5	1.2	18	0.6		
6	0.5	19	0.4		
7	0.0	20	0.1		
8	0.0	21	0.6		
9	0.1	22	0.7		
10	0.0	23	0.3		
11	0.6	24	0.0		
12	0.3	25	0.6		
13	1.4	26	0.2		

GROUP	N	Mean	Std Dev
Existing Database	26	0.4	0.38
Project Dataset	4	0.2	0.24

Variances	T	DF	Prob>|T|
Unequal	1.2272	5.7	0.2681
Equal	0.8723	28.0	0.3905

For variances equal,
F=2.56 DF=25,3 Prob>F=0.4790

Figure 7.12. Example of variance test procedure.

the agency guidelines for the determination of minimum inspection levels can be used. Agencies could make their own decisions based on their own specific needs.

The given inspection levels may have to be adjusted during the construction stage, or at the project level, if there is only one person available to perform inspection on a particular day. Figure 7.14 illustrates how the determination of a minimum inspection level could be achieved for this project with one available inspector. A summary of what the inspector could inspect and how often is provided in Table 7.15.

7.4 PRACTICAL APPLICATIONS

Putting this research into practice will take some planning and effort. There are many issues to consider within the structure of the agency and contractor organizations. Several practical issues arise that will decide the success of the implementation effort. Practical applications are largely the identification of barriers that must be overcome to successfully achieve implementation. The following practical applications are discussed: (1) training and certification, (2) technology, (3) statistics, (4) change, and (5) implementation.

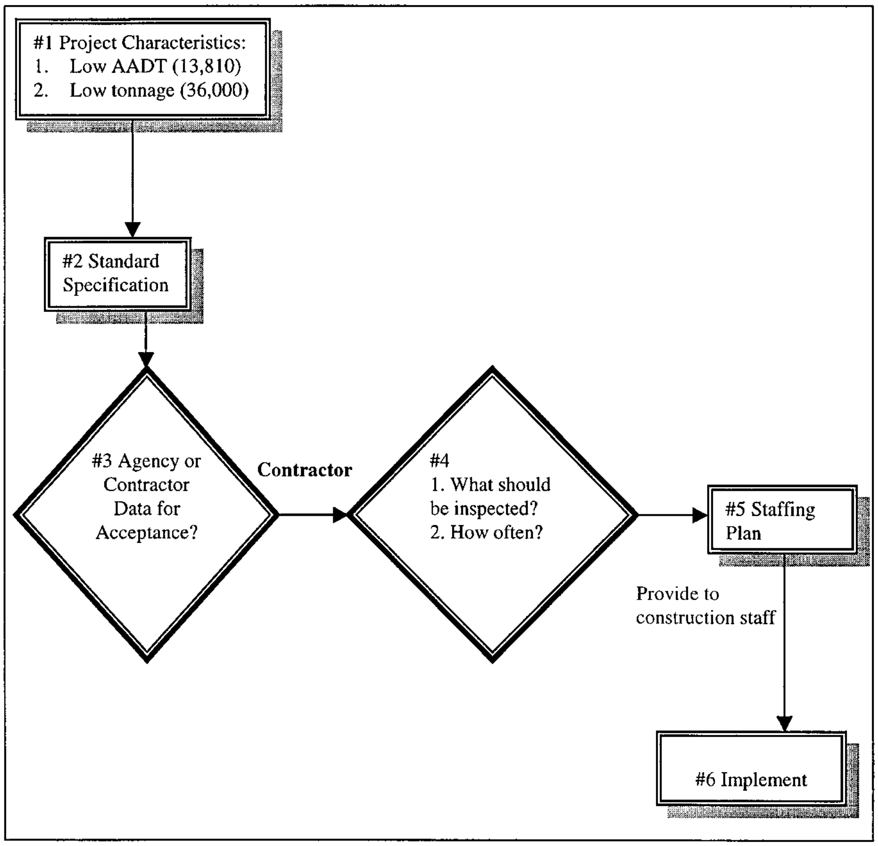


Figure 7.13. Determination of minimum inspection level during design phase.

TABLE 7.13 Implementation of minimum inspection levels

Process	Location	Inspection Task	Number of tasks per 10-hour work day *(How often)*
Raw Materials *Inspector One*	Aggregate Stockpiles	Segregation	1
		Contamination	1
		Discoloration	1
		Loading Rotation	1
Plant Mixing *Inspector One*	Aggregate Blending	Bin Contamination	1
		Bin Levels	1
		Segregation	1
		Feeder Rates	1
	Mix Plant	Contractor Sampling/Testing Procedures	Variable
	Storage	Storage Duration	As needed
Mix Transport *Inspector One*	Load-Out	Multiple Drops to Truck	1
		Tarp Use	As needed
Laydown	Base Condition *Inspector One*	Clean and Uniform Surface	10
		Dry Surface	As needed
	Paving *Inspector Two*	Collect Truck Ticket	90
		Calculate Yield	4
		Slope	10
		Grade	10
		Width	10
		Depth	10
		Inconsistencies/Segregation	10
	Rolling *Inspector One*	Maintain Rolling Pattern	5
		Surface Texture	5

TABLE 7.14 Staffing plan for construction

Location	Number of People	Required Time
Plant	1	1 hour
Laydown	2	1 day

7.4.1 Training and Certification

Each state in the study had some form of training and certification program for field-level staff. Although this may seem an obvious component of any QA program, the importance of training and formal certification must not be underestimated. Comprehensive hands-on training will largely decide the success of the program at the field level.

Certification is an important part of the training process because it validates the training program and determines the responsibilities of field-level personnel. In Arizona, training programs were certified through the National Institute for Certification in Engineering Technologies (NICET), while the other five states certified the training program through an agency program.

7.4.2 Technology

The sampling frequency for all HMA tests is largely governed by the physical time to perform the tests. The construction industry is currently limited in the ability to effectively measure and process HMA construction data for quality and productivity. There are emerging technologies that can reduce the amount of time and increase data for on-line process control. Automated aggregate gradation technology is being developed to provide real-time measurement for controlling aggregate size entering the asphalt plant mix chamber. Videoimaging technology for coarse aggregates has been successfully tested on asphalt production plants and measures on line and in real time every 5 min *(30)*. Advanced construction systems are being developed using global positioning systems (GPS) and computer simulation in order to provide more efficient path traversals by roller operators and to lead to significant reductions in total operation time *(31)*. Reduced operation time can result in labor cost savings and can ultimately reduce overall project time and total project costs.

Technology is an important practical issue during program implementation. Recent technological advances have accelerated the knowledge of HMA design and construction. Computer technologies are rapidly developing and are becoming increasingly common on construction projects. Newer database technologies are available that can provide standardized record-keeping of project checklists and daily diaries, possibly through the use of laptop computers and hand-held smart forms.

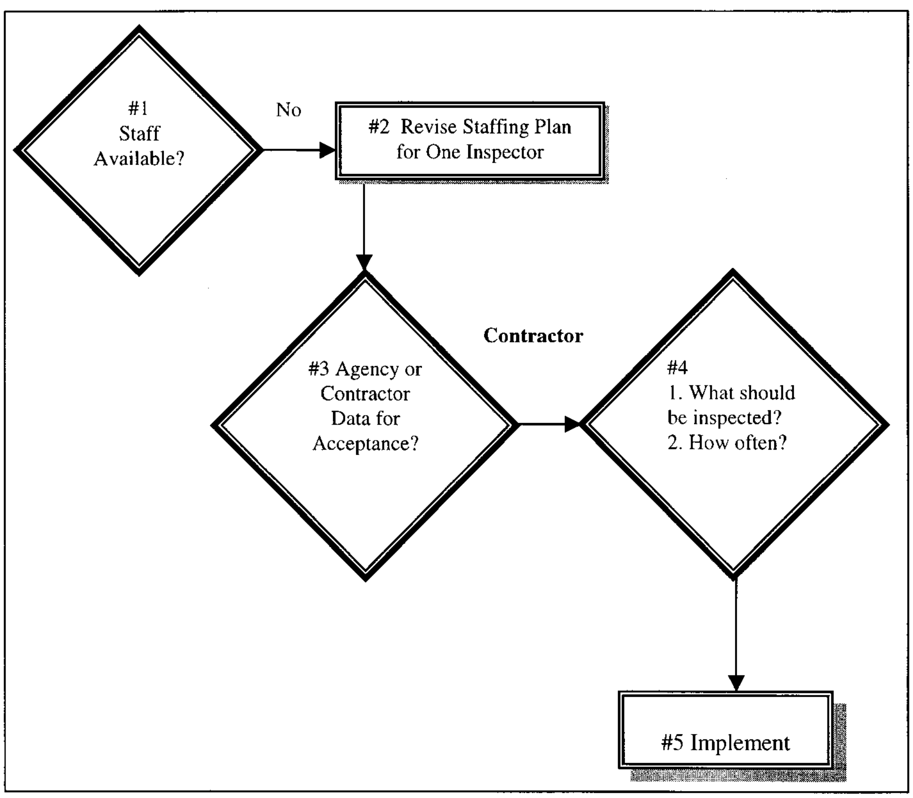


Figure 7.14. Determination of minimum inspection level for one available inspector.

TABLE 7.15 Inspection tasks for Minnesota I-90 for one inspector

Process	Inspection Tasks	Number of tasks per 10-hour work day *(How often)*
Plant Mixing	Contractor Sampling/Testing Procedures	As needed
	Storage Duration	As needed
Mix Transport	Tarp Use	As needed
Laydown	Clean and Uniform Surface	10
	Dry Surface	As needed
	Collect Truck Ticket	90
	Grade	10
	Width	10
	Depth	10
	Joint Construction	As needed
	Inconsistencies/Segregation	10

7.4.3 Statistics

The science of statistical methods and techniques is increasingly being used in construction. Statistics provide a scientific way of interpreting features found in the data unlike no other scientific method. Agencies and contractors should gain a necessary comfort level with statistics, and many organizations may already have in-house expertise in statistics. If in-house expertise is unavailable, consultants or local universities should be consulted to assist in this process.

7.4.4 Change

Personal resistance to change is a practical matter that must be overcome. Many management and field personnel are accustomed to the current program and may be reluctant to make changes. The old adage of "if it isn't broke, don't fix it" may be prevalent during any modifications to the existing program. This is not an easy barrier to overcome, given that the end-users of the program are responsible for its implementation.

There must be "buy-in" by management to support changes to the existing system, because management's initiative and leadership are necessary to implement the changes. Governmental agencies have a different working culture than private enterprise. If the agency takes a business approach to changes, it will be viewed by the public as responsive and adapting to the public's needs.

7.4.5 Implementation

The physical implementation of changes to the specification is a delicate process. Typically, an easy way to overcome full-scale implementation is through the use of pilot studies or projects. Problems to overcome in implementation can be resolved with minimal expense by piloting the new specification on several projects. Selecting a few projects allows a more detailed evaluation of each, as opposed to many projects that would make it difficult to assess the status of the implementation effort.

7.5 SUMMARY

This chapter provided implementation guidelines and a case application for the research. A system was described where decisions concerning testing and inspection levels are made throughout the life of the project. Then, a case application was provided using data from an actual project. Specific examples were given for quality control, acceptance, and verification testing levels. Developing a list of important inspections for a project based on attributes and available inspectors was discussed. Practical applications to consider during implementation were also described.

CHAPTER **8**

CONCLUSIONS AND FUTURE RESEARCH

8.1 RATIONAL METHOD FOR OTHER HIGHWAY CONSTRUCTION AREAS

The rational method for determining minimum testing and inspection levels to construct satisfactory HMA overlays can be applied to other areas of highway construction, such as Portland Cement Concrete (PCC), subgrades, and structures. The same processes described for HMA overlay testing and inspection can be used as a starting point and appropriately adjusted for the specific application area and project characteristics.

8.1.1 Testing Levels

The process for determining testing levels in other construction areas largely follows the methodology developed for HMA overlays. Primary areas to evaluate during this process are (1) sampling, (2) quality control, (3) acceptance, and (4) verification. Figure 8.1 shows the process for applying the findings of this research to other highway construction disciplines.

8.1.1.1 Sampling

Correct sampling procedures are fundamental to the testing specification. The integrity of obtaining a representative sample must be firmly established and must not be compromised by a conventional approach of simply acquiring material to perform tests. Randomization is a very important part of the sampling process because it allows each part of the population an equal chance of being selected and it protects against unsuspected sources of bias. With careful attention to the sampling design, estimates can be obtained that are unbiased for population quantities, such as the population mean and standard deviation.

There were four sampling attributes developed for HMA that will affect the ability to collect a representative sample: characteristics, resources, interaction of construction operations and sampling, and safety. Similar recognition should be given to those attributes for other highway construction areas. Identify other state-specific attributes that may influence the ability to collect a representative sample as well.

Sampling frequency is a fundamental component when determining testing levels. Two issues considered for HMA were time constraints and whether to use time or quantity for sampling frequency. Agencies and contractors should obtain time estimates for their specific testing procedures. A limited amount of test data is available because of the inherent time to perform the entire sampling and testing cycle.

Whether to use time or quantity to sample the material is a fundamental concern during specification development. There are inherent advantages and disadvantages with either time- or quantity-based sampling. The list of advantages and disadvantages enumerated for HMA overlays should be considered during this decision-making process. Time-based sampling frequencies can be easily obtained by using the testing time estimates collected by the agency and contractor. A practical way of determining a quantity-based sampling frequency is to match sampling and testing time with construction production rates.

8.1.1.2 Quality Control

Quality control is where the true quality is built into the product. The key to effective quality control for highway construction is understanding the process so that the specifications are met. In terms of field QC testing, this can be translated to meeting design targets with minimal variation. It should be a goal to continually meet the design targets and ensure that variation around the target value is minimized, because deviations from the target are known to affect quality. Thus, an understanding of variation is essential.

ANOVA models, developed earlier for HMA overlay construction, are the most powerful tool to determine where variation is found in the construction process. Control charts provide an excellent way to visualize variation, but they do not have the quantitative capacity to determine where the true sources of variation are found in the construction process. When determining testing levels, the models should be designed around a practical measure, such as days of production. If variation between days is significant, this would indicate to the contractor that the process is changing on a daily basis and that testing resources should be assigned to daily testing. If day-to-day variation does not exist, this would indicate that features within each day may be assignable to

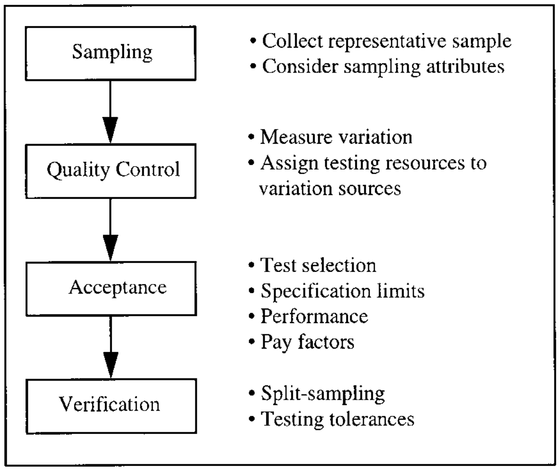


Figure 8.1. Process for determining testing levels for other types of highway construction.

variation. The ANOVA models would then be reconfigured to determine those variation sources within a day of production and then the testing resources would be assigned to those variation sources.

The ability to obtain any level of QC testing is governed by the time to conduct sampling and testing. Certain test procedures can be completed more quickly than others can and will produce larger amounts of test data. If a particular test can produce more than 20 test results per day, such as nuclear density gauges for measuring in-place density and moisture content, the average method could be used to determine the number of tests. Equation 4.4, which develops the confidence level for the average of several tests, was applied to pavement density data as a starting point to determine the number of daily density tests for quality control. There were four components to this equation that contractors can specify: (1) confidence limits of the average, (2) Z-statistic for the probability level, (3) estimated standard deviation, and (4) number of tests. Fewer tests are required when the confidence limits are increased, when the probability level is reduced, or when the standard deviation is small. Reliable estimates for the standard deviation can be found from historical data collected by the contractor.

8.1.1.3 Acceptance

Acceptance testing is central to any construction testing specification. As owner, the agency must have assurances that the acceptance plan will function satisfactorily and protect the interests of the agency as well as the contractor. There are three tangible areas to investigate when determining the level of acceptance testing: (1) test selection, (2) specification limits, and (3) pay factors.

Tests that could be specified for the acceptance of construction materials and workmanship range from aggregate tests to the ride index for pavement smoothness. Agencies should select those tests that have some relationship to performance and are practical. AASHTO recommends several construction tests in the QA publications, and agencies have specified certain tests unique to their states that are thought to have an influence on performance. The tests should be practical, well-understood by field-level staff, and have a limited duration so they can be conducted frequently during construction. Frequent testing results in more test data; this increases the reliability of estimating the work and minimizes risks to both the agency and contractor.

Specification limits define the allowable range for the selected construction tests. These limits have been established from AASHTO recommendations, agency studies, contractor process capability, and actual performance. Ideally, specification limits should be based on actual performance and the ability of the contractor to maintain process capability within the limits. Because it has been difficult to accurately translate the relationship between construction variability and performance, a reasonable alternative is to set specification limits using contractor variability. Using the contractor's variability provides a benchmark of project inputs during construction, which can be later translated to outputs in the form of actual field performance. Thus, assessing contractor inputs (e.g., meeting design target levels and project variability) is essential when establishing specification limits. It better defines the link between design tolerances (specification limits) and actual performance. In other words, as contractors achieve a certain level of variability, the resulting performance can be better understood and a relationship developed. The statistical methods used to evaluate the contractor's ability to achieve specification limits for HMA overlays should be applied to other construction areas.

Pay factors are integral to QA specifications. In concept, the pay factor is the difference between life-cycle costs from design and actual life-cycle costs from construction. The relationship between actual construction quality and life-cycle costs is better understood as data are collected. During the interim, the agency and contractor should develop an appropriate sample size and corresponding lot size for the pay factor. First, the sample size should be investigated by understanding how a certain number of samples will provide different risk levels. A system was described for HMA overlays where the contractor's population mean and standard deviation were developed, and different sample sizes were sampled from this distribution to obtain risk levels. This same procedure should be applied to other construction areas. More critical projects with higher life-cycle costs will typically require larger sample sizes for the lot, while less critical projects will have lower life-cycle costs. The lot size then becomes a function of the sample size. Once an appropriate sample size has been determined, the lot size is determined by the time necessary to complete testing of the samples. Either time or quantity can be used to construct the lot.

8.1.1.4 Verification

Verification testing is an important component of the QA program when contractor tests are used for acceptance. The required amount of agency testing can be reduced to the amount necessary to verify the contractor's results. This amount of testing is a function of several factors, including (1) whether to use split samples or independent samples for the verification, (2) testing tolerance between laboratories, (3) variation between tests, (4) confidence level for the comparison, and (5) probability of detecting a true difference between laboratory tests.

Whether to use split-sampling or independent-sampling will affect the number of verification tests. Split-sample verification can minimize the amount of agency verification because it eliminates the materials, production, and sampling variation between laboratories. If this variation is not removed, as is the case with independent-sampling, the comparison will be obscured and require nearly double the number of samples as split-sampling, provided the same probability levels are chosen. Given that the objective is to determine a minimum level of verification testing, split-sampling probably should be chosen. The agency should witness the contractor collect and split the sample and then take immediate possession of the agency's portion of the split sample for later testing.

Testing tolerances are common in those states where contractor data are used for acceptance. The concept is straightforward and does not require elaborate statistical calculations during construction. The differences between tests are considered acceptable if they do not exceed the tolerances set forth in a developed table. A statistical procedure was developed for HMA overlay construction that develops testing tolerances using several statistical parameters. This same procedure should be used for other highway construction areas.

The number of tests for verification testing within a lot is a function of the following: (1) variation between tests, (2) difference between means, (3) probability that the mean difference is contained within a defined acceptable region, and (4) probability that the mean difference is accepted outside the defined acceptable region. Equation 4.6 should be used to relate these different factors. Agencies should collect split-sample test data and develop estimates for the standard deviation between laboratory tests (σ_d). The estimate for σ_d is then applied to Equation 4.6 to determine the allowable testing tolerance for any number of tests. Values for the confidence level and probability of detecting a true difference between tests can be found in statistics textbooks and tables.

8.1.2 Inspection Levels

Figure 8.2 illustrates a process for determining inspection levels for other highway construction areas. First, a definition of the industry's process is necessary. The definition should include descriptions of the hierarchical divisions required for inspection. The hierarchy of HMA production inspection began with process and followed down through inspection location, activities, and tasks.

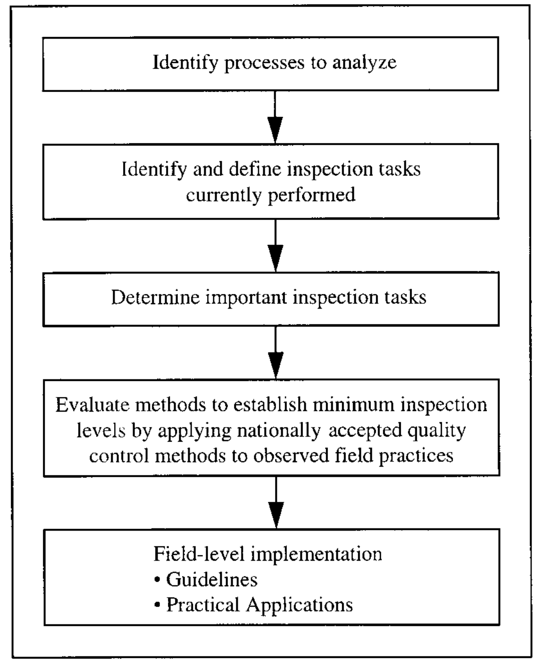


Figure 8.2. Process for determining inspection levels for other types of highway construction.

The inspection tasks, depending on the terminology a particular industry uses, should then be defined. The definitions should provide a thorough understanding of what that inspection task entails (e.g., its purpose; who should conduct it; and when, where, and how it should be completed). Industry literature, input, and current practices should be consulted for the inspection task definition.

Next, important inspection tasks should be determined through the use of field data collection and review of industry literature. It is necessary to determine what is currently thought to be important by the industry, as this gives a knowledge base of where to begin to make changes or improvements. It is also necessary to understand current practices, because there is likely justification for the current practice. These important inspections will also be referenced when forming guidelines for minimum inspection levels.

QC methods can then be used to analyze any collected field data representing current industry practices to determine minimum inspection levels. The four methods applied to HMA inspection were (1) Juran seriousness classification, (2) Pareto analysis, (3) fishbone diagrams, and (4) resource allocation. These methods provide direction for any improvements that may be made, as well as the factors that must be considered when making changes. Recommendations can then be made and should be based on key project factors.

Finally, guidelines should be created based on critical project characteristics specific to that industry. Important characteristics used for agency guidelines for HMA overlays included (1) whether agency or contractor test results were used for acceptance, (2) project AADT, and (3) availability of resources. Project size and the availability of limited resources are likely to affect inspection levels for any aspect of the highway industry. The developed guidelines can then be used as a starting point and adjusted to fit the state's culture, available resources, and the applicability of provided recommendations.

8.2 FUTURE RESEARCH

8.2.1 Relating Design and Construction with Performance

Future research must continue to address the relationship between project inputs from design and construction with pavement performance. This is a difficult challenge because of the numerous variables involved, but there must be some initiative to move this issue forward. It is recommended that a research project be developed to measure the actual performance from the 16 projects in this study to enable the design, construction, and actual pavement performance to be linked together. The performance of each project can be considered a response variable to many input variables from design and construction. Given that the traffic levels, temperature, and base condition (in certain projects) were consistent within projects, these variables could be blocked from the analysis. Then those variables that changed within the project could be evaluated. Such variables include deviations from design targets and variability for the various tests performed during construction.

Much of this analysis will involve reconstructing the projects on paper—similar to the approach used in this study. Distress survey data would be collected and compared with the as-built construction data. Figure 8.3 provides a schematic of how the construction data and distress survey data could be compared.

Designing a formal experiment on these 16 construction projects by adjusting key input variables to understand an output response variable was not possible. Because of this fact, the project could be treated as a random effect where several variables changed simultaneously without any specific treatments. Statistical regression and ANOVA models could provide an initial investigation into relationships between inputs and outputs. For example, if the design target for asphalt content was 5.0 percent, and there were significant deviations from the target in several pavement sections where raveling occurred, the regression and/or ANOVA models would detect some relationship. Once initial relationships have been disclosed, more careful investigation can be made. It may be learned that certain deviations from target are perfectly acceptable, while other deviations are critical to performance. The statistical area of multivariate analysis and response surface modeling would also be excellent tools for such an investigation.

8.2.2 Cost of Quality Principle

Future research is needed to apply the principles of cost of quality to highway construction. As a practical matter, agency quality assurance and contractor quality control consume

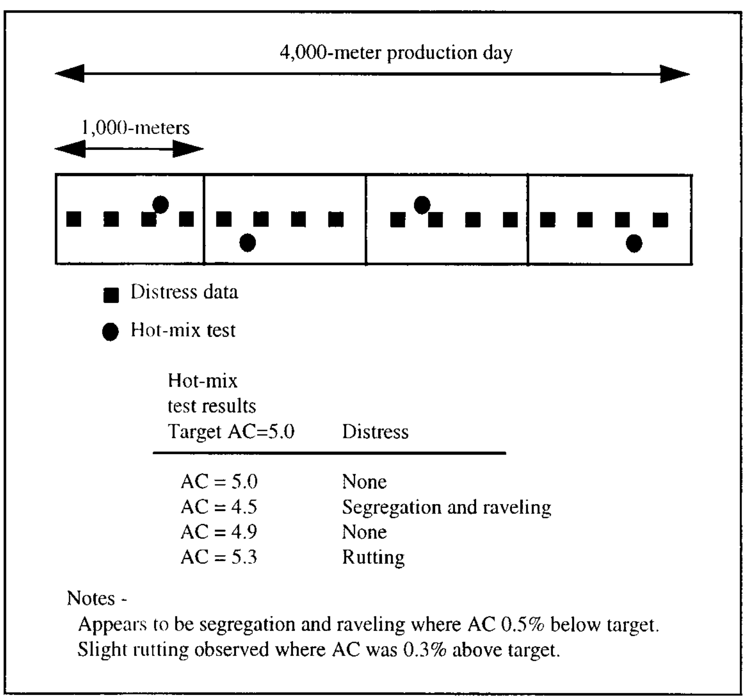


Figure 8.3. Comparison of construction data and distress data.

resources and cost money. It would be feasible to have a high level of testing and inspection on every highway construction project; however, there must be an acknowledgment of costs. Cost is a practical means to measure the use of resources during construction as well as increased or decreased performance. Cost defines the dollar value of limited resources during construction, and it also defines the initial capital to construct the highway facility and the long-term maintenance to operate the highway facility.

World leaders in quality, such as J.M. Juran, have defined the "cost of quality" as the sum of the appraisal, prevention, and failure costs *(32)*. Figure 8.4 illustrates how the cost of quality is described by these costs. Failure cost is defined by the decrease in correcting or replacing defective work, while appraisal plus prevention is the application of resources to increase the quality conformance level. The minimum cost of quality occurs at a level where the summation of failure costs and the costs of appraisal plus prevention are at a minimum. There is an upper region where the appraisal plus prevention costs provide diminishing returns; a large appraisal plus prevention cost exceeds the benefits of a reduced failure cost.

Applying the principle of cost of quality to highway construction stipulates that a certain level of appraisal plus prevention costs are necessary to reduce failure costs and minimize the cost of quality. As the quality of conformance increases, the construction quality level is thought to approach the expected level of performance. It is desirable that agencies apply an appropriate level of quality assurance and that contractors apply an appropriate level of quality control to minimize the cost of quality. There must be recognition that some failure costs will occur, despite efforts to prevent them.

8.2.2.1 Cost of Quality for Quality Control

The cost of quality principle can be adopted to highway construction that is specific to quality control, where quality control is a contractor function. A way of describing the costs of appraisal plus prevention for contractors is the added costs of personnel, equipment, materials, and methods to produce greater quality output and the costs of monitoring and verifying the quality of output. For example, the added costs to produce quality output would be construction of an additional aggregate stockpile during the crushing process to minimize stockpile segregation and variation in the distribution of particle sizes. This would require additional material handling and increase production costs, but it would provide increased control during plant mixing. The added costs for monitoring would be an additional QC technician to monitor the flow of aggregate processing with on-line testing and inspection. To achieve this increased quality of conformance, a contractor must invest greater resources in personnel, equipment, materials, and methods, and, therefore, increase additional costs.

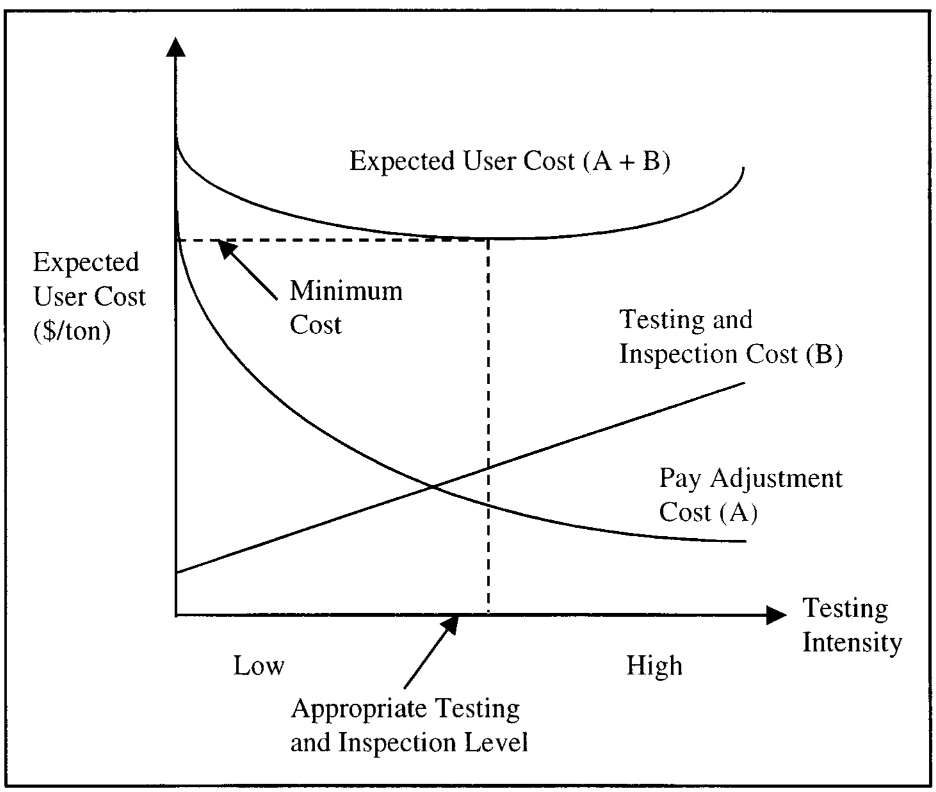


Figure 8.4. Classic model of quality costs (32).

Generally, an increase in appraisal plus prevention costs will cause an expected decrease in failure costs, assuming this cost is applied to those components of the process influencing performance. Using the previous example of greater aggregate control, there can be an expected decrease in variation and increased chances for meeting specification limits. By diminishing these detrimental effects, the cost of failure decreases and the quality of conformance increases. This increase correlates with higher levels of construction quality that will lead to higher levels of pavement performance.

A key factor that drives the quality of conformance for the contractor is the specifications. The level of quality received is a function of how the contractor understands the demands of the specification and chooses to respond to them. A common misconception is that quality can be tested into a product. Quality must be engineered into the product. Quality can be achieved through continued commitment to improving the design and construction of the product. Most contractors will continue to perform inspections, interpret test data, identify root causes of any defects, and generate options that could eliminate the quality deviations identified.

The key mechanism for enforcing specification compliance is pay factors. As shown in the synthesis of current practice, all agencies using a QA specification for HMA construction use some version of a pay factor. Statistically based specifications encourage the contractor to achieve design targets and reduce variability within lots to improve their likelihood of receiving full payment and/or incentive payments. Contractors failing to meet the specifications receive a pay deduction for the degree of quality of conformance.

8.2.2.2 Cost of Quality for Acceptance

The cost of quality principle can be adapted to highway construction acceptance. Acceptance is an owner function. A way of describing the costs of "appraisal plus prevention" for acceptance is from the costs of testing and inspection to assure and accept the construction product. Testing and inspection cost is simply the cost for administering the acceptance aspects of the QA program. The relative percentage of testing and inspection costs will be higher for smaller tonnage projects or those projects having higher testing frequency.

When setting acceptance testing and inspection levels, an answer must be provided to the question asked earlier—is it more appropriate to have relatively few construction projects with extensive testing and inspection, or a greater number of field projects with more limited testing on each? For example, on a particular project, should 2 hot-mix and 5 density samples, or 4 hot-mix and 10 density samples, be taken each day; and are one or two inspectors needed? Each testing alternative requires different levels of resources, leading to the question—what is the optimal use of limited resources? Agencies and contractors must work toward developing a level of acceptance testing and inspection that creates a rational specification that balances costs, risk, and quality.

A way of describing the failure costs for acceptance is through the additional cost incurred by the agency when there are insufficient levels of testing and inspection. For the agency, insufficient testing levels or inadequate lot sizes may lead to unacceptable work being accepted. In concept, the pay factor should represent the difference between life-cycle costs from design and the actual life-cycle costs from construction. Poor quality requires additional maintenance costs (e.g., pavement grinding and rut patching) to keep the overlay in service for the planned life cycle, or it requires reconstruction earlier than scheduled. Accepting poor work may occur from insufficient testing levels and severely disrupt the life-cycle cost formula for the project.

The contractor may also suffer the effects of insufficient testing levels by having acceptable work rejected or a reduced pay factor from an insufficient number of samples within a lot. It is desired that the contractor receive appropriate payment for acceptable work; however, there is a possibility that test results will indicate the contrary. For inspectors, a necessary level is needed to interpret and control those process components having one effect on pavement performance.

There must be careful consideration given to sublots and lots during the development of acceptance testing levels. One concept behind accepting on a lot-by-lot basis is an effort to maximize the likelihood that the material being considered is all from the same population. Combining adjoining sublot and lots would make it possible for unique and distinct populations to be combined into one lot. This could lead to multimodal populations that are not normal or lead to a normal distribution that has greater variability than that of the individual lots. This approach may defeat the purpose of accepting on a lot-by-lot basis. Thus, careful consideration must be given to defining lots.

There is a region where the level of testing and inspection provides diminishing returns for minimizing the acceptance risks for both the agency and contractor. An optimum level of acceptance testing and inspection would be defined where the summation of testing costs and the acceptance failure costs are at a minimum. An increase in appraisal plus prevention costs through greater acceptance testing levels will decrease failure costs. Using the previous example of two testing levels (2 or 4 hot-mix tests and 5 or 10 density tests), there can be an expected difference in risk of accepting a construction product. By increasing the testing level, the added costs from making incorrect decisions are decreased. An imprecise estimate is just as likely to underestimate the true quality level as overestimate it.

8.2.3 Variation Components

Four components of variation were identified in this research: (1) materials, (2) production, (3) sampling, and (4) testing. Experiments should be conducted on construction tests, in addition to laboratory air voids, to determine which variation component contributes the greater percentage to total variation. Performing a similar analysis on other tests, such as asphalt content and aggregate gradation, would require multiple-specimen testing similar to laboratory compaction of individual test specimens for determining air voids. It is proposed that similar analysis be conducted in future studies to determine the relative percentages of MPS and testing variation for the other common construction tests to understand their contribution to total variation. Experiments should be specifically designed to isolate and measure materials, production, and sampling variation components.

8.2.4 Compliance Measures

Five different measures are being used to determine specification compliance among the states: (1) Average, (2) QLA, (3) Average Absolute Deviation, (4) Moving Average, and (5) Range. Currently, AASHTO recommends the QLA method; however, a formal research study should be commissioned that compares the relative efficiency of these methods. A benefit of collecting data from states specifying different compliance measures is observing different perceptions of variation. Different feedback behaviors are created among contractors when operating under different compliance measures. For example, a given contractor may take a wait-and-see approach for changing a process with the Moving Average, while the same contractor may be inclined to make a rapid adjustment if the Absolute Average Deviation method is specified. A formal research study would require collecting data from several years within states having different compliance measures to investigate any changes in contractor behavior. Particularly, states switching compliance measures to QLA-type specifications would provide useful data.

REFERENCES

1. AASHTO. *AASHTO Implementation Manual for Quality Assurance*, American Association of State Highway and Transportation Officials, Washington, DC (1996).
2. AASHTO. *AASHTO Quality Assurance Guide Specification*, American Association of State Highway and Transportation Officials, Washington, DC (1996).
3. AASHTO. "Standard Recommended Practice for Acceptance Sampling Plans for Highway Construction," American Association of State Highway and Transportation Officials, Designation R9-90, Washington, DC (1990).
4. Anderson, D.A., D.R. Luhr, and C.E. Antle, "Framework for Development of Performance-Related Specifications for Hot-Mix Asphaltic Concrete," *NCHRP Report 332*. Transportation Research Board, National Research Council, Washington, DC (1990).
5. Luhr, D.R., B.F. McCullough, and A. Pelzner, "Simplified Rational Pavement Design Procedure for Low-Volume Roads," In *Transportation Research Record 898*, Transportation Research Board, National Research Council, Washington, DC (1983) pp. 202-206.
6. Witczak, M.W., "Development of Regression Model for Asphalt Concrete Modulus for Use in MS-1 Study," Asphalt Institute (1978).
7. FHWA. *Cost Effectiveness of Sampling and Testing Programs*. FHWA-85-29/30, U.S. Department of Transportation, Federal Highway Administration, Washington DC (1986).
8. Lybas, J.M., B.E. Ruth, and K.W. Kokomoor. "Importance and Cost-Effectiveness of Testing Procedures Related to Flexible Highway Pavement Construction in Florida," *Transportation Research Record 1034*, Transportation Research Board, National Research Council, Washington, DC (1985) pp. 123-131.
9. NCHRP 9-7 Final Report, *Field Procedures and Equipment to Implement SHRP Asphalt Specifications*, Transportation Research Board, National Research Council, Washington, DC (1996).
10. FHWA. *Practical Applications of Statistical Quality Control in Highway Construction*. FHWA-11-9435, U.S. Department of Transportation, Federal Highway Administration, Washington, DC (1979).
11. Juran, J.M. *Juran's Quality Control Handbook*, 4th Edition, McGraw Hill, Inc., New York, NY (1988).
12. AASHTO, FAA, FHWA, NAPA, U.S. Army Corps of Engineers. *Hot-Mix Asphalt Paving Handbook*, Lanham, MD (1991).
13. AASHTO, FAA, FHWA, NAPA, U.S. Army Corps of Engineers. *Hot Mix Asphalt Construction Participant Manual*, American Association of State Highway and Transportation Officials and National Asphalt Paving Association, Lanham, MD (1993).
14. NAPA. *Quality Control for Hot Mix Asphalt Operations*, National Asphalt Paving Association, Lanham, MD (1997).
15. NAPA. *Paver Operations for Quality*, National Asphalt Paving Association, Lanham, MD (1997).
16. NAPA. *Roller Operations for Quality*, National Asphalt Paving Association, Lanham, MD (1997).
17. NAPA. *Field Management of Hot Mix Asphalt*, National Asphalt Paving Association, Lanham, MD (1997).
18. Melan, E. H. *Process Management—Methods for Improving Products and Service*, McGraw-Hill, Inc., New York, NY (1993).
19. Summers, D. *Quality*, Prentice Hall, Upper Saddle River, NJ (1997).
20. Juran, J. M. *Juran on Planning for Quality*, The Free Press, New York, NY (1988).
21. Gitlow, H., A. Oppenheim, and R. Oppenheim. *Quality Management, Tools and Methods for Improvement, 2nd Edition*, Richard D. Irwin, Inc., Burr Ridge, IL (1995).
22. "Standard Test Method for Density of Bituminous Concrete in Place by Nuclear Methods," ASTM Designation D2950-91, *Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA (1991).
23. Hughes, C. S. *NCHRP Synthesis of Highway Practice 232*: *Variability in Highway Pavement Construction*, Transportation Research Board, National Research Council, Washington, DC (1996) 44 pp.
24. Thompson, S.K. *Sampling*, John Wiley and Sons, Inc., New York, NY (1992).
25. Snedecor, G.W., and W.G. Cochran. *Statistical Methods, 8th Edition*. Iowa State University Press, Ames, IA (1989).
26. Schmitt, R.L., et al. "Pavement Density Measurement and Comparative Analysis using Core and Nuclear Methods." *Journal of the Association of Asphalt Paving Technologists 1997*, Volume 66, pp. 379-406, St. Paul, MN (1997)
27. DeVor, R.E., T.-h Chang, and J.W Sutherland. *Statistical Quality Design and Control—Contemporary Concepts and Methods*, MacMillan Publishing Company, New York, NY (1992).
28. FHWA Demonstration Project 89, Quality Assurance Software for the Personal Computer, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., *Publication No. FHWA-SA-96-026* (1996).
29. *Life Cycle Cost Analysis in Pavement Design*, Draft Interim Technical Bulletin, FHWA, Washington, DC (1997).
30. Alba, F., M.Worischeck, and S. Madrigal."On-Line Real-Time Measurement and Control of Aggregate Gradation in Asphalt Plants," SHRP-IDEA project (1994).
31. Tserng, H.P., et al. "An Operations Planning System for Asphalt Pavement Compaction," Proceedings of the 13th International Symposium on Automation and Robotics in Construction, May 30—June 1, 1996, Tokyo, Japan (1996) pp. 349-358.
32. Juran, J.M., and F. Gyrna, *Quality Planning and Analysis*, 3rd Edition, McGraw-Hill, New York, NY (1993).
33. Nordheim, E.V., and M.K. Clayton, "Course Notes for Statistics/Forestry/Horticulture 571," Department of Statistics, University of Wisconsin—Madison, Madison, WI (1994).

BIBLIOGRAPHY

ASCE. "Quality in the Constructed Project—A Guide for Owners, Designers, and Constructors," American Society of Civil Engineers, New York, NY (1988).

Atkins, H. N. *Highway Materials, Soils, and Concretes, 2nd Edition*, Prentice Hall, Inc., Englewood Cliffs, NJ (1983).

Barrie, D.S., and B.C. Paulson. *Professional Construction Management, 3rd Editions*, McGraw-Hill, Inc., New York, NY (1992).

Benson, P.E. "Specification of Asphalt Concrete Using Total Quality Management Principles." 75th Annual Meeting of the Transportation Research Board, Washington, DC, *Proceedings* (January 7–11, 1996).

Roberts, F., et al. *Hot Mix Asphalt Materials—Mixture Design and Construction*, NAPA Research and Education Foundation, Lanham, MD (1991).

TRB "Glossary of Highway Quality Assurance Terms," Transportation Research Board, *Transportation Research Circular 457* (1996).

Walsh, L., R. Wurster, and R. J. Kimber. *Quality Management Handbook*, ASQC Quality Press, Milwaukee, WI (1986).

APPENDIX A

GLOSSARY OF TERMS

(Adapted from the following sources: (1) AASHTO, FAA, FHWA, NAPA, U.S. Army Corps of Engineers Hot Mix Asphalt Construction Participant Manual, (2) AASHTO Implementation Manual for Quality Assurance, (3) Melan, Eugene H. Process Management—Methods for Improving Products and Service, and (3) TRB "Glossary of Highway Quality Assurance Terms.")

Acceptable Quality Level (AQL)—That minimum level of actual quality that is considered fully acceptable as a process average for a single acceptance quality characteristic. For example, when quality is based on percent within limits (PWL), the AQL is that actual (not estimated) PWL at which the quality characteristic can just be considered fully acceptable.

Aggregate—Any hard, inert mineral used for mixing in graduated fragments. Aggregate includes sand, gravel, crushed stone, and slag.

Angle of attack—The angle at which the screed bottom travels throughout the asphalt material.

Asphalt—A dark brown to black cementitious material in which the predominating constituents are bitumens, which occur in nature or are obtained in petroleum processing.

Asphalt cement—A fluxed or unfluxed asphalt specially prepared as to quality and consistency for direct use in the manufacture of bituminous pavements and having a penetration at 25(C of between 5 and 300, under a load of 100 g applied for 5 s.

Asphalt overlay—One or more courses of asphalt construction on an existing pavement; the overlay may include a leveling course, to correct the contour of the old pavement, followed by a uniform course or courses to provide thickness.

Buyer's Risk (β)—Also called *type II* or β *error*. The probability that an acceptance plan will erroneously fully accept (at 100 percent pay or greater) RQL material or construction with respect to a single acceptance quality characteristic. It is the risk the highway agency takes in having RQL material or construction fully accepted.

Compaction—The process of removing the air voids in the material after placement.

Crown—The ability to change the transverse profile of the mat being placed. Also refers to the transverse profile or the opposite slopping sides of the existing grade.

Crown control—A device that shapes the screed to form a mat with the desired crown.

Density—The degree to which air voids have been removed from a material.

End-result specifications—Specifications that require the contractor to take the entire responsibility for supplying a product or an item of construction. The highway agency's responsibility is to either accept or reject the final product or apply a price adjustment that compensates for the degree of compliance with the specifications.

Expected Pay (EP) Curve—a graphic representation of an acceptance plan that shows the relation between the actual quality of a lot and its expected pay.

Grade—a) Refers to the surface over which paving is to be done. b) Refers to the longitudinal angle of rise or fall of the roadway. c) Refers to the elevation of the roadway.

Grade control—The electronic system for controlling the longitudinal elevation of the mat from a given reference.

Head of material—The given volume and level of material in front of and across the width of the screed.

HMA drum mix facility—A manufacturing facility for producing bituminous paving mixtures that continuously proportions aggregates, heats and dries them in a rotating drum, and simultaneously mixes them with a controlled amount of bituminous material.

Hopper—That section of the paver that receives the paving material from the external source.

Hot-Mix Asphalt (HMA)—A scientifically proportioned mixture of graded aggregate and asphalt cement, produced at a batch or drum-mixing facility, which is spread and compacted while at an elevated temperature. To dry the aggregate and obtain sufficient fluidity of the asphalt cement, both must be heated prior to mixing giving origin to the term "Hot Mix."

HMA construction inspection processes—First division of four in a hierarchical structure of HMA construction inspection including four parts: (1) raw materials, (2) plant mixing, (3) mix transport, and (4) laydown.

HMA construction inspection locations—Twelve physical locations that fit within the four HMA construction processes including: (1) aggregate stockpiles, (2) asphalt binder, (3) aggregate blending, (4) asphalt binder delivery, (5) mix plant, (6) dust collector, (7) mix storage, (8) load-out, (9) weigh scale, (10) base condition, (11) paving, and (12) rolling.

HMA construction inspection activities—Third division of four in a hierarchical structure of HMA construction inspection. Several inspection activities are identified at each of the 12 HMA construction inspection locations.

HMA construction inspection task—Fourth division of four in a hierarchical structure of HMA construction inspection. Inspection tasks are identified for each inspection activity and are to be performed by an inspector to help ensure that a quality pavement is constructed.

Inspection—The determination of whether a product conforms to a specification (Juran, 1998). The AASHTO *Implementation Manual for Quality Assurance* states that inspection is an activity as important to Quality Control as it is to acceptance for both production facilities and field observations.

Job Mix Formula—Target gradation and asphalt content for a particular specification.

Joint—The area where two mats meet or join.

Mat—Asphalt materials placed by the paver.

Materials and methods specifications—Specifications that direct the contractor to use specified materials in definite proportions and specific types of equipment and methods to place the materials.

Pavement performance—The history of pavement condition indicators over time or with increasing axle-load applications (AASHTO, 1993).

Operating characteristic (OC) curve—A graphic representation of an acceptance plan that shows the relationship between the actual quality of a lot and either (a) the probability of its acceptance (for accept/reject acceptance plans) or (b) the probability of its acceptance at various payment levels (for acceptance plans that include pay adjustment provisions) (TRB, 1996).

Operator—The person whose primary function is to control the paver's speed and direction.

Paver—A self-propelled construction machine (either rubber-tired or crawler-mounted) specifically designed to receive, convey, distribute, profile, and compact paving material by the free-floating screed method.

Percent Defective (PD)—Also called **percent nonconforming.** The percentage of the lot falling outside specification limits. It may refer to either the population value or the sample estimate of the population value.

Percent within Limits (PWL)—Also called **percent conforming.** The percentage of the lot falling above a lower specification limit, beneath an upper specification limit, or between upper and lower specification limits. It may refer to either the population value or the sample estimate of the population value. PWL = 100–PD.

Performance specifications—Specifications that describe how the finished product should perform over time. For highways, performance is typically described in terms of changes in physical condition of the surface and its response to load or in terms of the cumulative traffic required to bring the pavement to a condition defined as "failure." Specifications containing warranty/guarantee clauses are a form of performance specifications.

Performance-based specifications—Specifications that describe the desired levels of fundamental engineering properties that are predictors of performance and appear in primary prediction relationships.

Performance-related specifications—Specifications that describe the desired levels of key materials and construction quality characteristics that have been found to correlate with fundamental engineering properties that predict performance.

Process control—Testing and sampling for quality verification during production.

Quality—(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, or (3) The degree to which a product or service conforms with a given requirement.

Quality Assurance (QA)—All those planned and systematic actions necessary to provide adequate confidence that a product or service will satisfy given requirements for quality. Within an organization, QA serves as a management tool. In contractual situations, QA serves to provide confidence in the supplier.

Quality assurance specifications—Also called **QC/QA specifications**. A combination of end-result specifications and materials and methods specifications. The contractor is responsible for quality control (process control), and the highway agency is responsible for acceptance of the product.

Quality characteristic—That characteristic of a unit or product that is actually measured to determine conformance with a given requirement.

Quality Control (QC)—The sum total of activities performed by the seller (producer, manufacturer, and/or Contractor) to ensure that a product meets contract specification requirements. Within the context of highway construction, this includes materials handling and construction procedures, calibration and maintenance of equipment, production process control, and any sampling, testing, and inspection done for these purposes.

Quality Index (Q)—A statistic which, when used with appropriate tables, provides an estimate of either percent defective (PD) or percent within limits (PWL) of a lot. It is typically computed from the mean and standard deviation of a set of test results as follows:

$Q_L = (X\text{-}L)/S$	where $\overline{X}$ = sample mean
	S = sample standard deviation
Or	L = lower specification limit
$Q_U = (U\text{-}\overline{X})/S$	U = upper specification limit

Rejectable Quality Level (RQL)— That maximum level of actual quality that is considered unacceptable (rejectable) as a process average for a single acceptance quality characteristic. For example, when quality is based on percent defective (PD), the RQL is that actual (not estimated) PD at which the quality characteristic can just be considered fully rejectable.

Screed—The unit that is towed behind the paver to shape, smoothe, and control the depth of the material being placed.

Segregation—When the aggregates separate from the finer materials.

Seller's Risk (α)—also called *type I error* or α *error*. The probability that an acceptance plan will erroneously reflect AQL material or construction with respect to a single acceptance quality characteristic. It is the risk the contractor or producer takes in having AQL material or construction rejected.

Slope—Refers to the transverse angle of the grade or roadway.

Tack coat—A very light application of asphalt, usually asphalt emulsion diluted with water. It is used to ensure a bond between the surface being paved and the overlying course.

APPENDIX B

INSPECTION TASKS DESCRIPTIONS

TABLE B1 Inspection Tasks Descriptions for Raw Materials Process

Where to Inspect	What to Inspect		Who Performs		When to Inspect		How to Inspect		Why to Inspect
Location	Activity	Tasks	Agency	Contractor	Agency	Contractor	Agency	Contractor	
Aggregate Stockpiles	Aggregate Quality	Consistency, Particle size/shape, Moisture content, Washed aggregates	Inspector	Loader Operator; Testing Technician	Daily plant inspection	During loading of aggregate; As scheduled or needed	Visual inspection	Visual inspection	Ensure quality aggregate in mix
	Stockpile Construction	Clean foundation, Dry foundation, Pile separation, Screen subcontractor crushed aggregate	Inspector	Loader Operator; Testing Technician	Project start-up	Project start-up	Visual inspection	Visual inspection; Implementation of correct construction methods	Reduce variability in pile; Reduce mixing among piles; Reduce pile contamination
	Maintain Stockpile Quality	Segregation, Contamination, Consistency, Discoloration, Loading rotation	Inspector	Loader Operator; Testing Technician	Daily plant inspection	During loading of aggregate; As scheduled or needed	Visual inspection	Visual inspection; Work pile face with loader; rotate loading location	Prevent early pavement failure (segregation); Reduce mixing among piles; Ensure quality aggregate in mix
Asphalt Binder	Check Quantity	Tank sticking, Verify delivery invoice	-	Plant Operator; Plant Assistant	-	As scheduled or needed; During truck delivery	-	Conduct tank sticking; Collect invoice	Record quantity used; Verify quantity delivered
	Equipment Conditions	Clean tank, Dry tank (new delivery), Proper working condition	Inspector	Plant Assistant	Project start-up	Project start-up; New binder delivery; According to maintenance schedule	Visual inspection	Visual inspection	Prevent contamination of binder; Reduce equipment downtime
	Binder Quality	Verify invoice	Inspector	Plant Operator Plant Assistant	As scheduled or needed	During truck delivery	Collect invoice	Collect invoice	Ensure correct binder for mix

TABLE B2 Inspection Tasks Descriptions for Plant Mixing Process

Where to Inspect	What to Inspect		Who Performs		When to Inspect		How to Inspect		Why to Inspect
Location	Activity	Tasks	Agency	Contractor	Agency	Contractor	Agency	Contractor	
Aggregate Blending	Loading Practices	Bin contamination, Bin levels, Segregation	Inspector	Loader Operator	Daily plant inspection	During bin loading	Visual inspection		Ensure correct aggregate proportion in mix; prevent early pavement failure (segregation)
	Proportion-ing	Feeder Rates	Inspector	Plant Operator	Daily plant inspection	During production	Monitor plant controls		Ensure correct aggregate proportion in mix
	Equipment Conditions	Bin and belt condition, Belt/weigh scale calibration, Scalping screen condition/use	Inspector	Plant Assistant	Project start-up	Project start-up; According to maintenance schedule	Visual inspection; Verify correct scale readings		Reduce downtime; Ensure correct aggregate proportion in mix
	Aggregate Quality	Moisture Content	Inspector	Testing Technician	As needed	According to project schedule	Visual inspection	Testing procedures	Prevent binder stripping from aggregate in mix
Asphalt Binder Delivery	Binder Proportion-ing	Check flow, Verify meter reading	Inspector	Plant Operator	Daily plant inspection	During production	Monitor plant controls		Ensure correct proportion of binder in mix
	Binder Tempera-ture	Monitor temperature	Inspector	Plant Operator	Daily plant inspection	During production	Monitor plant controls		Prevent permanent aging of binder; Prevent additional compaction during laydown
	Equipment Conditions	Pump and pipeline calibration, Pump and pipeline condition	Inspector	Plant Assistant	Project start-up	Project start-up; According to maintenance schedule	Visual inspection; Verify correct scale readings		Reduced downtime; Ensure correct proportion of binder in mix

TABLE B2 (*Continued*)

Mix Plant	Plant Operations	Burner efficiency, Mix discharge, RAP operations, Efficiency	-	Plant Operator	-	During production	-	Monitor plant controls	Ensure proper mixing and quality of mix
	Equipment Conditions	Proper working condition, Calibration	Inspector	Plant Assistant	Project start-up	Project start-up; According to maintenance schedule	Visual inspection; Verify correct scale readings		Reduced downtime; Ensure correct plant settings for quality mix
	Mix Quality	Mix temperature, Sampling/Testing procedures	Inspector	Plant Operator; Testing Technician	Daily plant inspection	During production; During testing	Physical measurement; Visual inspection	Physical measurement; monitor plant controls; Correct testing methods	Avoid binder stiffening; Ensure accurate testing results
Dust Collector	Dust Collector Operations	Temperature of gases, Dust carryout	-	Plant Operator	-	During production	-	Monitor plant controls; Visual inspection	Prevent damage to equipment; Ensure proper amount of fines in mix
	Equipment conditions	Proper working condition, Fines return location, Proper housings/cover	Inspector	Plant Assistant	Project start-up	During production	Visual inspection	Visual inspection	Reduce downtime; Ensure proper emissions-control
Storage	Storage Operations	Storage duration, Loading method (segregation), Storage volume	Inspector	Plant Operator	As needed	During production	Visual inspection	Implement correct storage methods	Avoid deterioration of mix; Prevent early pavement failure (segregation); Keep production at level of demand
	Mix Quality	Mix temperature, Moisture content	-	Plant Operator; Testing Technician	-	As scheduled or needed	-	Physical measurement; Testing procedures	Avoid mix stiffness; Avoid compaction problems during laydown

TABLE B3 Inspection Tasks Descriptions for Mix Transport Process

Where to Inspect	What to Inspect		Who Performs		When to Inspect		How to Inspect		Why to Inspect
Location	Activity	Tasks	Agency	Contractor	Agency	Contractor	Agency	Contractor	
Load-Out	Loading Practices	Multiple drops (segregations), Use tarps	Inspector	Plant Operator; Truck Driver	Daily; As needed	During production	Visual inspection	Implementation of correct methods	Prevent early pavement failure (segregation); Maintain mix temperature; Protect mix from environment
	Truck Conditions	Proper working order (i.e. mechanical condition/tarps), Clean beds, Proper release agents	Inspector	Truck Driver	Daily inspection	According to maintenance schedule; During production	Visual inspection		Avoid deleterious material in truck bed; Avoid downtime
	Mix Quality	Mix Temperature	Inspector	Paving Crew	Daily; As needed	During production	Physical measurement		Avoid compaction problems during laydown
Weigh Scale	Equipment Conditions	Calibration, Proper working condition, Tare weight verification	Inspector	Plant Operator; Plant Assistant	Project start-up	Project start-up; According to maintenance schedule	Visual inspection; Verify correct scale readings		Ensure correct mix quantities recorded; Reduced downtime

TABLE B4 Inspection Tasks Descriptions for Laydown Process

Where to Inspect	What to Inspect		Who Performs		When to Inspect		How to Inspect		Why to Inspect
Location	Activity	Tasks	Agency	Contractor	Agency	Contractor	Agency	Contractor	
Base Condition	Surface Condition	Surface temperature, Dry surface, Clean and uniform surface	Inspector	Paving crew	During production		Visual inspection; Physical measurement		Ensure smooth final surface; Avoid water between layers (may result in stripping)
	Tack Coat	Tack coat temperature, Uniform tack coat	Inspector	Paving crew	During production		Physical measurement; Visual inspection		Ensure bond between tack and new surface; Ensure smooth final surface; Avoid waste
Paving	Mix Delivery	Check truck ticket, Calculate yield, Mix temperature	Inspector	Paving crew	During production		Collect ticket; Calculate yield; Physical measurement		Correct mix delivery, correct amount of mix is paved
	Paver Settings	Angle of attack, Strike-off plate, Head of material, Screed, Line of pull, Hopper wings, Material feed sensors	-	Paving crew	-	During production	-	Monitor paver settings	Correct settings for correct pavement specifications and quality pavement
	Equipment Conditions	Proper working order (maintenance)	Inspector	Paving crew	Project start-up	Project start-up; According to maintenance schedule	Visual inspection	Visual inspection	Equipment in good working order (reduced downtime)
	Pavement Specifica-tions	Establish grade reference, Grade control, Depth checking, Slope accuracy, Width	Inspector	Paving crew	During production		Implementation of proper method; visual inspection		Pavement according to specification/quality pavement
	Paving Operations	Truck to paver discharge (segregation), Constant paver rate, Paving inconsistencies, Crown, Joint construction	Inspector	Paving crew	During production		Visual inspection		Avoid segregation throughout process, smooth pavement, avoid segregation
Rolling	Rolling Operations	Continuous roller movement, Mat temperature, Air temperature, Establish rolling pattern, Maintain rolling pattern, Crown, Joint construction	Inspector	Rolling crew	During production		Visual inspection; Physical measurement		Avoid inconsistencies in pavement, effective compaction; joints cause weakness in pavement
	Roller Conditions	Proper working order (maintenance), Adequate water/release agent, Proper impact spacing, Tire pressure	Inspector	Rolling crew	Project start-up	Project start-up; According to maintenance schedule; During production	Visual inspection	Visual inspection; Physical measurement; Implement proper methods	Avoid mix sticking to rollers, Equipment in good working order (reduced downtime)
	Pavement Quality	Consistent surface texture, Density	Inspector	Density technician	During production		Visual inspection; Monitor contractor testing	Visual inspection; Monitor with nuclear gage	Smooth surface, meet specification/longer lasting pavement

APPENDIX C

SAMPLE SIMULATION PROGRAM

Figure C.1 provides the MINITAB™ program to generate a population distribution for 1,000 data points. The inputs to the normal distribution are a mean of 94 percent density and standard deviation of 1.7 percent density. The output file 'random.dat' is generated from these inputs with 1,000 data points that create the normal distribution.

```
MTB > random 1000 c1;
 SUBC>normal 94 1.7.
MTB > write 'random.dat' c1
 REPLACE EXISTING FILE? y
MTB > stop
```

Figure C.1. Program to develop population distribution of 1,000 data points.

Figure C.2 provides the SAS™ program to generate different sample sizes using the 'random.dat', then calculate the PWL and pay factor for each sample.

```
options nocenter ls=78 ps=150;

data a;
infile 'random.dat' missover;
input void;
uni=uniform(0);

proc sort; by uni;

data b; set a;
grp=ceil(_n_/4);

proc sort; by grp;
proc means noprint; by grp;
var void;
output out=c mean=mvoid std=stdvoid n=nvoid;

data values;
infile 'n4qla.dat' missover firstobs=2;
input q pwl;

proc sort; by q;

data v1; set values; rename q=qll;

data v2; set values; rename q=qul;

data d1; set c;
qll=((mvoid-3)/stdvoid);
qll=round(qll,.01);
drop _type_ _freq_;

proc sort; by qll;

data d2; set c;
qul=((6.5-mvoid)/stdvoid);
qul=round(qul,.01);
drop _type_ _freq_;

proc sort; by qul;


proc sort data=d1; by qll;

data part1; merge v1 d1(in=ind); by qll;
if ind;
rename pwl=pwll;
proc sort data=part1; by grp;


proc sort data=d2; by qul;

data part2; merge v2 d2(in=ind); by qul;
if ind;
rename pwl=pwlu;
proc sort data=part2; by grp;
```

Figure C.2. Program to calculate pay factor from different samples sizes.

(*Continued*)

```
data both; merge part1 part2;

data final; set both;
pwl=pwlu+pwll-100;
lpf=83+(.2*pwl);
mpf=55+(.5*pwl);
spf=10+pwl;
if pwl=. then do; pwl=100; lpf=103; mpf=105; spf=110; end;


proc print;
var grp mvoid stdvoid qll qul pwll pwlu pwl;

data frequent; set final;
npwl=0*(pwl<10)+10*(10<=pwl<20)+20*(20<=pwl<30)+30*(30<=pwl<40)
+40*(40<=pwl<50)+50*(50<=pwl<60)+60*(60<=pwl<70)+70*(70<=pwl<80)
+80*(80<=pwl<90)+90*(90<=pwl<100)+100*(pwl>=100);

nlpf=0*(lpf<10)+10*(10<=lpf<20)+20*(20<=lpf<30)+30*(30<=lpf<40)
+40*(40<=lpf<50)+50*(50<=lpf<60)+60*(60<=lpf<70)+70*(70<=lpf<80)
+80*(80<=lpf<90)+90*(90<=lpf<100)+100*(lpf>=100);

nmpf=0*(mpf<10)+10*(10<=mpf<20)+20*(20<=mpf<30)+30*(30<=mpf<40)
+40*(40<=mpf<50)+50*(50<=mpf<60)+60*(60<=mpf<70)+70*(70<=mpf<80)
+80*(80<=mpf<90)+90*(90<=mpf<100)+100*(mpf>=100);

nspf=0*(spf<10)+10*(10<=spf<20)+20*(20<=spf<30)+30*(30<=spf<40)
+40*(40<=spf<50)+50*(50<=spf<60)+60*(60<=spf<70)+70*(70<=spf<80)
+80*(80<=spf<90)+90*(90<=spf<100)+100*(spf>=100);

proc freq;
table npwl nlpf nmpf nspf;

proc means data=final noprint;
var pwl lpf mpf spf;
output out=e n=npwl mean=mpwl mlpf mmpf mspf;
proc print;

proc means data=final noprint;
var pwl lpf mpf spf;
output out=f n=npwl std=stdpwl stdlpf stdmpf stdspf;
proc print;

proc means data=final noprint;
var pwl lpf mpf spf;
output out=g n=npwl stderr=sepwl selpf sempf sespf;
proc print;
```

Figure C.2. (*Continued*)

The **Transportation Research Board** is a unit of the National Research Council, which serves the National Academy of Sciences and the National Academy of Engineering. The Board's mission is to promote innovation and progress in transportation by stimulating and conducting research, facilitating the dissemination of information, and encouraging the implementation of research results. The Board's varied activities annually draw on approximately 4,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purpose of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both the Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

Abbreviations used without definitions in TRB publications:

AASHO	American Association of State Highway Officials
AASHTO	American Association of State Highway and Transportation Officials
ASCE	American Society of Civil Engineers
ASME	American Society of Mechanical Engineers
ASTM	American Society for Testing and Materials
FAA	Federal Aviation Administration
FHWA	Federal Highway Administration
FRA	Federal Railroad Administration
FTA	Federal Transit Administration
IEEE	Institute of Electrical and Electronics Engineers
ITE	Institute of Transportation Engineers
NCHRP	National Cooperative Highway Research Program
NCTRP	National Cooperative Transit Research and Development Program
NHTSA	National Highway Traffic Safety Administration
SAE	Society of Automotive Engineers
TCRP	Transit Cooperative Research Program
TRB	Transportation Research Board
U.S.DOT	United States Department of Transportation

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

National Academy of Sciences
National Academy of Engineering
Institute of Medicine
National Research Council