STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV 10/98)

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA22-2650-1		

4. TITLE AND SUBTITLE	5. REPORT DATE
Developing Guidelines for Cracking Assessment for Use in Vendor Selection Process for Pavement Crack Data Collection/Analysis Systems and/or Services	March 2020
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Dennis A. Morian, P.E., Douglas Frith, P.E., Shelley Stoffels, D.E.Shervin, Jahangirnejad, Ph. D., P.E.	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NUMBER
Quality Engineering Solutions, Inc. 405 Water Street, PO Box 3004 Conneaut Lake, PA 16316	11. CONTRACT OR GRANT NUMBER 693JJ318P000034
12. SPONSORING AGENCY AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Federal Highway Administration Office of Technical Service – Resource Center 31 Hopkins Plaza, Suite 840 Baltimore, MD 21201	Final Report
	14. SPONSORING AGENCY CODE N/A

15. SUPPLEMENTARY NOTES

Andrew Mergenmeier, FHWA COTR

16. ABSTRACT

State and local transportation agencies collect pavement condition data to support their pavement management programs. The majority of state agencies collect network-level data with systems that operate at the prevailing speed limit. This report documents the development of technical guidelines for agencies consideration in the procurement process for automated and semi-automated pavement cracking data collection/ analysis systems and/or services. The guidelines include a statistically based process for the evaluation of the crack detection and analysis systems.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Pavement cracking distress, HPMS crack report criteria, paired t-test for equivalence testing, pavement management systems, automated and semi-automated crack identification	

19. SECURITY CLASSIFICATION *(of this report)*	20. NUMBER OF PAGES	21. COST OF REPORT CHARGED
Unclassified	123	None

DISCLAIMER STATEMENT

Developing Guidelines for Cracking Assessment for Use in Vendor Selection Process for Pavement Crack Data Collection/Analysis Systems and/or Services

Prepared for:

Federal Highway Administration
Office of Technical Services
FHWA Resource Center
Pavement & Materials Technical Services Team

Prepared by:

Quality Engineering Solutions, Inc.
PO Box 3004
Conneaut Lake, PA 16316

March 2020

U.S. Department of Transportation
Federal Highway Administration

FOREWORD

This project was initiated by the Transportation Pooled Fund study, TPF-5(299), to develop guidelines that include technical assessment protocols for automated and semi-automated pavement cracking data collection/analysis systems and/or services for use in agency vendor selection contracting documents. Key tasks in the completion of this project were to conduct a literature review, a review of previous TPF-5(299) documents, and other associated information. In addition, a data gathering effort consisted of obtaining information and meeting with the pooled fund study agencies to gain insight and experience with the procurement process. Information was also obtained from many of the vendors who are currently providing high speed cracking data technology and/or services.

Identified within this report are best practices currently being used by agencies to select vendors or to purchase equipment as well as information regarding the current automated data collection technology available. This report supports the guideline development, providing discussions around key issues considered necessary for procurement purposes. Identified within this report is the recommended method of using ground reference measurements to evaluate vendors and/or equipment and provides the statistical support needed to make sound decisions.

This report should be of interest to any local, county, state, or federal agency who is considering the procurement of an automated cracking distress data collection vendor for the acquisition of pavement distress information for network and/or project level pavement evaluations. The report is also valid for any agency who desires to purchase their own data collection and processing equipment.

Bernetta L. Collins
Director, National Resource Center

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA-RC-20-0005	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Developing Guidelines for Cracking Assessment for Use in Vendor Selection Process for Pavement Crack Data Collection/Analysis Systems and/or Services		5. Report Date March 2020 6. Performing Organization Code:
7. Author(s) Dennis A. Morian, P.E. Douglas Frith, P.E. Shelley Stoffels, D.E. Shervin Jahangirnejad, Ph.D., P.E.		8. Performing Organization Report No.
9. Performing Organization Name and Address Quality Engineering Solutions, Inc. 405 Water Street, PO Box 3004 Conneaut Lake, PA 16316		10. Work Unit No. 11. Contract or Grant No. 693JJ318P000034
12. Sponsoring Agency Name and Address Federal Highway Administration Office of Technical Service – Resource Center 31 Hopkins Plaza, Suite 840 Baltimore, MD 21201		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code N/A
15. Supplementary Notes: Andrew Mergenmeier, FHWA COTR		

16. Abstract

State and local transportation agencies collect pavement condition data to support their pavement management programs. The majority of state agencies collect network-level data with systems that operate at the prevailing speed limit. This report documents the development of technical guidelines for agencies consideration in the procurement process for automated and semi-automated pavement cracking data collection/analysis systems and/or services. The guidelines include a statistically based process for the evaluation of the crack detection and analysis systems.

17. Key Words Pavement cracking distress, HPMS crack report criteria, paired t-test for equivalence testing, pavement management systems, automated and semi-automated crack identification		18. Distribution Statement	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 123	22. Price N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

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

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AASHTO	-	American Association of State Highway Transportation Officials
ASR	-	Alkali Silica Reactivity
CALTRANS	-	California Department of Transportation
CRCP	-	Continuously Reinforced Concrete Pavements
DOT	-	Department of Transportation
FHWA	-	Federal Highway Administration
HPMS	-	Highway Performance Monitoring System
JCP	-	Jointed Concrete Pavement
LCMS	-	Laser Crack Measurement System
LiDAR	-	Light Detection and Ranging
LTPP	-	Long Term Pavement Performance
NCHRP	-	National Cooperative Highway Research Program
PCI	-	Pavement Condition Index
PMS	-	Pavement Management System
RPUG	-	Road Profiler Users' Group
SCANNER	-	Surface Condition Assessment of the National Network of Roads
TRB	-	Transportation Research Board

INTRODUCTION

This project was initiated by the Transportation Pooled Fund study, TPF-5(299), to develop guidelines that include technical assessment protocols for automated and semi-automated pavement cracking data collection/analysis systems and/or services for use in agency vendor selection contracting documents. Key tasks in the completion of this project were to conduct a literature review, a review of previous TPF-5(299) documents, and other associated information. In addition, a data gathering effort consisted of obtaining information and meeting with the pooled fund study agencies to gain insight and experience with the procurement process. Information was also obtained from most of the vendors who are currently providing high speed cracking data technology and/or services.

A summary of project objectives follows. The guidelines are to be applicable for automated 2D/3D systems; they should be:

- efficient and effective in providing the contracting agency the information needed to select a vendor
- sensitive to the vendor resources required to provide the information
- prepared in an appropriate format

In addition, the team was tasked with assessing the need to develop separate protocols for vendor services and vendor data collection/analysis systems.

The team recommends that two separate guidelines be developed, Guidelines for Vendor Selection and Guidelines for Equipment and Software Purchasing. The Guidelines for Vendor Selection are to be developed to answer the following questions:

1. Are high quality digital distress images that can be used to identify and classify cracking being delivered by the vendor?
2. Is the vendor capable of delivering acceptable cracking distress data?
3. Does the vendor have the capability and capacity to continue at full production speed and extent?

The Guidelines for Equipment and Software Purchasing will address the following questions:

1. Are high quality digital distress images that can be used to identify and classify cracking being collected by the system?
2. What is the reliability of the data collection system?
3. Does the software allow the agency to do what they need with the collected data?
4. Does the degree of automation meet the agency's needs in terms of speed, precision, and available personnel?

This report documents the findings and provides proposed guidelines in the appendix.

LITERATURE REVIEW AND DATA GATHERING EFFORT

The project team completed a literature review and information gathering effort. This process provided insight into the historical, current, and potential future automated pavement cracking data collection efforts. Key items of concern have been the collection of images which appropriately represent the cracking on the pavement as well as the interpretation of these images.

Consideration has been given to the determination of image quality measures and evaluations for images from other industries such as the medical and textile industries. The literature review was supplemented with agency specific questions for the agencies in the pooled fund study. The project team also met with the pooled fund study agencies at the TPF-5(299) meeting held in conjunction with the 2018 Road Profiler Users' Group (RPUG) meeting. The results of the information gathering efforts are provided in the following sections.

Literature Review

Literature regarding pavement condition data collection systems and processes from the 1980s to date were reviewed and a summary of the literature is presented in this section. Chronological order was employed in presenting the relevant literature in order to illustrate the progress in this field over time.

The 1980s

A paper presented to the Second North American Pavement Management Conference in 1987 presents an assessment of leading technologies (at the time) for automated pavement crack detection.(1) The paper cites a report dated 1981 which had mentioned that until that date, surface distress, particularly cracking, had not been well characterized by any automated means.

The paper presents three methods for crack detection and specific techniques used in each method, as well as each technique's merits and limitations. The three methods are:

1. Range finding methods including acoustic (e.g., ultrasonic) and optical (e.g., laser) techniques.

2. Reflective optical methods including photolog, short-exposure video capture, line scan, flying spot laser scanner, and directed light meter (slit integration) techniques.

3. Acoustic pickup techniques (e.g., microphonic).

The technologies are characterized by the data collection and processing systems, grouped into optical processing, analog electronic processing, digital electronic processing, and hybrid approaches.

The paper mentions that digital image processing is the most common automated treatment of pavement data for crack detection. It also mentions that image processing may be applied to different types of images, as derived from different instruments, cameras, electronically-scanned photographs, or tape recorders.

The 1990s

A Transportation Research Record publication from 1994 is dedicated to "Pavement Management Systems."[2] It includes 24 peer-reviewed papers, one of which presents a universal cracking indicator for pavements.[3] Of interest are sections discussing cracking characteristics and existing measures of cracking at the time. It mentions that cracking is characterized by five attributes: extent, severity, intensity, pattern, and location. It also enumerates some of the existing measures for cracking such as Texas Department of Transportation's (TxDOT) "cracking score," the American Association of State Highway Transportation Officials (AASHTO) "extent by severity class," the Long-Term Pavement Performance (LTPP) program's "extent by type and severity," and pavement condition index (PCI).

A couple of ASTM standards were developed in the late 1990s which include information relevant to this study. ASTM E 1778 presents standard terminology relating to pavement distress[4] and ASTM D 6433 presents standard practice for roads and parking lots pavement condition index surveys.[5]

The 2000s

Early in 2000, the Virginia DOT embarked on an effort to improve the inventory and management of all of their roadway related assets. As part of this effort, members of the project team developed a unique quality management plan for the pavement assets, which was presented in a poster session at the 2001 TRB.[6] Further details were presented and published in 2002 at the Pavement Evaluation Conference held in Virginia [7] This early work laid the foundation for subsequent quality management plan development.

The 2003 distress identification manual for the LTPP program provides a consistent, uniform basis for collecting distress data for the program. It also provides a common language for describing cracks, potholes, rutting, spalling, and other pavement distresses monitored by the program.[8] For each distress and pavement type, distress descriptions, severity levels, and measurement instructions are provided. There are appendices for faulting measurements and profile measurements using specific devices. The LTPP program brought an increased awareness of the need for greater specificity in quantification and reductions in rating subjectivity.

NCHRP Synthesis 334, published in 2004, presents techniques for automated pavement distress collection and processing. In addition to pavement surface distress, it includes pavement ride quality, rut-depth measurements, and joint-faulting measurement techniques. Contracting issues, quality assurance, costs and benefits of automated techniques, monitoring frequencies and sampling protocols in use, degree of adoption of national standards for data collection, and contrast between the state of the art and the state of the practice were also presented in the synthesis. The report also includes a survey of state transportation agencies, the Federal Highway Administration (FHWA), Canadian provinces, and the World Road Association on techniques and processes typically used in network-level pavement management.[9]

A series of documents from the United Kingdom outline Surface Condition Assessment of the National Network of Roads (SCANNER) surveys for local roads. Specification Volume 1, published in 2005, includes advice to local authorities about the services to be provided under the SCANNER specification. Volume 3 (Acceptance Testing and Accreditation) presents the requirements for testing survey vehicles to become accredited by site and network tests as well as the requirements for the reporting and delivery of survey data.[10]

In 2007, a TRB paper was published quantifying the effects of a complete and comprehensive quality monitoring plan, which includes quality control, quality assurance, and an independent validation and verification on pavement management data and the resulting budgetary estimates. Pre and post independent validation and verification results were analyzed to determine the effects of a comprehensive quality monitoring plan on pavement distress data collection.(11)

A California Department of Transportation (Caltrans) technical memorandum published in 2008 assesses automated pavement distress identification technologies in California. It evaluates the state of the practice for automated image capture and subsequent rating of pavement distresses.(12) It includes sections on pavement condition survey techniques used by Caltrans prior to 2008, current techniques at the time of publication, and future techniques.

A report sponsored by the New Jersey Department of Transportation (NJDOT) and the FHWA published in 2009 evaluates the automated distress survey equipment and software. Specific equipment and systems used by the NJDOT were used in the study. It was concluded that the equipment could be used to collect cracking distress data with quality control checks to ensure the accuracy of the collected, characterized, and analyzed cracking data. It was also concluded that the NJDOT needed to collaborate with the vendor to refine the data collection and analysis procedures to differentiate the location of cracking (within and outside of the wheel paths) and to provide quality control on the data collection and analysis.(13)

The 2010s

A 2010 report sponsored by the Indiana Department of Transportation (INDOT) discusses automated pavement condition data collection quality control, quality assurance, and reliability. The inherent variability of the automated data collection processes was studied and guidelines for an automated data collection quality management program in Indiana were proposed. It was recommended that a complete quality control plan be adopted for all phases of the data collection cycle including pre-project phase, data collection phase, and post-processing phase.(14)

A presentation from the 2010 Pavement Evaluation Conference presents challenges and opportunities regarding transition of automated pavement cracking analysis from two dimensional (2D) to three dimensional (3D).(15) Different methods and equipment types are presented and advantages of using 3D systems are outlined. A specific system is also introduced.

Research completed for NCHRP Project 20-74A provided insight into the development of service levels for the Interstate Highway System, which specifically included proposed Level of Service measures for the pavement assets.(16) This effort documented a survey of a number of highway agencies and the recommendations were applied to three pilot states.

A presentation from the 23rd Annual RPUG meeting in 2011 discussed the potential for comprehensive evaluation of pavement surface with 3D laser imaging.(17) It presents opportunities for a fully-automated cracking survey including data collection and data analysis, and discussed the ultimate goal of a 3D system.

A 2011 report published by the University of Arkansas discusses automated survey of pavement distress based on 2D and 3D Laser images. The goal was to produce a functional prototype with line lasers and 3D cameras that could overcome many existing limitations, with the capability of obtaining 3D pavement surface models at true 1mm resolution with full-lane coverage and conducting real-time analysis on rutting and cracking. The research team produced a prototype that could obtain 2D and 3D laser images of pavements and developed basic algorithms for measuring rutting and cracking distresses. However, it was stated that more funding was sought to establish virtual pavements at 1mm resolution and at highway speed.[18]

A 2011 circular published by the Transportation Research Board (TRB) presents automated imaging technologies for pavement distress surveys. The circular documents state-of-the-art techniques and technologies in the acquisition of pavement surface images, and basic requirements needed to automatically identify and classify pavement surface distresses.[19] Analog, digital, and laser-based imaging techniques are introduced.

The National Cooperative Highway Research Program (NCHRP) Report 748 published in 2013 presents guidelines for the use of mobile light detection and ranging (LiDAR) in transportation applications. LiDAR is a 3D measurement technology that can rapidly acquire a substantial amount of highly detailed geospatial information. Pavement condition assessment is listed as a potential LiDAR application. The report presents brief summaries from four different studies that incorporated LiDAR systems for pavement condition evaluation including rutting, ride quality (roughness/smoothness), rehabilitation, texture, and automated distress (e.g., crack detection).[20] LiDAR systems for pavement condition evaluation have not been used widely, however.

An FHWA report published in 2013 introduced an asset management system based on remote sensing that is GIS-enabled (RS-GAMS). This report covers several technologies, including automated sign detection, rut measurement, and automated crack detection.[21] Performance evaluation of crack detection algorithms is presented in Chapter 4. Included in the quantification methods are ones used in the medical fields, such as the Receiver Operator Characteristic and Hausdorff distance methods. One proposed method using buffered Hausdorff distance was evaluated and the report concludes that it is an improvement over previous efforts. Chapter 5 evaluates the use of 3D laser technology for crack detection, and reports it as successful.

Phase 2 of the final report mentioned above was published in 2014.[22] It stated that highway agencies' different applications are provided from different vendors at different times, and in order to make the data quality consistent and adequate, a standard calibration procedure is also required. It also states that a suitable data compression method is needed, along with the suggested File Exchange Format for line laser imaging data, which needs further study. It is stated that performance of a proposed automatic asphalt crack classification algorithm was validated. It is also stated that detection and measurements of various concrete pavements were validated. The report includes appendices discussing specifications for a 3D line laser imaging system laser profiler, literature review on automatic crack classification, algorithms (statistics-based approaches, neural network-based approaches, wavelet-based approaches, and graphics-based approaches), and summary descriptions of several distress identification protocols including LTPP, FHWA, AASHTO PP00 standard, AASHTO PP01 standard for automated interpretation, and state protocols, Texas and Louisiana, from the literature.

A 2014 study sponsored by TxDOT presents a field evaluation of automated distress measuring equipment in Texas. It evaluated the TxDOT automated visual distress 3D laser system van and three automated systems developed by automated visual distress data collection vendors.[23]

A paper presented at the 9th International Conference on Managing Pavement Assets in 2014 discusses an implementable framework for standardizing national pavement crack measures. It addresses the issue of significant diversity among state DOT's pavement distress protocols, which make it challenging to establish the consistent, nationwide crack measures essential to MAP-21.[24] The framework employs crack properties that are independent from state DOT's protocols. A second paper at this same conference documented the effectiveness of independent verification and validation in identifying systematic errors, correcting those, and taking steps to prevent further recurrence of such errors, based upon 10 data collection cycles using pavement monitoring information collected by a single vendor using automated data collection equipment and a semi-automated rating process.[25] Finally, a third paper from this same conference discusses the means and methods utilized by NCDOT and the quality assurance contractor to develop statistically valid quality assurance limits, which are also meaningful in terms of pavement management decision-making impacts, for the collected and processed asphalt pavement distress data. The paper describes the strategic selection of control sections to include a range and mix of the distresses with impact in the current decision trees, the data collection on the control sections, the rating methodology, and the rater pools and preparation to develop predicted limits for the control of the data. The paper discusses the consideration of multiple control indices, and the need to also reflect a range of values for those aggregate indices for multiple distresses, and presents the statistical analysis from the asphalt concrete control sites.[26]

An article published by the Institute of Electrical and Electronics Engineers in 2015 reviews 3D imaging technologies for pavement distress detection and measurements. It presents available 3D imaging methods in order to help make a quick initial technology selection and deployment by the owner agencies. It includes a taxonomy of 3D imaging technologies such as stereo imaging, shape from focus, shape from defocus, shape from shading, photometric stereo, laser scanning, interferometry, structured light systems, time-of-flight camera systems and flash LiDAR, and Microsoft Kinect as a cheap 3D alternative.[27]

A 2015 report sponsored by TxDOT presents an analysis of network-level data from automated distress surveys.[28] For this evaluation, two pavement distress data collection vendors were asked to collect full network-level semi-automated data according to TxDOT specifications for two districts and the results were compared.

A 2015 paper published in the Journal of Traffic and Transportation Engineering discusses network-level pavement evaluation with a 1-mm 3D survey system. It presents a specific commercial system which is said to obtain true 1-mm resolution 3D data at full lane coverage in all 3 directions at highway speed up to 60 mph. It is also stated that the technology presented allows highway agencies to use the 1-mm 3D system for their design and management purposes, particularly to meet the data needs for pavement management system (PMS), pavement ME design, and highway performance monitoring system (HPMS). [29]

A paper published from Ireland in 2015 presents a method for automated pavement condition assessment on airfield pavements using a laser crack measurement system (LCMS). Pavement condition expressed in the U.S. Army's Corps of Engineers PCI was compared for different data collection methods, i.e., manual walking surveys, visual surveys from forward view digital video, and LCMS 3D imagery. It is concluded that in general, there is a very good consistency between the type, quantity, and severity of distress data identified by these three methods.[30]

An NCHRP innovations deserving exploratory analysis (IDEA) final report published in 2015 outlines an inexpensive vision-based method for pavement defect detection, localization, and quantification. The report presents the design and development of the hardware (depth cameras, accelerometers, and an Inertial Navigation System) and software modules (data and hardware synchronization schema, and color and depth camera calibration). It also presents the field tests and data analysis that were carried out as part of the project and discusses the results.(31)

Results from a 2016 FHWA study funded by the Transportation Pooled Fund study, TPF-5(299) are presented in a series of task reports. These reports present development of standard data format for 2D and 3D pavement image data used to determine pavement surface condition and profiles.(32,33,34) The Task 2 report presents current practices, while the Task 3 report presents an evaluation of data items and formats. The Task 4 report discusses the development of metadata and also presents the proposed standards.

Part G of a technical method for highways titled "Automated Road Condition Assessments" and published by the South African National Roads Agency in 2016, presents guidance and methodologies on the planning, execution and control of automated road condition surveys using imaging technologies.(35) It is focused on pavement itself but incorporates other related right-of-way assets as well. The following systems are introduced:

- Frame imaging
- Scaled frame imaging
- Line scan imaging systems
- Range imaging systems
- 3D laser scanning

For each of these systems, devices, image validation, and operational/quality control procedures are outlined.

A presentation made at the TFP-5(299) meeting during the 2017 RPUG meeting discusses artificial intelligence in the context of digital image library ground truthing and the ongoing NCHRP project 1-57A (Developing Standard Definitions for Comparable Pavement Cracking Data). Challenges of cracking automation and ultimate objectives for true automation are discussed along with an image library for training and data augmentation. The presentation also includes a status briefing on the NCHRP 1-57A.(36)

A pavement condition data collection and processing request for proposals (RFP) advertised in 2017 by the Michigan Department of Transportation (MDOT) discusses the data requirements in detail. Information on data collection, data formats, data processing, and equipment requirements are provided.(37) It is a good recent example showing different components of a pavement condition survey RFP.

A paper published by the American Society of Civil Engineers in 2017 discusses a comprehensive, quantitative performance evaluation system for crack detection algorithms.(38) The paper proposes an approach to verification of automated image technology for use by agencies in evaluating/confirming the reliability of image identification. It provides an outline for this process. It goes on to evaluate several models for crack identification, using a scoring system offered to evaluate automated crack interpretation system accuracy.

Agency Information Gathering

The technical aspects of pavement management data collection have continued to evolve over time. Initially, agencies used personnel to collect data manually. This transitioned to video technology with raters working at an electronic workstation to identify distresses. By 2000, digital 2D image technology was developed and began to be adopted by agencies for use. The technology evolved from image capture cameras to line scan cameras. Early development was restricted by available hardware limitations on data processing speed and storage capacity. The available hardware improved over time, increasing the capability to collect and store pavement images at traffic speed. Within the last 10 years the use of 3D imaging equipment has become commonplace. This has significantly increased the possible applications of distress data from digital images, but work remains to take full advantage of the technology.

A review of relevant literature seems to suggest that when it comes to capturing images for crack detection, ongoing technological advancements have resolved challenges such as lighting conditions, image resolution, image storage, and compression as the current systems have continuously improved. For example, current imaging techniques can identify crack widths of 1 to 3 mm. The owner agencies' challenges according to the literature are establishing quality control, reliability, and crack definitions/interpretations compatible with image detection/ identification systems. That is, can the equipment adequately identify the cracking and subsequently, can the software system adequately classify and summarize the cracking distresses. In discussions with the vendors, their challenge is to provide summarized cracking distress information that is compatible with an agency's expectations based upon the agency's cracking definitions. The recent draft/unpublished NCHRP 1-57A Developing Standard Definitions for Comparable Cracking Data and NCHRP project 20-05 Synthesis Topic 49-15, Automated Pavement Condition Surveys, questionnaires and information were reviewed. Based upon this information, pooled fund study agency specific questions were developed with the intent of filling in missing information that is needed to fully develop a set of procurement guidelines.

The purpose of this information gathering effort was to help the project team define the data quality needs of State Agencies as a part of the process of developing appropriate criteria for equipment or vendor services selection. This project and the scope of these questions focus specifically on cracking distress and will aid in the production of guidelines for cracking assessment for use in the vendor selection process for network-level pavement crack data collection and analysis systems or services.

The following summary information was gleaned from 16 of the pooled fund study agencies, unless otherwise noted.

1. Type of system primarily used for network-level distress data collection on the NHS system?

   - ☐ Digital imaging
     - ☐ 2D System
     - ☐ 3D System
     - ☐ Other Digital system
   - ☐ Manual Crack (windshield/walking) Surveys
     - ☐ In-House personnel
     - ☐ Vendor staff
   - ☐ Other

Several agencies identified multiple systems being used in their agency, typically on different systems for some portions of their data collection; results are shown in Figure 1 for the 21 pooled fund agencies. As shown, 17 of the 21 agencies are currently using 3D systems. Those indicating manual surveys typically noted that these were on lower volume routes or for control section evaluations.

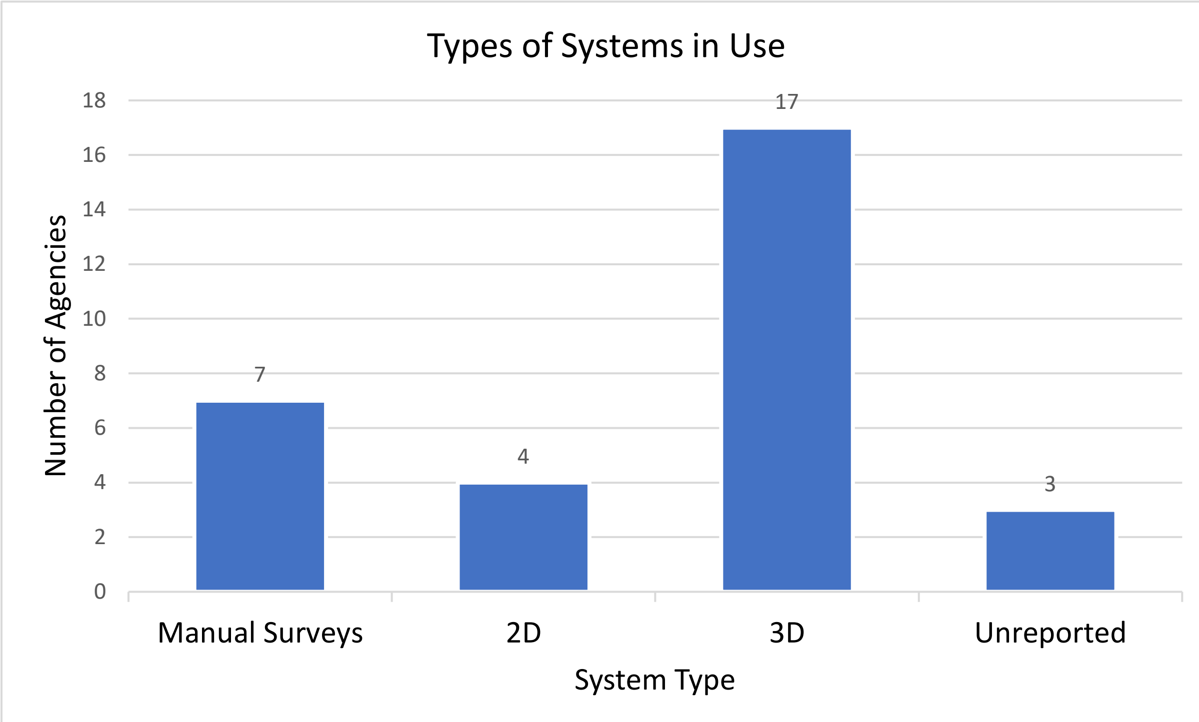


Figure 1. Type of system primarily used for network-level distress data collection on the NHS system.

2. When using digital technology for data collection, how is the cracking data being interpreted?
   ☐ Manually rating (reviewing and classifying the cracks by viewing the digital images)
   ☐ Semi-automated rating (rating is performed by computer processing of raw distress data, algorithms or other, and verified by manual checks)
   ☐ Fully automated rating (final distress classifications are produced by computer processing using algorithms or other methods).

As illustrated in Figure 2, over half of the agencies use a semi-automated crack classification process. Four agencies identified using a fully automated system. Those agencies that are using fully automated do not have data supporting the accuracy of the reported cracking data.

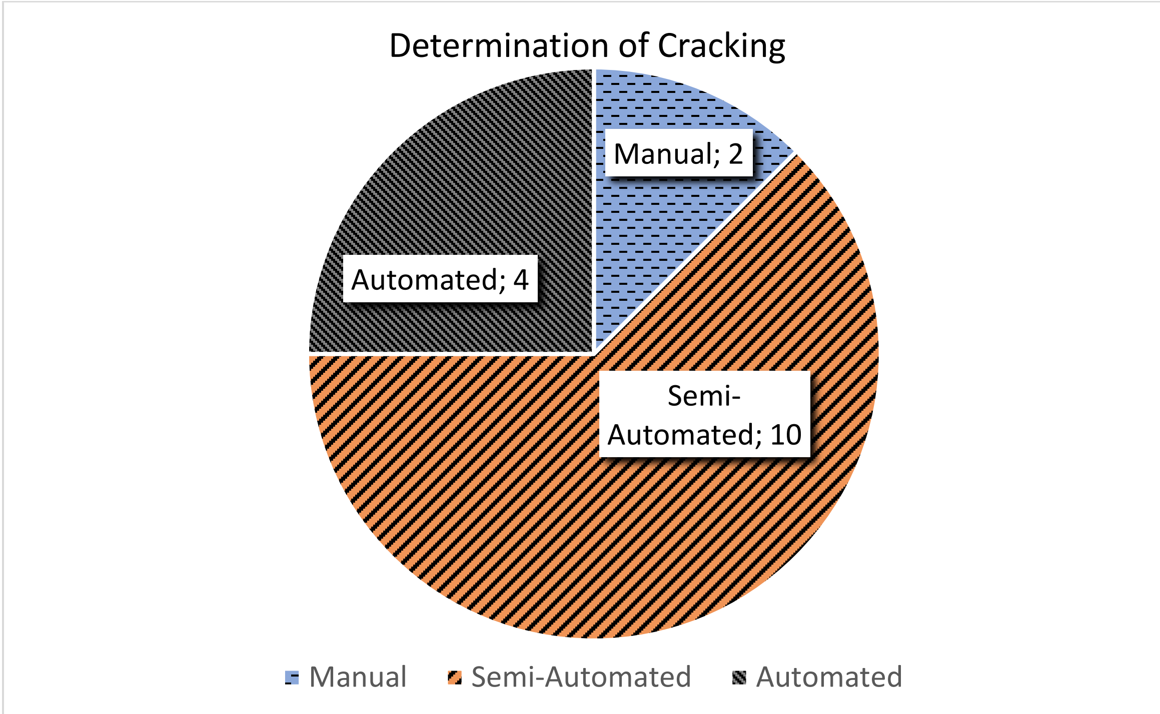


Figure 2. Indication of the method of determining cracking data from images.

3. Minimum crack width identification required as a criterion for vendor data collection equipment?
   - ☐ <1mm
   - ☐ 1 mm
   - ☐ 2 mm
   - ☐ 3 mm or greater
   - ☐ Other:

Nearly 1/3 of the agencies indicated that a minimum crack width of 1 mm is required, another 1/3 require a 2 mm crack width and the remaining 1/3 allow 3 mm or greater. This information is illustrated in Figure 3.

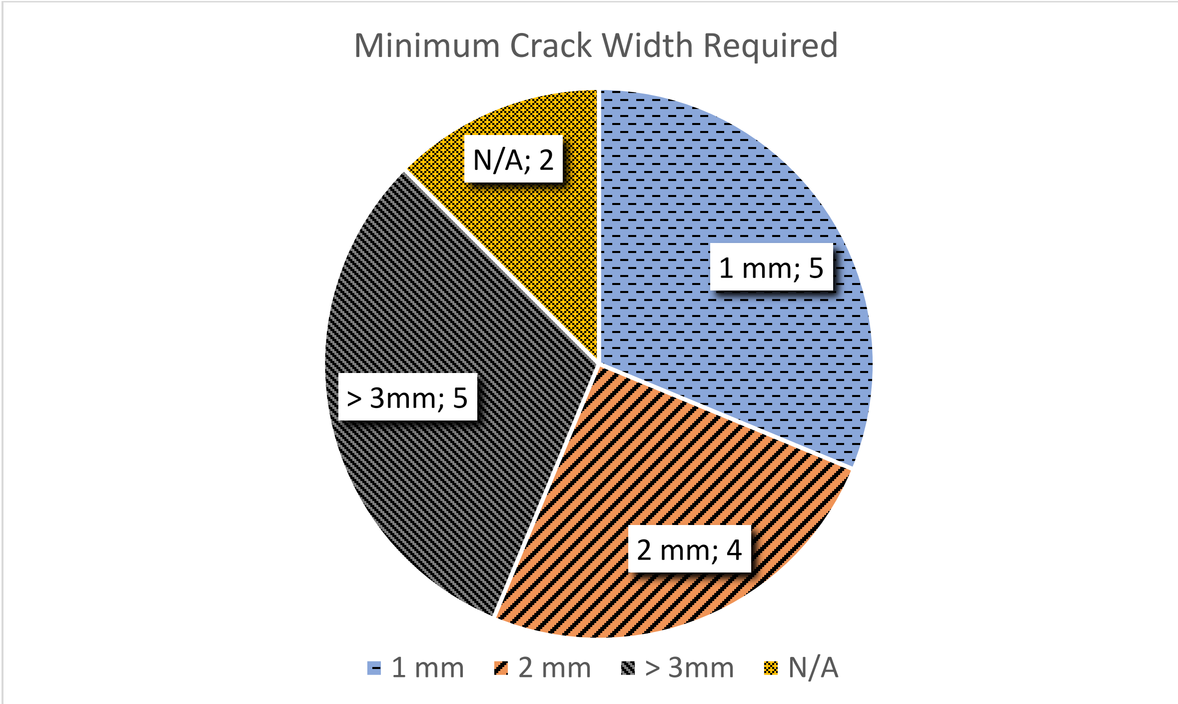


Figure 3. Minimum crack width identification required as a criterion for vendor data collection equipment.

4. Criteria currently used to define an acceptable digital image. (Example: Identify 1 mm wide crack, clear lighting and contrast, image size/width, etc.)

Criteria included statements such as images must be continuous, clear, color balanced, in focus, consistent illumination, stitched together properly and cover the full lane width. Small or low severity cracks must be visible.

5. When assessing data provided by the vendor for procurement purposes, field data collection/analysis needs to take into account cracking distress reference measurements. (Reference values are used as the basis for comparison of ratings by others to assess uniformity of interpretation.)

   Method used to determine reference values for cracking distress quality measurements?

   - ☐ Manual surveys on the roadway
   - ☐ Manual ratings from images
   - ☐ Vendor supplied comparisons
   - ☐ Other:

One agency is in the process of developing their reference criteria, and thus does not have a response. As illustrated in Figure 4, five agencies use manual ratings from the road, five agencies use manual ratings from images, three use both road and image ratings, and two reported other means.

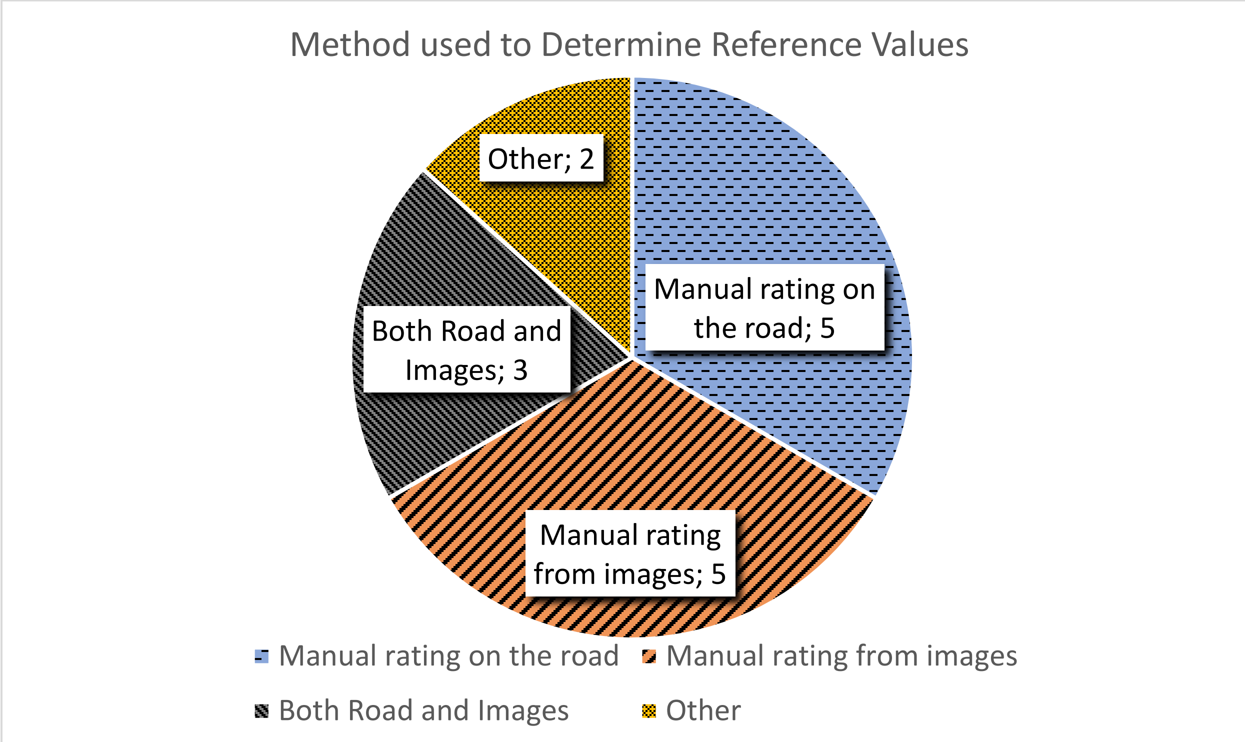


Figure 4. Method to determine reference values for cracking distress quality measurements.

6. It is possible that some agencies have tighter quality acceptance criteria for cracking in the procurement process as compared to production.

   a. Process followed to assure the reliability of production level cracking data.
      ☐ Training (all raters attend a single training event, or multiple events)
      ☐ Training and testing (raters are tested after training to verify that they meet the objective of data collection)
      ☐ Statistical comparison of rating results (a statistical evaluation of comparable data is completed to identify limits of acceptable data)
      ☐ Other:

Seven agencies use rater training. Five agencies use a statistical comparison of rating results.

   b. Method used to certify, accept, or calibrate production level cracking distress data. (Examples: control sites, blind sites, comparison to previous year's data, comparison of section-specific distress index values, comparison of results for individual distress types, etc.)

Nine agencies use a year-to-year comparison of cracking data. In addition, three agencies use control sections prior to data collection, two have blind sections where they compare the data, and eight agencies do some type of random check on the cracking data.

7. When defining what is acceptable data, either for procurement or production, a certain level of variability is to be expected.

a. Does the quality management process identify variability among distress raters (either for manual or semi-automated rating)? ☐ YES ☐ NO

If YES, what is the typical range of variation?

The results on this question were evenly split 50/50 between the agencies, as shown in Figure 5. For those that do identify variability, the range of variation reported was between 5 and 10%.

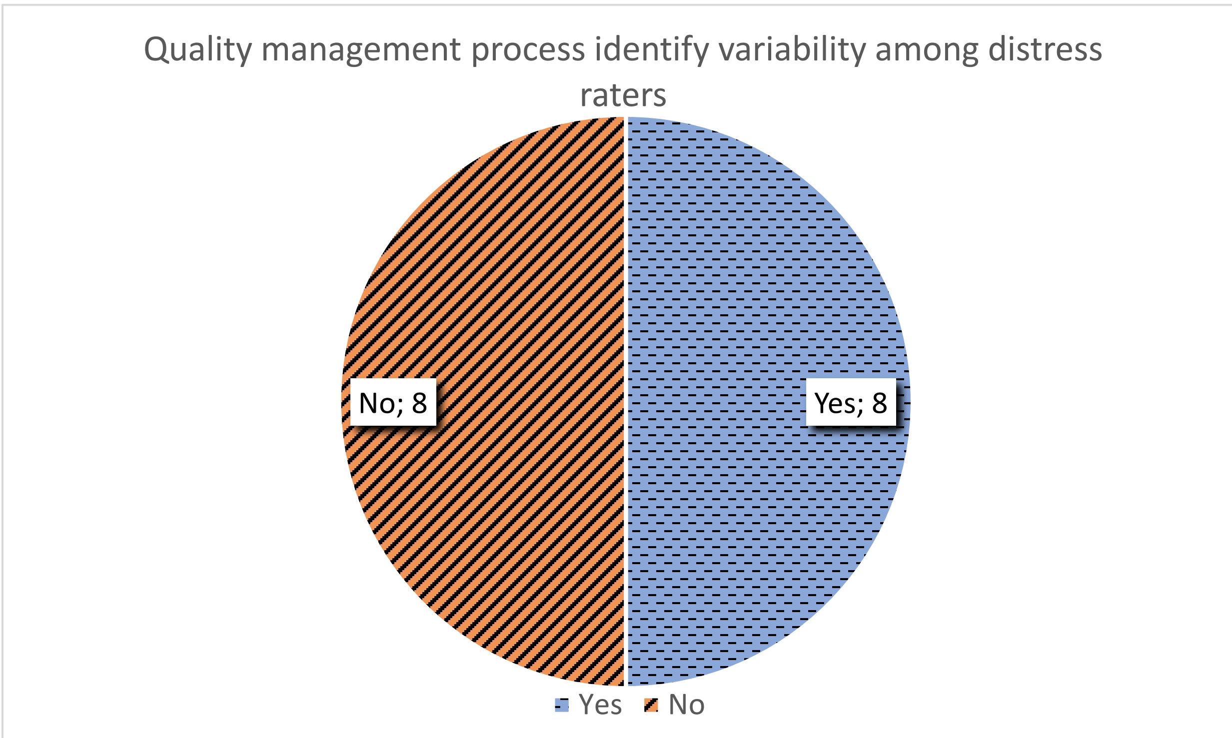


Figure 5. Indication of whether the quality management process identifies variability among distress raters.

8. Raters are assessed over time for consistency and comparisons are made between raters?
☐ YES ☐ NO

As illustrated in Figure 6 over half of the agencies do not assess raters for consistency over time.

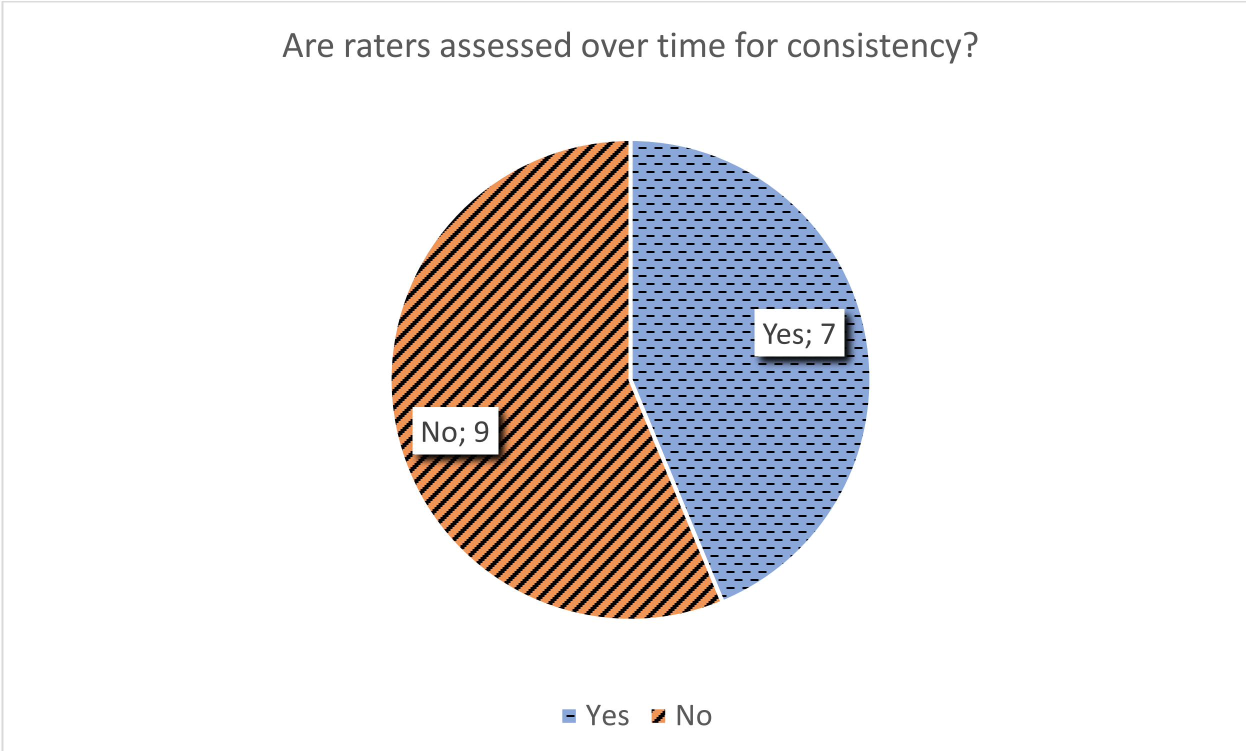


Figure 6. Indication as to whether distress raters are assessed over time for consistency and comparisons.

9. What is the allowable tolerance when accepting cracking distress data for input into the pavement management database? (Example: 90% of all cracking data must pass the quality monitoring check as defined by …)

As shown in Figure 7, seven of the eleven agencies that have this criterion require at least 90% of the data checks pass the acceptance criteria.

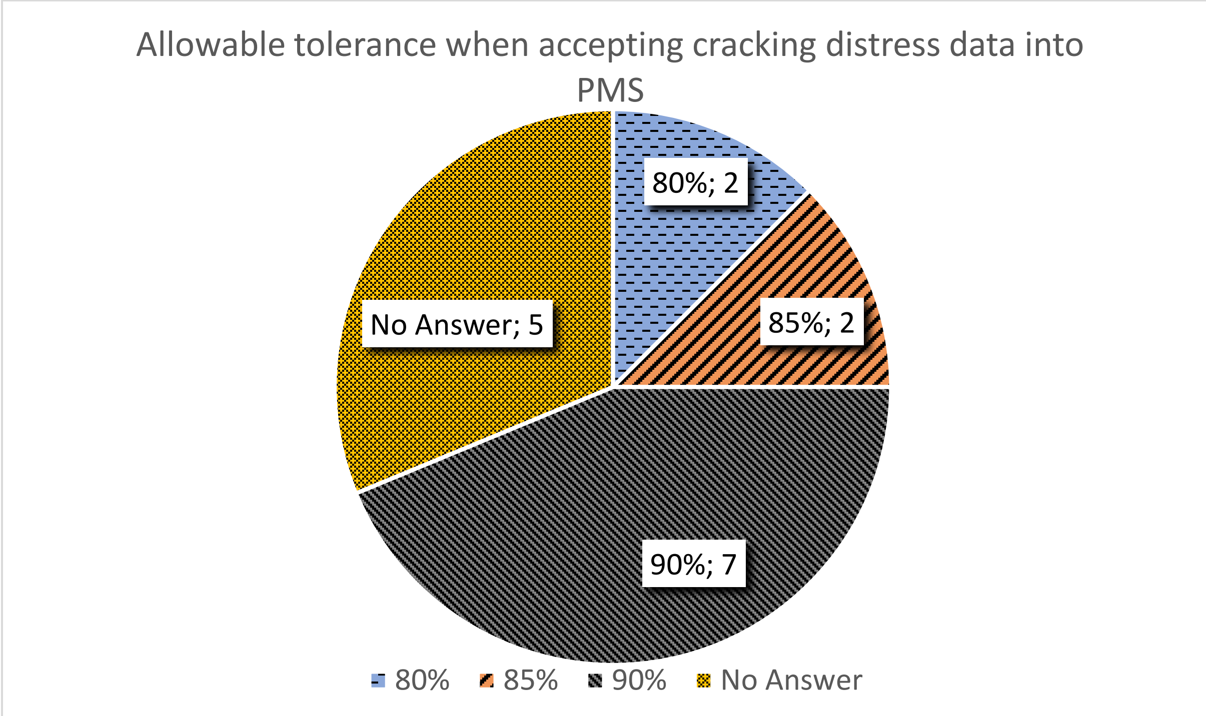


Figure 7. Allowable tolerance when accepting cracking distress data for input into the pavement management database.

10. Within most PMS systems, recommended treatments are often driven by certain critical distresses (distress type and/or severity level). Agencies identified the critical distresses that most frequently drive rehabilitation and maintenance recommendations.

Asphalt Concrete Pavements

☐ Longitudinal Cracking	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Transverse Cracking	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Block Cracking	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Patching	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Potholes	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Surface Deterioration (Raveling/Weathering/Polishing)	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Bleeding	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Joint Deterioration	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Rutting	☐ All severities	☐ Medium and high severity	☐ Only high severity

Information on fourteen of the agencies were compiled for asphalt concrete pavement. Nearly all identify fatigue cracking, transverse cracking and rutting as key decision making distresses. In addition, the majority identify that all severity levels are important, not merely the higher severities. Bars in Figures 8, 9, and 10 represent the number of agencies identifying specific distress-severity combinations as critical for rehabilitation and maintenance recommendations.

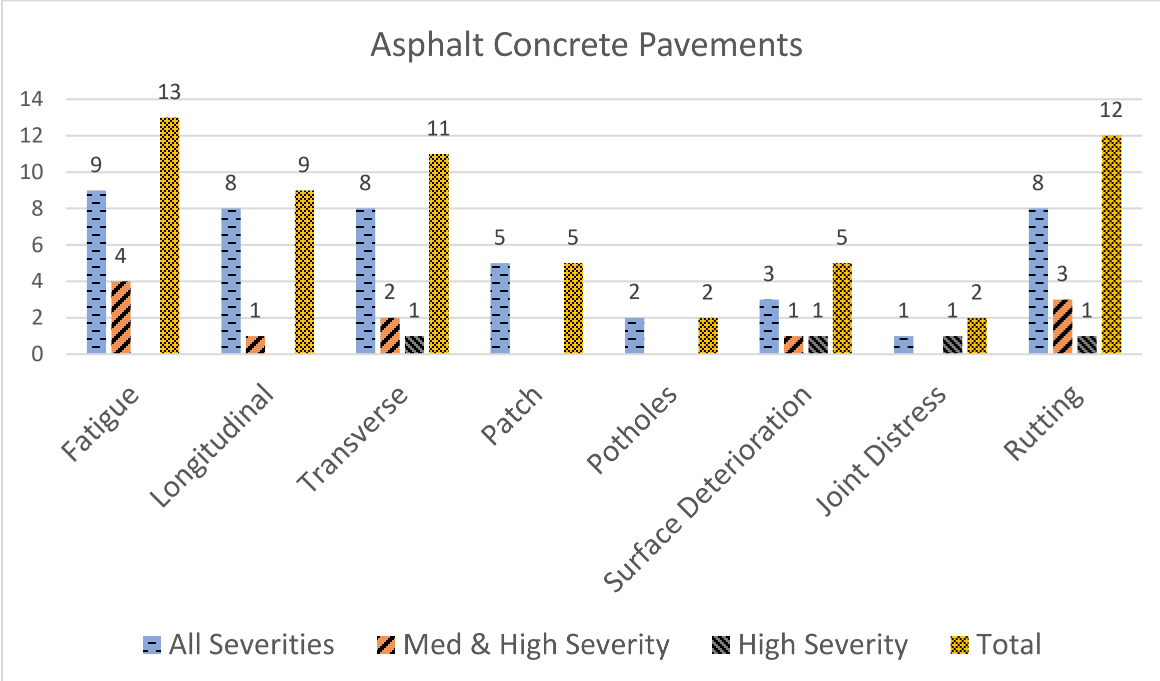


Figure 8. Key decision-making distresses for asphalt concrete pavements.

<u>Jointed Concrete Pavements</u>

☐ Corner Breaks	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Longitudinal Cracking	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Transverse Cracking	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Divided Slab	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Durability Cracking	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Joint Deficiencies (Seal damage, spalling)	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Surface Deducts (Map Cracking, Scaling, Polished Aggregate, Popouts)	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Faulting	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Patching	☐ All severities	☐ Medium and high severity	☐ Only high severity

Information on twelve of the agencies were compiled for jointed concrete pavement. As illustrated below all agencies identify longitudinal and transverse cracking as key decision making distresses. Over half of the agencies identify corner breaks and faulting as key factors as well.

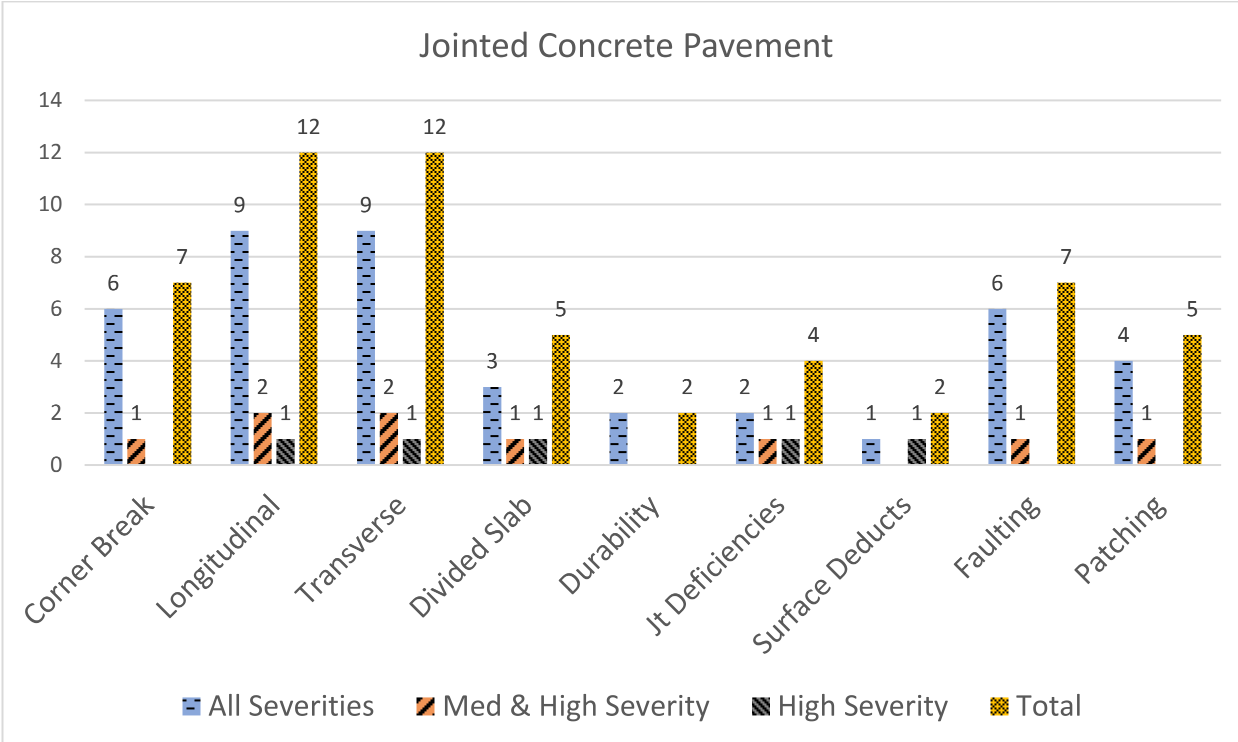


Figure 9. Key decision making distresses for jointed concrete pavements.

<u>Continuously Reinforced Concrete Pavements</u>

☐ Longitudinal Cracking	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Transverse Cracking	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Durability Cracking	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Surface Deducts	☐ All severities	☐ Medium and high severity	☐ Only high severity
(Map Cracking, Scaling, Polished Aggregate, Popouts)			
☐ Patching	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Punchout	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Cluster Cracking	☐ All severities	☐ Medium and high severity	☐ Only high severity
☐ Patching	☐ All severities	☐ Medium and high severity	☐ Only high severity

Information on seven of the agencies were compiled for continuously reinforced concrete pavement. Key distresses include longitudinal cracking and transverse cracking, followed by patching and punchouts. As with the previous pavement types, most agencies indicate that all severity levels are critical.

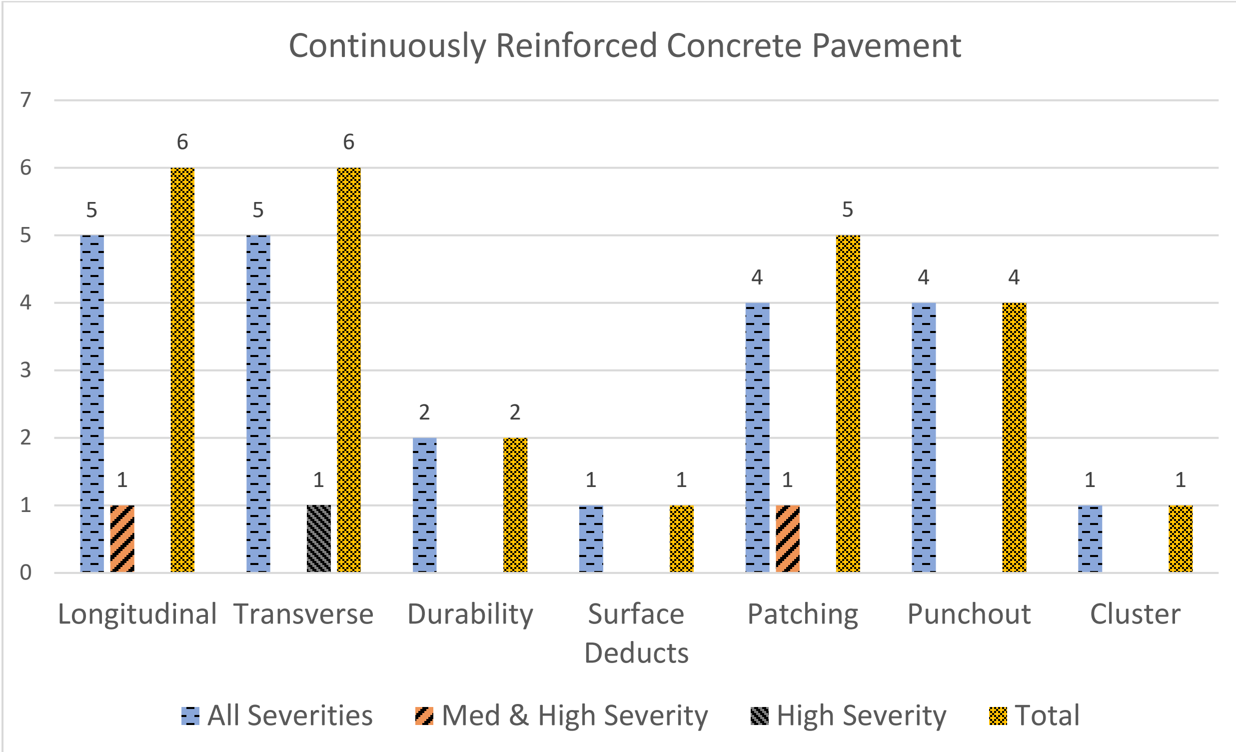


Figure 10. Key decision making distresses for continuously reinforced concrete pavements.

11. RFP requirements can vary widely, to include specifying specific equipment details, requiring site collection, and data delivery with the proposal. The pooled fund agencies ranked the following items in order of importance during the procurement process (number 1 is highest priority and 7 is lowest priority).
    a. (_____) Accurate crack identification
    b. (_____) Clear images that can be used for distress interpretation and/or quality monitoring
    c. (_____) Equipment specifications (e.g., 3D data collection system, automated classification, etc.)
    d. (_____) Proof of cracking meeting your requirements
    e. (_____) Proof of crack classification based on your agency's protocols
    f. (_____) Proof of crack classification based on industry standard cracking protocols
    g. (_____) Actual distress collection and delivery for your agency during the RFP period

    The results from 12 agencies are summarized in Figures 11 and 12, the first showing the average value (the lower the value, the higher the importance) and the second chart showing the mode value (i.e., the value that was most often reported).

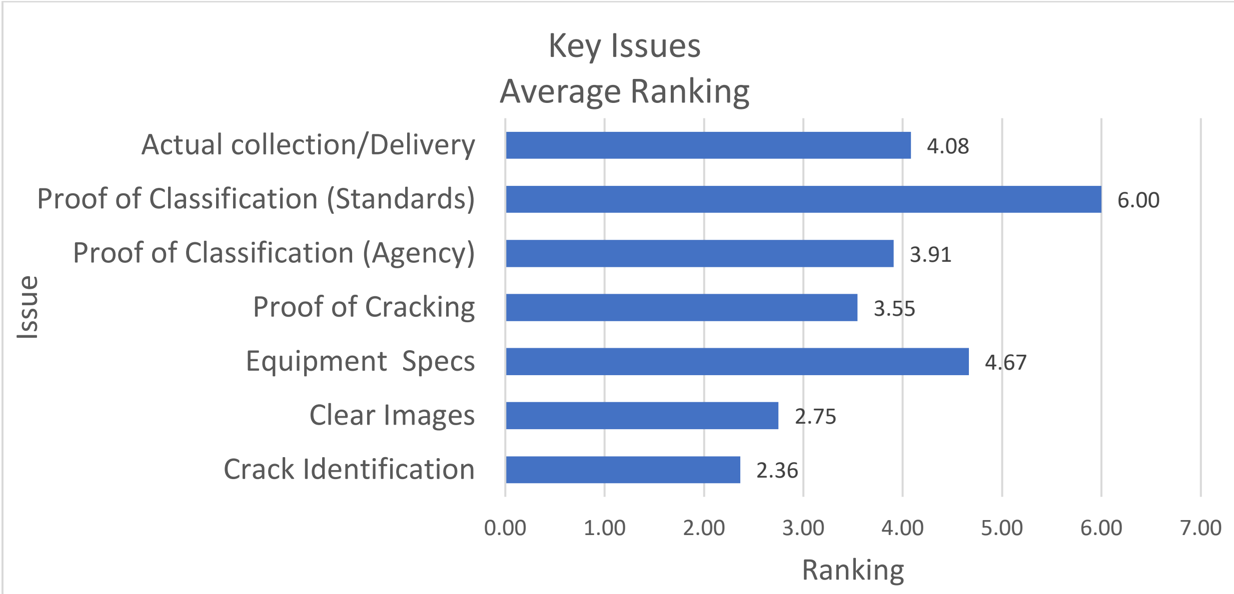


Figure 11. Average value of the key issues ranked in order of importance during the procurement process (number 1 is highest priority and 7 is lowest priority).

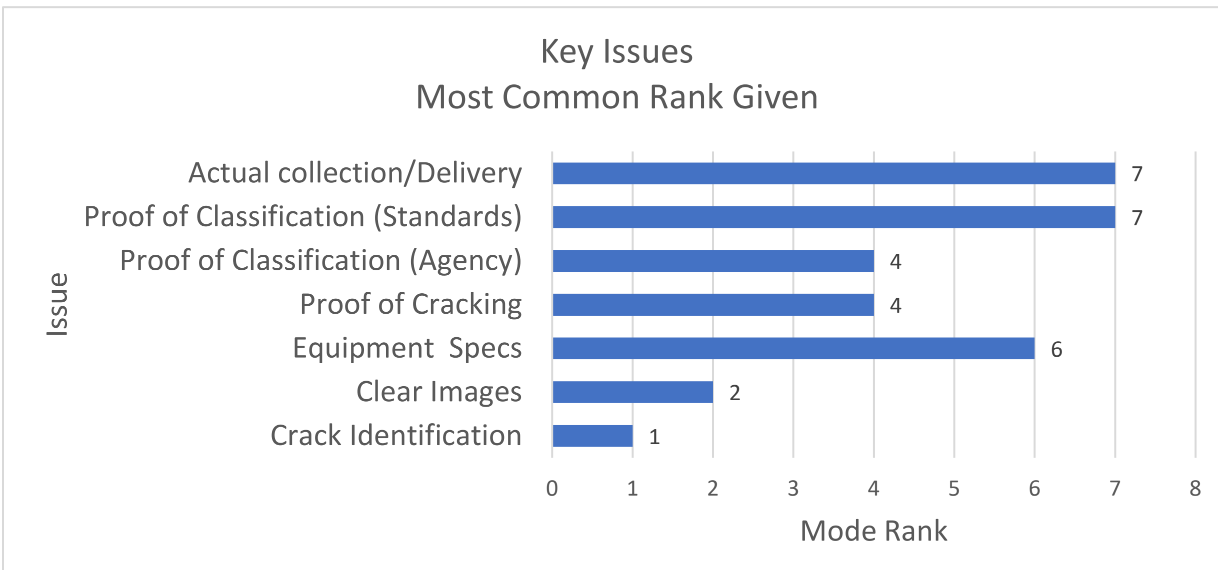


Figure 12. Mode value of the key issues ranked in order of importance during the procurement process (number 1 is highest priority and 7 is lowest priority).

In order to fulfill the objectives of this project there must be some standardization of the approach to distress definitions. At the September 2018 TPF-5(299) meeting, this issue was discussed and the consensus was to use the HPMS cracking criteria for the guideline template. Thus, the HPMS cracking criteria will be used as an example guideline, which agencies can adapt for their own specific distress definitions if they deem appropriate.

Several of the pooled fund agencies provided their distress identification guides, their Data Quality Management Plans, PMS decision trees, and most recent Request for Proposals for services or equipment. In addition, the pooled fund agency websites were reviewed for additional information.

When comparing the distress identification guides (see references 39, 40, 41, 42, and 43), the largest takeaway is that there is limited consistency between cracking definitions between agencies. These differences exist in both the cracking definitions, the severity levels and the method of reporting the distress. None of the distress guides reviewed address the reporting of the HPMS cracking criteria. This leaves interpretation of how the agency distresses get translated into the HPMS reported distress cracking percentages to the individual agency.

The provided data quality management plans were reviewed to gain insight into how the agencies are controlling the distress data reported, and specifically how the HPMS cracking percentages are reported. The HPMS data reporting requirements are mentioned in all six of the agency DQMP reviewed, however, none of them specifically addressed details for these submissions (see references 44, 45, 46, 47, 48, and 49). All agencies reviewed have a number of quality checks performed on the cracking distress data prior to acceptance by the agency, with most agencies requiring 90% or better match of the data.

The PMS decision trees were reviewed in detail for five agencies, specifically looking to identify common cracking distress measures which triggered maintenance and rehabilitation (see references 50, 51, 52, 53, 54, and 55). For asphalt concrete pavements, only one agency used individual distress types to trigger treatments (treatments are triggered by alligator cracking, transverse cracking, rutting, patching, bleeding and raveling). This agency also used a composite index which is comprised of cracking, both wheel path and environmental, rutting, ride, raveling, and bleeding. Two additional agencies use index values that trigger various treatments (which include alligator cracking, transverse cracking and patching) and one agency is only based upon the ride quality (IRI value). For jointed concrete pavements, all agencies used an index specific to their agency, which are typically based upon the number of slabs with distress to include transverse cracking, longitudinal cracking, corner cracking, spalling and patching. Based upon this review, it is clear that the agencies are using similar, yet different distress measures in the determination of maintenance and rehabilitation recommends.

Recent (within the last three years) RFPs from nine agencies (seven of them were pooled fund agencies) were reviewed (see references 56, 57, 58, 59, 60, 61, 62, 63, and 64). The requirements in the RFPs varied significantly. Three agencies required previous proven state experience which ranged from 3 to 10 years. Five agencies held a mandatory preproposal conference and three agencies required preproposal data collection and reporting efforts to be completed, which ranged from 8 to 200 segments of data (4 to 100 miles of data collection). None of the agencies identified specific specifications of the distress camera's needed for data collection, although two did specify the system to be a 3D system. Most agencies had a required camera resolution specified for the forward image camera(s), but not for the downward camera. All agencies defined the cracking that was required to be reported, most often by reference to their Distress Rating Manual or protocols. Five agencies required the vendor to pass a control site evaluation step after contract award. The three agencies that required preproposal data collection and submission did not indicate that additional control sites would be collected. All of the RFP's reviewed included mention of the HPMS data collection and reporting, although not in much specific detail. An example of a typical HPMS reference is "These data items shall conform to the latest version of HPMS Field Manual. Cracking quantities in HPMS units will be part of the pavement distress evaluation requirements."

Vendor Information

Discussions were held with several vendors to obtain information regarding their experience with the procurement process and suggestions for improvement. The following are highlights of these conversations.

- Vendors are concerned with the time and effort expended in collecting, rating, and processing pre-award sample sets.

- Vendors expressed concern in the time and effort to understand agency cracking protocols during the procurement process, stating that the task often takes a couple of data collection cycles once under contract to get it correct, therefore it is unrealistic to expect a vendor to fully understand the agency definitions and delivery requirements during the procurement phase.

- The previous bullet leads to a potential advantage of incumbents on re-bids due to express knowledge of cracking and delivery protocols.

- Vendors indicated they perform testing, verification, and validation of each system in-house before sending equipment to collect data. Each agency then requires their own set of validation and verification steps, which is seen as a duplicate effort.

- Often the procurement and contract language are cumbersome to understand.

- Many of the tolerance limits identified in the procurement documents, and subsequent contracts can be unrealistic, or unjustified, but continue to be included.

Summary of Findings

Based upon the literature reviews and the information gathered the following statements can be made:

- Having standard data collection procedures would make the procurement process simpler for both agencies and vendors.

- Agencies have identified a need for improved quality management (quality assurance, quality control, and reliability).

- The technology for capturing high quality images is here, but still not readily validated and continues to evolve.

- Vendors are transitioning from 2D to 3D systems whether or not the agencies are requesting this change.

- For procurement purposes the industry standard is 3D.

- More than one-half of the agencies are using semi-automated crack detection methods. A few indicated they are using fully automated crack detection and a few require manual interpretation of the cracking data from the images.

- The specified minimum crack width required varies among agencies.
- Agencies see value associated with comparing the automated or semi-automated cracking data to cracking data from manual road surveys. This is primarily done at project initiation using reference sites.
- Agency and vendor understanding and application of rating variability remain uncertain.
- On asphalt concrete surfaces, fatigue, transverse and longitudinal cracking, and rutting are the primary decision making factors, generally at all severity levels.
- On jointed concrete pavements, transverse, longitudinal, corner cracking and faulting are the primary decision making factors, generally at all severity levels.
- On continuously reinforced concrete pavements, punchouts, longitudinal and transverse cracking are the primary decision making factors, generally at all severity levels.
- When agencies were asked to prioritize several potential procurement items, accurate crack identification was rated as the highest priority item. This was followed by the need for clear images, which was followed by proof of the cracking meeting an agency's needs being identified. The lowest ranked option was proof of the classification of the cracking based on standard protocols.

HPMS Data Reporting

The preliminary results of the information gathering and concepts of data quality management were presented to the pooled fund agencies at the TPF-5(299) meeting on September 18, 2018. Responses from the discussions, and subsequent confirmation from the FHWA Technical Point of Contact was that the project objective should be to provide recommendations based on the HPMS Field Manual reporting criteria.[65] Since all agencies are requested to submit cracking information in accordance with HPMS it is reasonable that the agency procurement requirements would require the vendors to demonstrate that ability. Given the near term product schedule for this project, discussion during the meeting, and information in the literature about future technical developments, clarification was requested regarding how the image quality parameters should be framed for this purpose. It was determined that the recommendations must be immediately implementable, so discussion of future developments is contained in a separate section of this report.

The MAP-21 reporting requirement includes a cracking metric as defined by the HPMS cracking percent. To date, this has represented a challenge to both state highway agencies and distress data collection vendors. Each state must find a way to compute this cracking percent from the data being obtained from their pavement management data collection process, or specifically collect this information. Likewise, data collection vendors asked to provide the HPMS cracking percent face the challenge of computing this measure from the pavement management cracking information they are asked to report to each state. Since each agency has its own unique set of distress information and distress definitions, this calculation is often unique for individual agencies, and is created by translating the pavement management distress data into the HPMS cracking percent. At this time, some agencies are not requiring the vendors to specifically collect HPMS cracking criteria, but rather are asking the vendor to translate the HPMS data from the PMS collected distress data. This currently requires the vendor to use a custom process for computing the HPMS cracking statistic for each state.

The 2016 HPMS Field Manual requires that the percent cracking be continuously collected in a manner that will allow for reporting in nominally uniform section lengths of 0.1 mile (528 ft.); shorter sections are permitted only at the beginning of a route, end of a route, at bridges, or other locations where a section length of 0.1 mile is not achievable; the maximum length of a section shall not exceed 0.11 mile in length.

Asphalt Concrete Pavement Surfaces

For asphalt concrete pavement surfaces, the Cracking Percent is the percentage of the total area exhibiting visible fatigue type cracking for all severity levels in the wheel path in each section.(65)

The Manual goes on to define wheel path as follows:

For purposes of reporting cracking data to HPMS, the wheel path width is to be 39 inches (1.0 m) and located as described in the Standard. Regardless of the method of data collection, the percentage of cracking to be reported is the total area of the wheel paths where cracks are detected divided by the total area of the 0.1 mile section.(65) An example is provided in the HPMS Field Manual illustrating this calculation. In essence, the linear length of cracking along both wheel paths is needed.

To calculate this metric it is necessary to have two areas, the total area of the sample section, and the total area of wheel path fatigue cracking. The total area of the sample section is simply the product of the length and width of the section. Potential differences in the reporting of this area of fatigue type cracking in the wheel path come when an agency measures the actual area of fatigue cracking in the wheel paths as compared to when an agency measures the length of wheel path cracking and then multiplies this length by the entire width of the wheel path. As discussed above HPMS defines the wheel paths to be 39 inches (1 meter) wide. In other instances, an agency may use a standard reported width of cracking for various fatigue cracking severities which is then multiplied by the length of cracking. There may also be differences in the type of wheel path cracking that an agency considers as "fatigue type cracking." Some agencies will include any cracking that falls within the wheel path, to include longitudinal, fatigue, transverse, block and miscellaneous cracking. In further discussion with FHWA staff, they have clarified that all cracks in the wheel path are to be included. In order to standardize the reporting for these guidelines the following will be requested:

- Vendor is to report the lineal length of cracking in each wheel path (all cracks should be included).
- The length of cracking from each wheel path is multiplied by 39 inches, and converted to square feet.
- Percent Cracking is the calculated area of wheel path cracking divided by the total area of the sample section multiplied by 100. The Cracking Percent is to be reported to the nearest one percent.

JCP Cracking

The procedure described in the HPMS Field Manual provides the following definitions regarding reporting of the cracking percentage for JCP pavements.

The percentage of cracking reported is calculated as the number of slabs containing one or more transverse cracks extending for at least one-half the lane width, divided by the total number of slabs in the section. [(65)]

This can be challenging for data collection and analysis systems due to the other distresses and features of jointed concrete pavements. The HPMS manual states that reported cracking for jointed concrete pavements excludes longitudinal cracks, corner breaks, D-cracking, and Alkali Silica Reactivity (ASR) cracking that may occur on a slab.

In the case of jointed concrete pavements, the Cracking Percent is the percentage of slabs within the section that exhibit transverse cracking extending for at least one-half the lane width. For adjoining sections that share a slab, the section that contains the majority of the slab length will include the count for that slab. Two key variables contribute to the potential differences in reporting methodology; determining the total number of slabs and determining which slabs exhibit transverse cracking extending for at least one-half the lane width. The first item can be simplified by reporting the actual number of slabs in the reporting section (generally 0.1 miles in length). Historically, some agencies calculate the total number of slabs by dividing the length of the section by the typical joint spacing. Other factors that can contribute to the difference in the number of slabs is how full slab replacements and/or joint repairs are reported. Are they counted as patched slabs or counted as new slabs? Often a portion of an existing slab may have been replaced which now effectively created two slabs where only a single slab originally existed. Regarding the determination of cracked slabs, HPMS defines a cracked slab as a slab containing one or more transverse cracks extending for at least one-half the lane width. The majority of agencies in their standard procedures report the number of slabs with transverse cracking, however, they don't always limit these counts to transverse cracks that are at least ½ the lane width. There are a few agencies that report the number and length of transverse cracks in a section, rather than reporting by slab. In order to standardize the reporting for these guidelines the following will be requested:

- The actual number of slabs be identified and reported for the sample section (this includes all full width concrete patches or slab replacements)
- The number of slabs with transverse cracks at least ½ the slab width also must be reported.
- The Cracking Percent is then simply the number of slabs with transverse cracking divided by the total number of slabs in the reporting section multiplied by 100.
- When a partial slab is at the beginning, end or both of a section, the section with the largest portion of the partial slab should include the count for that slab.

HPMS states "Cracks in the Concrete Slabs may be detected using manual observations, imaging, or other methods that identify at least 85% of all cracks present in the slabs."

CRCP Cracking

For continuously reinforced concrete pavements (CRCP), the Cracking Percent is the percentage of the area of the section exhibiting longitudinal cracking (any severity), punchouts, and/or patching. Transverse cracking shall not be considered in the Cracking Percent for CRCP. Percentage of Cracking for CRCP pavements is determined as the area of pavements where cracking or distresses are detected divided by the total area of the section. For longitudinal cracking, the cracked area is determined as the length of the crack multiplied by one foot width. For punchouts, the area is determined by the two transverse cracks and the edge of the pavement or longitudinal joint, this includes the area of a spalled Y-crack. The total area of the pavement section is determined by the product of the lane width multiplied by the length of the section.

From the description of cracking percent described above it is clear that multiple distress types are to be considered in this HPMS criteria. It can be reasonably assumed that agencies are identifying punchouts, longitudinal cracks, and patches as a part of their distress surveys; however, they may not currently be reporting the area of punchouts or patches. In order to standardize the reporting for these guidelines the following will be requested:

- Report the length of longitudinal cracking and multiply by one foot to convert to an area
- Report the area of punchouts
- Report the area of patching

Percent Cracking is the summed area of longitudinal cracking, area of punchouts and area of patching divided by the total area of the sample section multiplied by 100.

HPMS states "Cracks and related distresses in the CRCP pavement surface may be detected using manual observations, imaging, or other methods that identify at least 85% of all distress present in the surface."

FUTURE DIRECTIONS

Industry continues developing and enhancing technology for fully automated distress data collection and analysis. For the purposes of this effort, it has been assumed that for the immediate future most agencies will be using some combination of manual ratings and manually assisted ratings of captured images to obtain cracking distress data. However, at some time in the future, the use of fully automated systems seems feasible. So, how will this change the process needed to prequalify vendors and vendor provided systems? The elements of the recommended process will still remain valid. Cracking distress definitions will still be applicable. It will still be necessary to develop the reference values to be used for cracking distress interpretation comparisons.

One example of the application of automation is the work of Dr. Tsai at Georgia Tech.

In order to represent the real-world complex crack patterns, Tsai et al. proposed the multi-scale Crack Fundamental Element model to provide rich and systematic crack properties for crack classification. [24] This model provides a methodology for interpreting cracking from images. The process can be simplified by standardization of distresses and distress definitions, but can be applied to individual agency procedures. In another report Dr. Tsai demonstrates an evaluation of crack pattern recognition. This may be useful in distinguishing individual types of cracking distress. [22]

Dr. Tsai has also developed a methodology for evaluating the reliability of distress images collected with automated equipment. This concept will be useful in determining the acceptability of images produced by various vendors. To achieve this objective, it will be necessary for all vendors to provide data in a format which can be interpreted by others. Thus, a universally accepted format will be needed.(38)

Dr. Wang, Oklahoma State University, continues his work on an artificial intelligence system that shows promise and is based upon deep learning concepts that include Generative Adversarial Networks which are a class of artificial intelligence algorithms used in unsupervised machine learning that are implemented by a system of two neural networks contrasting with each other in a zero-sum gain framework. These systems can generate photographs superficially authentic to human observers. Dr. Wang claims to have achieved a fully automated crack detection system with the 2018 version of CrackNet and anticipates that a processing speed of 60 miles per hour may be achieved in 2019.(66)

OUTLINE FOR GUIDELINES FOR VENDOR SELECTION

The guidelines developed for vendor selection will be based on the following key items:

- Illustrate that the system can produce high quality digital distress images that can be used to identify and classify cracking.
- Illustrate that the system can produce cracking distress summaries meeting the HPMS standard.
- Illustrate that the system is robust enough to achieve the above criteria at production levels.
- Develop a standard that can be used for agency procurement.
- Provide the framework of the standard such that an agency can modify to meet their specific needs, if desired.

Two Step Protocol Process

The recommendation is to provide a two-step process in the procurement guidelines. The first step is to verify that the equipment is capable of collecting images from which the cracking can be identified to a specified level. The second step is to then verify that the necessary cracking distresses can be quantified to the FHWA HPMS reporting standard and/or to meet an agency's needs.

Equipment Verification Requirements

1. Imaging System Clarity
   It is anticipated all collection will soon be based on 3D imaging systems. Therefore, the recommendations provided will primarily focus on this type of system. Although either line scan or frame type digital cameras can be used to collect pavement images, the line scan type has proven most useful and is generally expected to be used for pavement distress data collection. The use of frame type cameras can result in some distortion of the image along the image edges, whereas the line scan camera produces a series of single pixel images stitched together to provide the second dimension, similar to a fax machine, and therefore does not have this issue.

a. Image size – pixels by pixels
   For a 3D image 4,096 pixel transverse resolution will theoretically support the identification of a 1 mm wide crack. Similarly 1,300 pixel transverse resolution theoretically supports the identification of a 3 mm wide crack, and 2,048 pixels a 2 mm crack width. These theoretical resolution levels are best achieved when the camera is still or moving slowly. Finer cracks can be identified by higher resolution cameras. For 3D pavement imaging, 16-bit images are recommended. Crack identification is also affected by the 256 shade levels in an 8-bit image, making crack width identification more complex.[20,32] The vendor should be requested to provide their proposed image size.

b. Image dynamic range check
   Dynamic range is the ratio between the brightest to darkest signal levels. It determines how many levels of difference in digital values exist in a given image. For binary (black and white) images, 8-bits (256 levels) is usually sufficient to represent visual information. Images of 16 bits are the minimum required to handle 3D images, with some systems using as much as 32 bits. Color images with dynamic range of 24-bits are available, and may be needed to capture detailed features such as surface texture. The vendor will be requested to provide their image dynamic range factor.

c. Percent fill factor
   Pixel fill factor indicates the light gathering area of the photo detector being used. The proportion of the pixel area insensitive to light is indirectly indicated. The minimum recommended fill factor is 90%. This indicates that 10% of the pixel is insensitive to light. The photo detectors commonly used are most often a silicon chip (CCD type) or a metal oxide semiconductor (CMOS type). A lower fill factor may be acceptable if the resulting image quality is sufficient to identify the desired distresses.

2. Image Resolution
   3D line scan data collection produces higher resolution, dynamic range, and higher fill factor than earlier technologies. It also reduces blurred images of fast moving objects. The application of the 3D range data with laser image lighting has resulted in a major reduction in image clarity issues from variable lighting conditions and improved interpretation of poor intensity contrasts such as oil stains on the pavement surface.

   According to the referenced report by Dr, Tsai, the Georgia Tech 3D system is able to identify a 1 mm crack at low speeds, and a 2 mm crack when images are collected above 60 mph. [22] Current agency RFP requirements range from 3mm to 1mm systems, with no clarifications on how that is measured or evaluated. These guidelines recommend a system capable of producing images where a 1mm crack is visible when confirmed at slow speed and/or a 3 mm crack at speeds exceeding 60 mph.

3. Compression Type And Storage Size Requirements Vendor Uses To Meet Data Delivery
   Currently, the size of the original data collection files and compression method used varies from one software developer to another. However, FHWA is in the final publication phase of a report titled "Evaluation of Proposed Standard Data Format and Compression Algorithms for 2D/3D Pavement Surface Image" [67] which proposes a standard data format for the 2D/3D image systems. An agency may consider requiring the 2D and 3D images provided by the vendor comply with the proposed standard data format. An example of the current minimal rules and validation procedures are listed in Table 1.

Table 1. Example of minimal rules and validation procedures.

Properties	Sub-rules	Validation Procedure
File Integrity	The file signature is present	Check if the last four bytes of the file is "psi."
File Integrity	The file trailer is present	Check if the last four bytes of the file is "@@@@."
File Integrity	The file's checksum equals to the given one	If a checksum is given, calculate the checksum based on the file content and check if it equals to the given checksum.
Header Correctness	The values in the required header fields are valid	For each value in required fields, if the field takes only assigned value, check if the value is in the "assigned values list". For example, version must follow the format "X.YY" where X and YY are numbers.
Header Correctness	The size of the 2D/3D data is correct	If the data are not compressed, check if the following condition holds: "datasize = bitdepth / 8 * width * length"
Data Correctness	The data in the 2D and 3D sections can be extracted using header information	Extract and decode the 2D and 3D data using header provided information. Check if the extracted data can be fit into a width * length matrix of that given data type.

4. Image capture width
   The width of image should cover the entire driven lane in the travel direction, accounting for vehicle wander, which is typically 13.5 to 14 ft. wide.

5. Cracking distress interpretation type
   Definitions:

   - Manually rating (reviewing and classifying the cracks by viewing the digital images)

   - Semi-automated rating (rating is performed by computer processing of raw distress data, algorithms or other, and verified by manual reviews of the processed distress)

   - Fully automated rating (distress classifications are produced by computer processing using algorithms or other methods, with little to no human intervention)
     Software algorithms are generally used for automated interpretation of cracks and the creation of crack maps. Algorithms filter images, removing noncracking data with the remaining crack information in binary form. Noise is removed, and cracks linked as considered appropriate. 3D imaging provides increased image resolution, and should support improved crack interpretation using automated interpretation techniques. In some cases, 3D data collection images are converted using 2D tools for interpretation.

   Work continues on the refinement of technology for a fully automated distress interpretation process. However, at this time it is not considered ready for full implementation for all distress types without additional post processing or manual assistance. Consequently, it is assumed that for the near term, semi-automated or manual interpretation from images will be utilized. An agency can choose which of the three cracking distress interpretations they desire to specify.

6. System capabilities
    a. Illumination source – The Downward Perspective Image shall be collected with a uniform and consistent form of illumination applied to the pavement surface. The illumination shall be regulated to provide sufficient contrast and crack-shadows for the clear discernment of cracking and patching. Images bearing ambient and/or vehicle shadows that obscure pavement features will not be accepted.

    b. Data collection speed – Data should be collected at or near prevailing highway speed, typically 25 to 65 miles per hour.

    c. Required system storage – The file size requirement is partially determined by the size of data stream collected and compression ratio used. For example, 1-mm resolution imaging for a 4 meter lane width requires a data flow of approximately 120 MB per second prior to compression [20]. Data collection systems differ as to whether data is processed in real time or post processed. Therefore storage requirements can be determined as a part of the system used, contingent upon being able to meet the requirements for image clarity and interpretation of the information stored in the retrieved image.

       The size of storage required is also significantly affected by requirements to collect full survey data vs. sample data. General requirements should be that the data storage file be self-describing and self-contained. Self-description here means that the data can be interpreted by different systems at different points in time. Self-contained means that all data needed for interpretation is included within the data "container." For example, metadata will not only provide the identifying information such as location and date, but also a string of information linking the header to all the accessible data stored for that specific location. System storage capacity need not be verified by the agency, rather the acceptability of storage capacity and compression taken as a whole can be determined based on the final image quality produced by the system.

7. Define a quality image
   Until now individual agencies have developed their own definition of image quality, and tests to verify that their PMS data is derived from quality digital images of the pavement surface. One objective of this study is to provide recommendations for determining image quality which can be uniformly applied by all agencies. Based upon the literature review, other industries, such as the medical field and textile industry, have identified quality images. However, the major difference is that they are looking for images to be identical when compared to another image. For pavement surfaces, this is never the case and as such the criteria used in other fields does not seem to be practically adaptable.

   Image quality should provide the following characteristics: pavement surface is visible without shadows, reflection from wet surface, or other conditions of the imaging process resulting in images which cannot be clearly viewed. This cannot address obstacles which obscure the pavement surface, such as dirt or other contaminants. In the discussion of image resolution, it has been stated that for a 3D imaging system, a crack width of 1 mm should be identifiable (for a stationary or low speed system), with a 2 mm width identifiable when the image is collected while the collection vehicle travels at 60 mph. The use of lasers has virtually eliminated problems related to shadows and damp pavement surfaces. A wet pavement surface can still provide an obstacle to successful identification of cracks, and therefore must be controlled as a part of the data collection QC process. As previously discussed, the image size must be sufficient to accommodate some vehicle wander while data is being collected. It is not practical to totally avoid vehicle wander, particularly when traveling on an active highway.

8. Validation of Image Quality
   The vendor should be requested to provide images and viewing software to the agency for the validation of image quality. Images from the reference sites (described later in this report) will be evaluated by the agency for conformance with image quality. The following guidelines are taken from the Standard Practice for Collecting Images of Pavement Surfaces for Distress Detection, AASHTO Designation R86-18, and are proposed as a reference for use in the guidelines:

   - The images must provide sufficient difference between data point values representing distressed and non-distressed areas that, subsequently, distress detection techniques can delineate a minimum of 33 percent of all cracks under 3mm (0.12 in.), 60 percent of all cracks present from 3mm (0.12 in.) and under 5mm (0.2 in.) wide, and 85 percent of all cracks 5mm (0.2 in.) wide or wider regardless of orientation or type. The determination of this capability will be made utilizing a minimum of ten 0.03-km (100-ft.) samples containing an average of at least five such cracks per sample.

   - The images should be sufficiently void of erroneous differences between data point values that a section of pavement without distress, discontinuities, or pavement markings contains less than 3m (10 ft.) total length of detected false cracking in $50m^2$ ($540 ft^2$) of pavement. The determination of this capability will be made utilizing a minimum of ten 0.03-km (100-ft.) samples of various types that meet the criteria.

   - Average crack width for each crack detected in the previous bullet must be within 20 percent or 1mm (0.04 in.), whichever is larger, of the actual width with at least 85 percent confidence.

Cracking Distress Data Verification Steps

1. Identification of Cracking Distress Types to be Verified
   The HPMS Field Manual defines cracking as "a fissure or discontinuity of the pavement surface not necessarily extending through the entire thickness of the pavement." The AASHTO R86 "Standard Practice for Collecting Images of Pavement Surfaces for Distress Detection" defines a crack as "a fissure of the pavement material at the surface with minimum dimensions of 1mm (0.04 in.) width and 25 mm (1 in.) length. The HPMS reporting criteria that require the reporting of the Cracking Percentage for Asphalt Concrete, Jointed Concrete, and Continuously Reinforced Concrete are as follows:

   - For Asphalt pavements, Cracking Percent is the percentage of the total area exhibiting visible fatigue type cracking (any cracking in the wheel path) for all severity levels in the wheel path in each section.

     - Cracking should be measured and reported for both wheel paths. Measuring and reporting cracking outside of the wheel path areas is not required for HPMS reporting. Report the lineal length of cracking for each wheel path.

     - Any and all severity levels (sealed and unsealed) should be reported.

- Multiply the lineal length of wheel path with cracking by 39 in. to get the total area of wheel path cracking.

- Cracking Percent is reported as the area of wheel path cracking divided by the total area of the section, multiplied by 100. To limit the potential variation in reported lane width, for the purposes of this cracking comparison the lane width is set at 12 ft. wide.

- Report Cracking Percent to the nearest one percent.

- For Jointed Concrete Pavements, Cracking Percent is the percentage of slabs within the section that exhibit transverse cracking. Partial slabs shall contribute to the section that contains the majority of the slab length.

  - Reported cracking for jointed concrete pavements excludes longitudinal cracks, corner breaks, D-cracking, and ASR cracking that may occur on a slab.

  - Report the actual total number of slabs in the section.

  - Report the total number of slabs containing one or more transverse cracks extending for at least one-half the lane width. This is the number of cracked slabs.

  - The percentage of cracking reported is calculated as the number of cracked slabs, divided by the total number of slabs in the section, multiplied by 100.

  - Report Percent Cracking to the nearest one percent.

- For CRCP, the Cracking Percent is the percentage of the area of the section exhibiting longitudinal cracking, punchouts, and/or patching. Transverse cracking shall not be considered in the Cracking Percent for CRCP.

  - Cracking and distresses may occur anywhere on the pavement. Transverse cracks that are at or near right angles to the direction of travel in the lane should not be included in the calculation.

  - Distresses to be included are longitudinal cracking (any severity), punchouts, and patched areas.

  - For longitudinal cracking, the cracked area is determined as the length of the crack multiplied by 1 ft. width.

  - For punchouts, the area is determined by the two transverse cracks and the edge of the pavement or longitudinal joint.

  - Report the area of all patches.

  - Percentage of Cracking for CRCP pavements is determined as the area of pavement where longitudinal cracking, punchouts, and patches are detected divided by the total area of the section, multiplied by 100. To limit the potential variation in reported lane width, for the purposes of this cracking comparison the lane width is set at 12 ft. wide.

- Report Percent Cracking to the nearest one percent.

2. Reference Site Selection and Weighting

   a. Selection of reference site surface types and conditions

   To determine the statistical reliability of HPMS cracking distress data provided by a vendor, an agency must have some reference set of data with which to make a comparison. The use of reference sites has proven to be very useful in establishing a basis for comparison of cracking distress data. Reference site pavement sections should be selected to represent the types and severity levels of distress normally experienced by the specific agency for each pavement type to be evaluated. The reference sites should consist of asphalt concrete pavements, jointed concrete pavements, and continuously reinforced concrete pavements (if present in the agency). A minimum of three sections of each pavement type should be included, although as many as six sections will enhance the statistical reliability. Asphalt concrete and continuously reinforced concrete sections should be a minimum of 0.3 mile (1,584 ft.) in length. Jointed plain concrete pavement sections should be a minimum of 0.5 mile (2,640 ft.) or 100 slabs in length, whichever is greater. The sections should have cracking distresses representing a range of types and severities typical for the agency. Sections will be subdivided to create subsections for the verification process.

   Additional sections may be considered for different roughness values (as roughness may affect image quality), surface textures, tined or grooved conditions, and the presence of other distresses on the pavement section that may affect the interpretation of the images.

   b. Weighting Criteria

   The percentage distribution of the reference sections by pavement type and condition does not have to correspond to the distribution of types and conditions on the network as a whole. Since each section will be statistically evaluated independently, the relative importance of each section can be weighted for the final interpretation of the statistical results. Thus, an agency can have a reference site with a set of conditions of interest, but not as critical as other conditions, without it skewing the overall decision process. This is an advantage of the site-by-site statistical evaluation.

   Therefore, for HPMS Cracking Percent, the significance of the reference sections to the selection process should be weighted consistently with the approximate distribution of those pavement conditions on the agency's network, based upon prior year information. The number of reference sites, subsections, and weighting criteria should be provided to all potential vendors prior to any vendors conducting surveys.

   c. Optional Regional Reference Sites

   As an alternative to individual agency reference sites, regional reference sites could be established to be used by grouping several agencies together. This will support the use of consensus ratings, which can be used to apply the paired t-test for equivalence to establish the cracking identification for reference sections. It is proposed that the ratings be carried out on a minimum number of three sites for each representative pavement surface type making sure that these sections include representative cracking distresses such as low, medium, and high levels of crack severity.

3. Ground Reference Rating
   It is recommended that the reference rating be a highly accurate consensus rating completed by experienced distress raters under traffic control and controlled conditions. The rater or raters should be uniformly trained in the specific distress interpretation before performing reference section ratings.

   Should an agency choose to conduct replicate ratings (either by using multiple independent raters or by having a single rater rate the section multiple times) for determination of the ground reference value, then equivalence testing of each independent rating should be assessed, following the same procedure that will be used to compare vendor ratings to the ground reference. Either a single rating or a consensus rating should be used as the ground reference. Replicates should not be averaged in determining the final reference rating, but rather only one of the "equivalent" replicate ratings should be used, or a consensus on the ground reference be determined. A vendor will only be supplying one rating per section and this provides for a valid one-to-one determination of equivalence between the ratings.

   If used, the consensus rating team should be composed of the team of persons who will be involved in the distress rating process. This should include any in-house agency reviewers, and any independent QA reviewers who will be responsible for reviewing data. In the situation where an agency is interested in qualifying a vendor system for equipment acquisition, it may be appropriate to assemble a consensus rating team including representatives from the prospective vendor(s).

   It is important that the data be collected and summarized on the specific subsections needed for the statistical analysis. Minimum subsection lengths of 0.03 miles are recommended for asphalt concrete and continuously reinforced pavements. For jointed concrete pavements, the recommended subsection length is 0.05 miles or 10 slabs, whichever is greater. Typically, 10 subsections will be necessary for each section. Shorter subsection lengths increase variability and, therefore, a tradeoff exists between subsection length, the number of subsections (N), and total length that must be surveyed by the agency. A more robust discussion on section and subsection lengths is provided in the Statistical Evaluation Discussion contained in Appendix A.

   All cracking equal to or greater than 1mm in width should be reported. The sections and subsections must be clearly marked with Start, End, and Intermediate Points. Mark or use a template to identify wheel paths, following the HPMS directives regarding location. The width of each wheel path should be 39 inches wide. In the case of jointed concrete pavements, the joints and/or slabs should be marked. Examples are provided in Figures 13, 14, and 15.

Figure 13. Example of wheel path marking prior to crack survey.

Figure 14. Example of a template used to define the wheel path areas surveyed.

Figure 15. Example marking of joints (red #2) and intermediate points (40 ft. station mark) on jointed concrete pavement.

4. Conduct Vendor Data Collection
   As a part of the prequalification process, the vendor is required to collect and report HPMS cracking data within a finite period of time (typically 2 to 10 days), to include the detailed measures as well as the HPMS Cracking Percentage on all designated reference sections. Data shall be reported at the specified subsection lengths.

   a. For asphalt concrete pavements, the vendor shall collect and report the lineal length of all cracking in each wheel path. They shall also report the Cracking Percent as previously defined.

   b. For jointed concrete pavements, the vendor shall collect and report the number of slabs in the section and the number of slabs with transverse cracks that extend at least ½ the width of the slab. They shall also report the Cracking Percent as previously defined.

   c. For continuously reinforced concrete pavements, the vendor shall collect and report the area of punchout and patches and the length of longitudinal cracking for all severity levels. They shall also report the Cracking Percent as previously defined.

5. Statistical Equivalence Evaluation
   Results from vendor data collection should be reviewed to assure that all the elements necessary to provide useful distress data are included. Following receipt of the vendor's data, the data comparison process should be conducted to evaluate the acceptability of the distress results produced. Detailed guidance on how to complete the paired t-test for equivalence is provided in Appendix A. An example is provided in Appendix B.

As explained in Appendix A, the agency will need to balance the acceptable power of the test and corresponding alpha value based upon the number of subsections and the standard deviation of the paired differences. Based on the current level of image technology, it is suggested that the possibility of greater discrepancy between reference and vendor ratings be deemed acceptable for the identification of low severity cracking. In the paired t-test for equivalence, this would mean the agency accepts a higher alpha value. With these known values, the p-value is then calculated and compared to the alpha value to determine equivalence.

Each reference site is evaluated on an independent basis. The pass/fail criteria are then applied to the section weighting factor and an overall vendor score is achieved.

6. Interpreting the Results
   Results from the statistical evaluation of the reference site data should be summarized into a report form indicating the acceptability of the information collected and any specific problems should be identified.

   In cases where the acceptability of images or data is in question, a detailed review of the information should be included. It may be possible to determine from this in-depth assessment that the results could be improved after making adjustments to the data interpretation process and re-evaluating the results. Time should be built into the review process to accomplish the adjustments and re-evaluation of the data. The size of the control dataset will affect the amount of time potentially needed for this process.

   An applied example of the reference site comparison evaluation process and interpretation is provided in Appendix B.

Alternative Cracking Distress Data Verification Options

In lieu of the above methodology of using reference sites and consensus ratings as a pre-selection verification method of the vendor's capability of determining cracking distress, two additional alternatives are provided. One is to allow a vendor to provide verification tests from another agency and the second is to require the selected vendor to pass a set of Certification Sites which the agency will establish post selection.

1. Proof of Verification
   An agency may elect to allow a vendor to provide results of verification tests from another agency as a means of proving both their equipment capabilities and their ability to produce acceptable cracking distresses from the images. The vendor should be asked to submit documentation, similar to the process described above, from a reputable agency that is in a similar climatic region and has similar pavement types.

2. Post Award Reference Sites
   In an effort to reduce the pre-selection time and efforts required of the vendor, an agency may desire to select a vendor based upon that vendor's proposed equipment and experience. In this case, all vendor equipment considered should meet or exceed the criteria required in the "Equipment Verification Requirements" section above. The selected vendor would then be required to pass a number of agency established reference sites prior to beginning production level data collection.

The process would be similar to the process outlined above in ***Cracking Distress Data Verification Steps***; however, in this case the risk would need to be balanced between the agency and vendor. The statistical parameters would have to be carefully determined in advance of the vendor's ratings. To accomplish the calibration, multiple attempts may be required by the vendor.

Framework of the Standard for Agency Modification

Probably the most critical question which must be answered in executing a pavement management system is how to make a determination (assessment/evaluation) of the accuracy of distress determination, whether data is collected manually in the field (boots on the ground) or using an automated system. Accurate identification of distresses is vitally important to achieving useful pavement management recommendations and budget estimates.

A number of obstacles have been identified which can affect the outcome of distress identification. These include factors which physically impact the apparent distress observed in a pavement. The most prominent of these is pavement temperature, which can change the size and affect the visibility of cracks on the pavement surface, and affect slab curling which is sometimes measured as joint faulting. This is the underlying factor behind the following recommendations:

- ***Make reference and image comparisons under similar conditions to prevent the observable distress from actually being different between the two.***

- ***Conduct annual distress surveys during the same seasonal conditions from year to year to avoid introducing climatic induced differences into the observable pavement condition.***

Moisture conditions can also result in changes in observed distress levels, particularly in the visibility of fine cracks. For jointed concrete pavements, wet conditions could increase active pumping which is indicative of poor slab support, as well as slab warping which could be manifested in ride quality.

The distress identification process can potentially introduce a significant amount of variation into the distress interpretation process. There are several elements which will all help reduce the variability in the distress identification process. These include the following:

- ***Clear, concise distress definitions*** – An important element of consistent distress interpretation is clear, concise distress definitions. In certain cases, discrepancies in distress interpretation can be eliminated by improving distress definitions. A prime example of this was the identification of well-defined wheel paths.

- ***Consistency between raters*** – For field collected manual distress rating (boots on the ground) and the manually assisted rating processes, it is vital that all raters are consistent in the identification of distresses. If rater A is reporting cracking differently from rater B, the desired consistency of the data cannot be achieved. A strong rater training program is best for addressing this issue.

- ***Consistency of rating over time*** – It is equally important for achieving consistent, reliable PMS data that individual raters do not change how they interpret cracking over time, as most state surveys require several months to accomplish. This issue can be addressed by checking individual raters to assure they have not changed how they interpret cracking over time by requiring them to rate five segments which they previously rated and compare the differences two times per month.

- ***Automated distress interpretation*** – It is anticipated that automated distress identification systems, when coupled with a robust method of verification of results, can minimize the rater related discrepancies.

Distress Data Verification Steps

1. Identification of Distress Types to be Identified
   To support improved data quality for state PMS applications, additional recommendations are provided. These recommendations are consistent with the original scope of the project. Each agency must identify the range of distress types and severity levels needed for inclusion in their reference sites.

   a. Review the PMS process to identify critical distress types and levels of major impact. Rather than attempting to confirm that the vendor can collect and report all distress types and severities from an agencies distress protocol, it is appropriate to focus on those distresses or indices which are key to decision making. An examination of the decision trees used by a state will typically reveal there are a specific number of distresses used to determine a PMS recommendation for treatment. Further, within this list of critical treatments, it is common to find a system of weights between distresses, which determines the most important distresses in terms of triggering a treatment recommendation. For example, probably the most common distress which triggers treatment in asphalt pavements is fatigue cracking. In most situations, the presence of fatigue cracking will trigger a rehabilitation consisting of repair and overlay. In jointed concrete pavement, distresses such as corner breaks or severe spalling will drive different decisions. While for continuously reinforced concrete pavements a critical distress is generally punchouts.

      If distress types are being collected which do not trigger an identified need, it is likely that the distress is not a significant factor in managing the pavement system.

      In the early days of pavement management, it was believed to be necessary to identify all the distress types which might be present in a pavement. However, experience has demonstrated that not all individual distress types will control treatment recommendations. Therefore, this step of evaluating the impact of distress type on treatment recommendations is a very effective step toward simplifying the requirements for a distress data collection process.

   b. Develop acceptable ranges of ratings (either individual distresses or indices) which comply with the recommended outcomes in the pavement asset management process.

   c. Determine the need for ranges of:

      - Surface type (hot mix asphalt, seal coats, chip seal, open-graded, jointed concrete pavement [JCP], CRCP, etc.)

      - Texture (rough, smooth, transverse tined, longitudinally tined, etc.)

      - Condition (a range of Good, Fair, and Poor based upon PMS condition rating)

A historical review of the PMS will allow the agency to determine the most significant pavement types, surface types, and distress types present on their system. It is not practical that all pavement surfaces and all distress combinations be included. The purpose of this step is to identify the most prominent surface types, surface textures and pavement conditions, to ensure these are included in the reference sites. At least one reference site should be selected to represent a pavement type, surface type, and condition when 15% or more are included in the system.

Determine the allowable acceptance limits, or statistical quality levels based on the paired t-test for equivalence. This represents a method for verifying that rating results (or distress interpretation results) are consistent enough to use as the basis for treatment and budget recommendations. Since it is impractical to eliminate all the potential variability in the distress interpretation process, it is important to determine how good is good enough, or how accurate is sufficient to provide consistent PMS recommendations. To answer this, consider the question, "At what point does incorrectly identifying a distress affect the subsequent PMS recommendation?"

The significance of distress types should be weighted consistently with the importance of the specific distress in the agency decision making process. Those distress types which drive the majority of treatment decisions should be weighted more heavily than those with less impact on overall PMS treatment recommendations. For example, for most agencies, fatigue cracking of asphalt pavements triggers the more extensive treatments, e.g., overlay or reconstruction. Therefore, this distress would be weighted more heavily (perhaps two times) than some others.

Framing statistical verification, that data is acceptable is not an easy challenge. Considerable effort has been put into this by many with limited success. Frequently used statistics, such as coefficient of variation and standard deviation, have not produced good results. In the case of validation of pavement distress the determination of reference distress values is a difficult challenge, for all the reasons previously discussed. Given two sets of data, how can a determination be made that one set provides the "correct" answer, or is more reliable than the other? For the problem of qualifying a vendor rating system, or distress rating results, the acceptability of data can be effectively evaluated using the paired t-test for equivalence process described in Appendix A.

An agency may find other statistics useful, but they should be careful that the application is sufficiently statistically valid to be relied upon for the acceptance of pavement distress data.

2. Reference Site Selection and Weighting

   a. Selection of reference site surface types and conditions

      Reference site pavement sections should be selected to represent the types and severity levels of distress normally experienced by the specific agency for each pavement type to be evaluated. The reference sites should consist of asphalt concrete pavements, jointed concrete pavements, and continuously reinforced concrete pavements (if present in the agency). A minimum of three sections of each pavement type should be included, although as many as six sections will enhance the statistical reliability. Asphalt concrete and continuously reinforced concrete sections should be a minimum of 0.3 mile (1,584 ft.) in length. Jointed plain concrete pavement sections should be a minimum of 0.5 mile (2,640 ft.) or 100 slabs in length, whichever is greater. The sections should have cracking distresses representing a range of types and severities typical for the agency. Sections will be subdivided to create subsections for the verification process.

Additional sections may be considered for different roughness values (as roughness may affect image quality), surface textures, and the presence of other distresses on the pavement section that may affect the interpretation of the images.

b. Weighting Criteria

The percentage distribution of the reference sections by pavement type and condition does not have to correspond to the distribution of types and conditions on the network as a whole. Since each section will be statistically evaluated independently, the relative importance of each section can be weighted for the final interpretation of the statistical results. Thus, an agency can have a reference site with a set of conditions of interest, but not as critical as other conditions, without it skewing the overall decision process. This is an advantage of the site-by-site statistical evaluation.

A field review of these initially selected reference sites is recommended to quantify the type, severity, and extent of distress represented by each reference section. Once this has been completed, the agency should provide the best distress interpretation possible for each reference site. This may take the form of rating by a single expert rater, completing replicate ratings, or generating a consensus rating developed by multiple raters.

3. Ground Reference Rating
It is recommended that the reference rating be a highly accurate consensus rating completed by experienced distress raters under traffic control and controlled conditions. The rater or raters should be uniformly trained in the specific distress interpretation before performing reference section ratings.

If replicate ratings are conducted for determination of the ground reference then equivalence testing of ratings should be assessed, following the same procedure that will be used to compare vendor ratings to the ground reference. Either a single rating or a consensus rating should be used as the ground reference. Replicates should not be averaged in determining the final reference rating, but rather only one of the "equivalent" replicate ratings should be used, or a consensus on the ground reference be determined. A vendor will only be supplying one rating per section and this provides for a valid one-to-one determination of equivalence between the ratings.

The consensus rating team should be composed of the team of persons who will be involved in the distress rating process. This should include any in-house agency reviewers, and any independent QA reviewers who will be responsible for reviewing data. In the situation where an agency is interested in qualifying a vendor system for equipment acquisition, it may be appropriate to assemble a consensus rating team including representatives from the prospective vendor(s).

It is important that the data be collected and summarized on the specific subsections needed for the statistical analysis. Minimum subsection lengths of 0.03 miles are recommended for asphalt concrete and continuously reinforced pavements. For jointed concrete pavements the recommended subsection length is 0.05 miles or 10 slabs, whichever is greater. Typically, 10 subsections will be necessary for each section. Shorter subsection lengths increase variability and, therefore, a tradeoff exists between subsection length, the number of subsections (N), and total length that must be surveyed by the agency. A more robust discussion on section and subsection lengths is provided in the Statistical Evaluation Discussion contained in Appendix A.

a. Agency-determined reference limits
   The statistical comparison of the rating results from the reference sections provides the basis for determining data acceptability. This comparison can be approached differently depending upon how the distress data is used in the agency PMS. When a distress index is used to develop treatment recommendations, comparing distress rating results on the basis of the index is appropriate. If individual distresses are used in decision trees to arrive at treatment recommendations, the paired t-test for equivalence can be applied to the difference between the reference ratings and the vendor ratings. The objective is always to control the variability of the data at an acceptable level so the agency can have confidence in using it to make PMS decisions. This approach recognizes there is not a "perfect" solution to the distress identification challenge. In some cases, a combination of a distress index and key distresses may best be used.

b. Distress Data Interpretation
   Once an agency has developed its reference rating data, these can be used to prequalify vendor collected distress data from a set of reference sections provided for that purpose, or to evaluate the acceptability of a specific data collection system operated by agency personnel. From the example for HPMS reporting, if it is determined that the difference between subsection values should be assessed for eight subsections with an alpha of 0.1 and a power of 0.8, the vendor, or system, should be able to consistently demonstrate the ability to produce results within this tolerance. Similar criteria should be determined for use with the agency PMS process. During actual data production, any bias identified can be addressed by discussions with the vendor. Experience has shown that they can make adjustments to the process which is useful in removing bias, or misidentification issues. However, at this time, these adjustments are often specific to the requirements of the agency.

4. Conduct Vendor Data Collection
   As a part of the prequalification process, the vendor shall collect and report pavement condition data according to the agency's pavement distress identification guide and/or data dictionary. This report should be submitted within a finite period of time (typically 2 to 10 days). It is important that the vendor be provided ample opportunity to discuss the distress definitions and reporting criteria before the vendor data is reported. Historically, a difference between the vendors interpretation of the distress definitions and data dictionary as compared to the agencies intention have been identified to be the cause of differences in reported cracking distresses, which had nothing to do with the equipment or the identification of cracking, but rather on the difference of interpretation for reporting.

   Agencies contemplating fully automated image interpretation at this time should verify that the data being produced fulfills the user's needs. In addition to the quality of images and accuracy of data interpretation, the distress types used by the agency will also play a role in how effective automated interpretation is. Certainly, a lessor number of required distress types helps simplify the distress interpretation process. Similarly, limiting the number of severity levels would also simplify the interpretation of cracking data. However, in many cases these changes would not adequately support existing state pavement management data needs.

5. Statistical Equivalence Evaluation
   Results from vendor data collection should be reviewed to assure that all the elements necessary to provide useful distress data are included. Following receipt of the vendor's data, the data comparison process should be conducted to evaluate the acceptability of the distress results

produced. Detailed guidance on how to complete the paired t-test for equivalence is provided in Appendix A and an example using HPMS Cracking Percent is provided in Appendix B.

As explained in Appendix A, the agency will need to balance the acceptable power of the test and corresponding alpha value based upon the number of subsections and the standard deviation of the paired differences. Based on the current level of image technology, it is suggested that the possibility of greater discrepancy between reference and vendor ratings be deemed acceptable for the identification of low severity cracking. In the paired t-test for equivalence, this would mean the agency accepts a higher alpha value. With these known values, the p-value is then calculated and compared to the alpha value to determine equivalence.

Each reference site is evaluated on an independent basis. The pass/fail criteria are then applied to the section weighting factor and an overall vendor score is achieved.

6. Interpreting the Results
   Results from the paired t-test for equivalence evaluation of the reference sites data should be carefully reviewed. Reference sites where the paired t-test is determined to not be equivalent will indicate where major differences between the agency distress ratings and the vendor ratings occur. Specific problems should be identified, and the vendor given the opportunity to improve upon their results.

   In cases where the acceptability of images or data is in question, a detailed review of the information should be included. It may be possible to determine from this in-depth assessment that the results could be improved after making adjustments to the data interpretation process and re-evaluating the results. Time should be built into the review process to accomplish the adjustments and re-evaluation of the data. The size of the reference site dataset will affect the amount of time potentially needed for this process.

   A final determination of the quality level of the vendor collected data, or vendor equipment collected data for agency use should result from the evaluation performed. If the data results are found to be within the agency determined acceptance limits, the vendor process or their equipment can be accepted for use by the agency.

7. Develop Guidelines That Can Be Used For Vendor Procurement
   The proposed guideline, included as Appendix C, is intended to aid agencies in the development of the procurement documents needed when hiring a vendor for the collection and reporting of high speed cracking distresses. This guideline provides the appropriate format for collecting and preparing data reports under the current HPMS Field Manual reporting requirements.

OUTLINE FOR GUIDELINES FOR EQUIPMENT AND SOFTWARE PURCHASING

The guidelines developed for Equipment and Software Purchasing will be based on the following key items:

- Illustrate that the system can produce high quality digital distress images that can be used to identify and classify cracking
- Illustrate that the system can produce cracking distress summaries meeting a standard (agency, national, or other)

- Robustness of equipment and software
- Flexibility/Adaptability of the software
- Degree of automation meeting the agency's needs
- Develop a standard that can be used for agency procurement

Equipment Verification Requirements

1. Imaging System Clarity
   It is anticipated all collection will soon be based on 3D imaging systems. Therefore, the recommendations provided will primarily focus on this type of system. However, the intent is not to limit technology, so any technology that can provide an image with enough clarity to identify the cracking distresses would be acceptable.

   a. Image Size – Pixels by Pixels
      For a 3D image 4,096 pixel transverse resolution will theoretically support the identification of a 1 mm wide crack. Similarly 1,300 pixel transverse resolution theoretically supports the identification of a 3 mm wide crack, and 2,048 pixels a 2 mm crack width. These theoretical resolution levels are best achieved when the camera is still or moving slowly. Finer cracks can be identified by higher resolution cameras. For 3D pavement imaging, 16-bit images are recommended. Crack identification is also affected by the 256 shade levels in an 8-bit image, making crack width identification more complex. [20,32] The vendor should be requested to provide their proposed image size.

   b. Image Dynamic Range Check
      Dynamic range is the ratio between the brightest to darkest signal levels. It determines how many levels of difference in digital values exist in a given image. For binary (black and white) images, 8-bits (256 levels) is usually sufficient to represent visual information. Images of 16 bits are the minimum required to handle 3D images, with some systems using as much as 32 bits. Color images with dynamic range of 24-bit are available, and may be needed to capture detailed features such as surface texture. The vendor will be requested to provide their image dynamic range factor.

   c. Percent Fill Factor
      Pixel fill factor indicates the light gathering area of the photo detector being used. The proportion of the pixel area insensitive to light is indirectly indicated. The minimum recommended fill factor is 90%. This indicates that 10% of the pixel is insensitive to light. The photo detectors commonly used are most often a silicon chip (CCD type) or a metal oxide semiconductor (CMOS type). A lower fill factor may be acceptable if the resulting image quality is sufficient to identify the desired distresses. The vendor should be requested to provide their proposed minimum fill factor.

2. Image Resolution
   3D line scan data collection produces higher resolution, dynamic range, and higher fill factor than earlier technologies. It also reduces blurred images of fast moving objects. The application of the three-dimensional range data with laser image lighting has resulted in major reduction in image clarity issues from variable lighting conditions and improved interpretation of poor intensity contrasts such as oil stains on the pavement surface.

According to the referenced report by Dr. Tsai, the Georgia Tech 3D system is able to identify a 1 mm crack, and a 2 mm crack when images are collected above 60 mph. [22] Current RFP requirements range from 3mm to 1mm systems, with no clarifications on how that is measured or evaluated. The guidelines will recommend the vendor provide the speed at which a 1mm crack is detected, the speed at which a 2 mm crack is detected and the speed at which a 3mm crack is detected.

3. Compression Type And Storage Size Requirements Vendor Uses To Meet Data Delivery
   Currently, the size of the original data collection files and compression method used varies from one software developer to another. However, FHWA is in the final publication phase of a report titled "Evaluation of Proposed Standard Data Format and Compression Algorithms for 2D/3D Pavement Surface Image" [67] which proposes a standard data format for the 2D/3D image systems. An agency may consider requiring the 2D and 3D images provided by the vendor comply with the proposed standard data format. An example of the current minimal rules and validation procedures were previously listed in Table 1.

4. Image Capture Width
   The width of image should cover the entire driven lane in the travel direction, accounting for vehicle wander, which is typically 13.5 to 14 ft. wide.

5. Does the equipment hardware and software have demonstrated durability and utility to provide the agency needs for distress data collection?
   If the equipment has been used to collect data for systems of a similar size or larger, maintaining a reasonable delivery schedule, this should demonstrate the durability of the equipment. If new developments introduce components which have not been so demonstrated, is there reason to be concerned about the durability under extended field use? If so, identify concerns and ask the vendor to address these concerns regarding the use of the equipment. This could take the form of a warranty to perform, or support if equipment or software problems do arise.

6. Can the equipment be configured to provide the data required by the agency?
   If the vendor equipment and software have met the previous requirements, it will be necessary to demonstrate that the system can provide the distress data in the format required by the agency. This could include the thoroughness and adaptability of software delivery to provide the agency distress types and severity levels. It should also include the ability to adjust any software tools used to identify cracking and/or interpret cracking patterns. This item may take the form of vendor support, at least for an initial period while agency personnel learn to perform this task alone.

7. Does the equipment provide the degree of automation which meets the agency's needs?
   The agency should specify their expectations for automated distress interpretation. This could take several forms such as automated interpretation of wheel path fatigue and transverse cracking. Or it could request additional distress identification capability. At the present time, the demonstrated ability of software to reliably provide accurate distress data is limited. It is expected that this capability will improve over time, and adjustments could be made in what is expected from a system as the technical capacity of systems change. The agency should have a backup plan for obtaining distress interpretation of its collected data in the event it is found that partial or full manual interpretation assistance is needed.

8. Develop Guidelines That Can Be Used For Agency Equipment Procurement
   The proposed guideline, included as Appendix D, is intended to aid agencies in the development of the procurement documents needed when purchasing high speed cracking distress data collection and processing equipment. This guideline provides the appropriate format for collecting and preparing data reports under the current HPMS Field Manual reporting requirements.

Cracking Distress Data Verification Steps

1. Identification of Cracking Distress Types to be Verified
   The HPMS Field Manual defines cracking as “a fissure or discontinuity of the pavement surface not necessarily extending through the entire thickness of the pavement.” The AASHTO R86 “Standard Practice for Collecting Images of Pavement Surfaces for Distress Detection” defines a crack as “a fissure of the pavement material at the surface with minimum dimensions of 1mm (0.04 in.) width and 25 mm (1 in.) length. The HPMS reporting criteria that require the reporting of the Cracking Percentage for Asphalt Concrete, Jointed Concrete, and Continuously Reinforced Concrete are as follows:

   - For Asphalt pavements, Cracking Percent is the percentage of the total area exhibiting visible fatigue type cracking (any cracking in the wheel path) for all severity levels in the wheel path in each section.
     - Cracking will be measured and reported for both wheel paths. Measuring and reporting cracking outside of the wheel path areas is not required for HPMS reporting. Report the lineal length of cracking for each wheel path.
     - Any and all severity levels (sealed and unsealed) will be reported.
     - Multiply the lineal length of wheel path with cracking by 39 in. to get the total area of wheel path cracking.
     - Cracking Percent is reported as the area of wheel path cracking divided by the total area of the section, multiplied by 100. To limit the potential variation in reported lane width, for the purposes of this cracking comparison the lane width is set at 12 ft. wide.
     - Report Cracking Percent to the nearest one percent.

   - For Jointed Concrete Pavements, Cracking Percent is the percentage of slabs within the section that exhibit transverse cracking. Partial slabs shall contribute to the section that contains the majority of the slab length.
     - Reported cracking for jointed concrete pavements excludes longitudinal cracks, corner breaks, D-cracking, and ASR cracking that may occur on a slab.
     - Report the actual total number of slabs in the section.
     - Report the total number of slabs containing one or more transverse cracks extending for at least one-half the lane width. This is the number of cracked slabs.
     - The percentage of cracking reported is calculated as the number of cracked slabs, divided by the total number of slabs in the section, multiplied by 100.

- Report Percent Cracking to the nearest one percent.

- For CRCP, the Cracking Percent is the percentage of the area of the section exhibiting longitudinal cracking, punchouts, and/or patching. Transverse cracking shall not be considered in the Cracking Percent for CRCP.

    - Cracking and distresses may occur anywhere on the pavement. Transverse cracks that are at or near right angles to the direction of travel in the lane should not be included in the calculation.

    - Distresses to be included are longitudinal cracking (any severity), punchouts, and patched areas.

    - For longitudinal cracking, the cracked area is determined as the length of the crack multiplied by 1 ft. width.

    - For punchouts, the area is determined by the two transverse cracks and the edge of the pavement or longitudinal joint.

    - Report the area of all patches.

    - Percentage of Cracking for CRCP pavements is determined as the area of pavement where longitudinal cracking, punchouts, and patches are detected divided by the total area of the section, multiplied by 100. To limit the potential variation in reported lane width, for the purposes of this cracking comparison the lane width is set at 12 ft. wide.

    - Report Percent Cracking to the nearest one percent.

2. Reference Site Selection and Weighting
    a. Selection of reference site surface types and conditions
    To determine the statistical reliability of HPMS cracking distress data provided by a vendor's equipment, an agency must have some reference set of data with which to make a comparison. Reference site pavement sections should be selected to represent the types and severity levels of distress normally experienced by the specific agency for each pavement type to be evaluated. The reference sites should consist of asphalt concrete pavements, jointed concrete pavements, and continuously reinforced concrete pavements (if present in the agency). A minimum of three sections of each pavement type should be included, although as many as six sections will enhance the statistical reliability. Asphalt concrete and continuously reinforced concrete sections should be a minimum of 0.3 mile (1,584 ft.) in length. Jointed plain concrete pavement sections should be a minimum of 0.5 mile (2,640 ft.) or 100 slabs in length, whichever is greater. The sections should have cracking distresses representing a range of types and severities typical for the agency. Sections will be subdivided to create subsections for the verification process.

    Additional sections may be considered for different roughness values (as roughness may affect image quality), surface textures, tined or grooved conditions, and the presence of other distresses on the pavement section that may affect the interpretation of the images.

b. Weighting Criteria
The percentage distribution of the reference sections by pavement type and condition does not have to correspond to the distribution of types and conditions on the network as a whole. Since each section will be statistically evaluated independently, the relative importance of each section can be weighted for the final interpretation of the statistical results. Thus, an agency can have a reference site with a set of conditions of interest, but not as critical as other conditions, without it skewing the overall decision process. This is an advantage of the site-by-site statistical evaluation.

Therefore, for HPMS Cracking Percent, the significance of the reference sections to the selection process should be weighted consistently with the approximate distribution of those pavement conditions on the agency's network, based upon prior year information. The number of reference sites, subsections, and weighting criteria should be provided to all potential vendors prior to any vendors conducting surveys.

c. Optional Regional Reference Sites
As an alternative to individual agency reference sites, regional reference sites could be established to be used by grouping several agencies together. This will support the use of consensus ratings, which can be used to apply the paired t-test for equivalence to establish the cracking identification for reference sections. It is proposed that the ratings be carried out on a minimum number of three sites for each representative pavement surface type making sure that these sections include representative cracking distresses such as low, medium, and high levels of crack severity.

3. Ground Reference Rating
It is recommended that the reference rating be a highly accurate consensus rating completed by experienced distress raters under traffic control and controlled conditions. The rater or raters should be uniformly trained in the specific distress interpretation before performing reference section ratings.

Should an agency choose to conduct replicate ratings, (either by using multiple independent raters or by having a single rater rate the section multiple times), for determination of the ground reference value, then equivalence testing of each independent rating should be assessed, following the same procedure that will be used to compare vendor ratings to the ground reference. Either a single rating or a consensus rating should be used as the ground reference. Replicates should not be averaged in determining the final reference rating, but rather only one of the "equivalent" replicate ratings should be used, or a consensus on the ground reference be determined. A vendor will only be supplying one rating per section and this provides for a valid one-to-one determination of equivalence between the ratings.

If used, the consensus rating team should be composed of the team of persons who will be involved in the distress rating process. This should include any in-house agency reviewers, and any independent QA reviewers who will be responsible for reviewing data. In the situation where an agency is interested in qualifying a vendor system for equipment acquisition, it may be appropriate to assemble a consensus rating team including representatives from the prospective vendor(s).

It is important that the data be collected and summarized on the specific subsections needed for the statistical analysis. Minimum subsection lengths of 0.03 miles are recommended for asphalt concrete and continuously reinforced pavements. For jointed concrete pavements, the recommended subsection length is 0.05 miles or 10 slabs, whichever is greater. Typically, 10 subsections will be necessary for each section. Shorter subsection lengths increase variability and, therefore, a tradeoff exists between subsection length, the number of subsections (N), and total length that must be surveyed by the agency. A more robust discussion on section and subsection lengths is provided in the Statistical Evaluation Discussion contained in Appendix A.

All cracking equal to or greater than 1mm in width should be reported. The sections and subsections must be clearly marked with Start, End, and Intermediate Points. Mark or use a template to identify wheel paths, following the HPMS directives regarding location. The width of each wheel path should be 39 inches wide. In the case of jointed concrete pavements, the joints and/or slabs should be marked. Examples of such markings were provided in Figures 13, 14, and 15.

4. Conduct Vendor Data Collection
   As a part of the selection process, the vendor is requested to collect and report HPMS cracking data, to include the detailed measures as well as the HPMS Cracking Percentage on all designated reference sections. Data shall be reported at the specified subsection lengths.

   a. For asphalt concrete pavements, the vendor shall collect and report the lineal length of all cracking in each wheel path. They shall also report the Cracking Percent as previously defined.

   b. For jointed concrete pavements, the vendor shall collect and report the number of slabs in the section and the number of slabs with transverse cracks that extend at least ½ the width of the slab. They shall also report the Cracking Percent as previously defined.

   c. For continuously reinforced concrete pavements, the vendor shall collect and report the area of punchout and patches and the length of longitudinal cracking for all severity levels. They shall also report the Cracking Percent as previously defined.

5. Statistical Equivalence Evaluation
   Results from vendor data collection should be reviewed to assure that all the elements necessary to provide useful distress data are included. Following receipt of the vendor's data, the data comparison process should be conducted to evaluate the acceptability of the distress results produced. Detailed guidance on how to complete the paired t-test for equivalence is provided in Appendix A. An example is provided in Appendix B.

   As explained in Appendix A, the agency will need to balance the acceptable power of the test and corresponding alpha value based upon the number of subsections and the standard deviation of the paired differences. Based on the current level of image technology, it is suggested that the possibility of greater discrepancy between reference and vendor ratings be deemed acceptable for the identification of low severity cracking. In the paired t-test for equivalence, this would mean the agency accepts a higher alpha value. With these known values, the p-value is then calculated and compared to the alpha value to determine equivalence.

   Each reference site is evaluated on an independent basis. The pass/fail criteria are then applied to the section weighting factor and an overall vendor score is achieved.

6. Interpreting the Results
   Results from the statistical evaluation of the reference site data should be summarized into a report form indicating the acceptability of the information collected and any specific problems should be identified.

   In cases where the acceptability of images or data is in question, a detailed review of the information should be included. It may be possible to determine from this in-depth assessment that the results could be improved after making adjustments to the data interpretation process and re-evaluating the results. Time should be built into the review process to accomplish the adjustments and re-evaluation of the data. The size of the control dataset will affect the amount of time potentially needed for this process.

   An applied example of the reference site comparison evaluation process and interpretation is provided in Appendix B.

CONCLUSION

This report documents the efforts completed by the project team in determining current best practices in the collection and comparison of pavement cracking distresses. A detailed literature review of over 60 documents, including international publications, was completed. This was further supplemented with input from the 21 pooled fund agencies and discussions with most of the major vendors. Additionally, distress data from both vendors and agency reference ratings were provided for agencies, which were used as validation data for the recommended procedure.

A two-step process is recommended; step one addresses the equipment verification requirements while step two addresses the cracking distress data verification. Based upon input from the Pooled Fund Agencies, the HPMS Cracking Percentage was to be used as the reference value for comparison. In an effort to ensure that the vendors can collect and report accurate cracking data, best practices are to perform a consensus rating in the field, under closed traffic conditions resulting in high quality cracking data. After a thorough review of several different statistical comparison tests, including the two-sigma limit test, Pearson's coefficient, t-test, and paired t-test, the recommended method is to use a two one-sided paired t-test for equivalence.

Based on experience with all three methods of collecting cracking distress information (manual, 2D systems, and 3D systems), the manual survey method from boots on the ground was selected as the most appropriate method for ground reference rating. Based on experiences with these three methods, it is expected that the manual survey method will report fine cracking possibly not identified by the automated imaging systems. It is also expected that the 3D imaging systems will do better at capturing fine cracking than the 2D systems. Regardless of whether imaging system data is interpreted manually, or by a fully automated process, the limitation identified in this study is a function of the clarity of fine cracking in the image. As a result, the distress survey methods could result in different quantities of fine cracking. This difference will be evident when the paired t-test for equivalence is performed on sections with large amounts of fine cracking. Most likely, equivalency will only be found acceptable as the agency accepts a higher alpha value, which means a higher risk. Conversely, the weighting factor applied to sections with fine cracking may be adjusted depending upon an agency's belief in the criticality of identifying fine cracking.

APPENDIX A – STATISTICAL FIELD VERIFICATION BACKGROUND AND STATISTICAL EQUIVALENCE FORMULATION

Project Background

A statistical quality assurance process was developed in 2000 for the evaluation of pavement data interpreted by manual ratings from digital images. A validation of the distresses reported from the digital images was compared with rater pool assessment from selected reference sections exhibiting specific distress types.(68) This process was originally proposed for the calibration of data collection equipment. This included application of the two sigma limits, or a narrower confidence interval as a potential method of statistical analysis, in association with a rating distribution developed using a pool of raters. Based on feedback from the project panel, the availability of a qualified rater pool large enough to establish variability for confidence intervals was identified as a significant limitation for many agencies. A one-to-one comparison based on variabilities established on other pavement sections using historical data is not adequate for vendor or equipment selection.

Subsequently, a different approach was taken considering a limited number of possible state pavement raters to develop a reference rating value for pavement sections. Some states indicated that there is only a single rater available to evaluate reference sections, so this was considered in statistical approaches moving forward. During the review process, the suggestion was made that a larger statistical sample size (N) could be achieved by subdividing the reference sites rather than by increasing the number of raters.

The initial thought in discussions was to compare all reference sites in a single hypothesis test (t-test). However, that posed several limitations, including the need for the distribution of reference site conditions to directly mirror that of the agency's network and the need to use the same statistical assumptions (variability, limits) on all reference sites collectively. This led to several key decisions in the development of the proposed process:

- To analyze the reference sites individually so their importance can be weighted and appropriate statistical parameters can be set for each site.

- To utilize a paired t-test on each site, directly pairing the reference and vendor ratings on each subsection.

This site-by-site paired t-test approach was tested on data submitted by several of the pooled fund agencies. Using the paired t-test, and typical assumptions for statistical parameters, the trial results were successful on some data. But, as would be expected, the results also showed:

- Ratings are statistically different if variability of the differences is small but the value of the differences is smaller.

- Ratings are statistically the same if the variability of the differences is large.

Neither conclusion is desirable. The first would result in rejection of vendor ratings that are statistically significantly different than the reference ratings, but have no practical impact on the use of the data because the magnitude of the significant difference is small. To counter this effect, a two-tiered approach was developed in which the magnitudes of differences that make a practical impact on reported results was also considered--for example, the magnitudes of differences in HPMS Cracking Percent that would cause a pavement section to be rated as qualitatively Good versus Fair or as Fair versus Poor. For example, for asphalt concrete sections, the alternate approach would have been to compare vendor to reference ratings of HPMS Cracking Percent as follows:

- For reference ratings between 0-30%, acceptable difference between vendor and the reference rating is ±4%

- For reference ratings greater than 30%, acceptable difference between vendor and the reference rating is ±10%

Vendor ratings would have had to either pass the t-test as not having a statistically significant difference from the reference rating or alternatively to fall below the threshold for magnitude of difference that affects outcomes.

This approach was also tested on the data provided by the pooled fund agencies. However, there were still complications and the desired results were not consistently produced. A key reason was the second observation—that vendor ratings were found to be statistically the same as the reference rating if the variability of the differences in the paired data were large. That meant that a vendor whose results were less consistent could have an advantage. That is a logical result for typical applications of hypothesis testing, in which the purpose of the test is to determine if there is a *statistically significant difference* between the reference (or control) and experimental data.

In the current application, however, the need is to determine if the results are *statistically the same*. In hypothesis testing, such as the paired t-testing, the absence of not being statistically different is not at all the same as being statistically the same. Therefore, the project team returned to the fundamentals of statistical hypothesis testing and considered alternate hypothesis formulations.

Introduction to Statistical Equivalence Testing

One promising formulation of statistical tools for informing decisions as to whether results of two methodologies are statistically the same can be found in biostatistics, in the formulation of statistical tests developed for the acceptance of generic pharmaceuticals. (69,70,71) Unlike statistical tests formulated to seek significant differences between methods or for variables significantly affecting outcomes, these tests are formulated to assess equivalence or noninferiority between methods, thus also holding promise for assessing if new equipment or vendors are equivalent (or noninferior) to current accepted standards. Very recently, these methods have also been proposed for use in the social sciences and in quality and manufacturing engineering.(72,73)

While no prior applications to infrastructure asset management data have been located in the literature, the equivalence method is more straightforward in this application than the two-tiered approach previously considered. Subsequently, the research team determined that using a paired t-test for equivalence provided a better approach for the needed data analysis.

- Our goal is to identify vendors or equipment that give the **same** results as the ground reference or results close enough to **not** affect the outcomes.

- Equivalence tests are hypothesis tests formulated for when equivalence rather than significant difference is the goal.

- Using a **paired t-test** for the equivalence testing, with the ratings carefully paired on the same rating subsections, offsets the variability in the pavement along the length of the control sites.

Because it is testing the hypothesis of whether ratings are the same, rather than whether they are different, the equivalence formulation of hypothesis testing:

- Directly addresses the magnitude of differences that is practically different, without requiring a two-tiered approach or alternate criteria for acceptance.

- Is more likely to find that the ratings are statistically equivalent if the variability of the differences between the reference ratings and vendor ratings are small.

- Is less likely, rather than more likely to find that the ratings are not statistically equivalent if the variability of the differences between the reference ratings and vendor ratings are large.

Paired t-testing can still be utilized, retaining the advantages of the one-to-one pairing of subsections, while resolving the main difficulties with its application. However, a reformulation of the null hypothesis is required, as well as a reconsideration of the statistical parameters (alpha, N, limits, power) used for the test. Equivalence testing lends itself especially well to considering the differences needed for shifts in defined infrastructure asset condition states directly when setting the equivalence limits.

Statistical Parameters for Paired t-test for Equivalence

The paired t-test for equivalence is included in many statistical software packages. The overview of statistical parameters presented here assumes that the user has access to such a statistical software package for conducting the analysis. The paired t-test for equivalence is often paired with a confidence interval approach as a function of the regulatory requirements associated with its development. At this time, the confidence interval approach is not recommended within the constraints of the current problem.

The examples and values presented here are based on using HPMS Cracking Percent as the decision variable. If the vendor or equipment selection is based on agency indices or individual distresses, the equivalence limits will be different. The variability of the differences between the vendor ratings and reference ratings may also be different. Therefore, the recommended values for all statistical inputs may differ. At a minimum, power curves would need to be generated for the specific decision variables and associated equivalence limits.

Reference Site Subsection Length

The reference and vendor rating values are determined at the subsection level. If the subsections are too small, this will result in additional variability in the ratings being introduced, which could obscure the true differences between the reference and vendor ratings. This can mean that a much higher N (number of subsections) is required for a meaningful statistical test, and that a total greater reference site length would be needed.

Since HPMS requires 0.1-mile reporting lengths, the use of 0.1-mile reference sites was logical. But how finely can we subdivide the 0.1-mile reporting lengths? Alternatively, if 0.1-mile subsections were used (for N subsections), the overall section length required to be rated would become onerous, especially for the ground reference ratings.

Therefore, 0.01-mile subsections were considered. However, based on the literature, it was surmised that there will likely be more variability between the 0.01-mile HPMS cracking percentages than there would be at the 0.1-mile HPMS reporting length, due to the offsetting errors in the greater length. In the 0.01-mile length, a very small difference in actual length of measured cracking could produce a difference in the HPMS Cracking Percent. An example of this effect can also be seen at the 0.1-mile to 1-mile reporting lengths, as shown in Figure 16, although the effect is expected to be less than at shorter intervals.

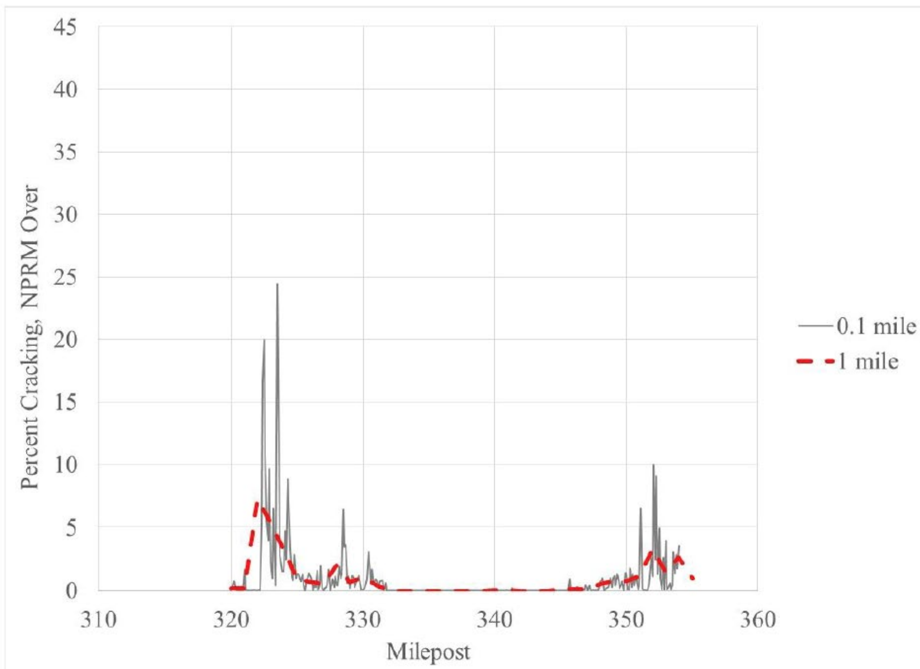


Figure 16. Impact of section length on percent cracking. [74]

The variability for paired t-testing is the variability of the differences between the reference and vendor data, not the variability in the cracking measurements between the subsections themselves. The issue of subsection length in this context was examined using data provided by the state agencies. Data collected by one of the agencies for 0.3-mile sections, divided into 0.01-mile subsections, was analyzed at various subsection lengths. Table 2 includes average values of Standard Deviation/Mean for that agency's cracking distress measurement, for values of N (number of subsections) and length of subsection. The 0.01-mile data was aggregated to produce the different subsection lengths. The agency had collected data from four vendors.

Table 2. Examination of the impact of subsection length and number of subsections on cracking data variability (standard deviation of the differences between vendor and reference ratings/mean rating) using 2015 reference site data from one state agency and four vendors.

SECTION	N=10; 0.01-mile subsections[1]	N=30; 0.01-mile subsections	N=15; 0.02-mile subsections	N=10; 0.03-mile subsections	N=6; 0.05-mile subsections
A	3.78	4.11	3.44	2.88	1.82
B	3.85	3.11	2.13	1.80	1.76
C	1.70	1.52	1.29	1.10	0.89
D	1.42	1.15	0.89	0.82	0.84
E	1.31	1.71	1.42	1.25	0.99
F	2.62	2.23	1.72	1.46	1.29
G	1.17	1.19	0.94	0.87	0.85
H	3.66	3.03	1.98	1.32	1.04
I	2.78	2.28	1.46	1.33	1.11
J	1.19	1.40	1.15	0.94	0.63
AVG	*2.35*	*2.17*	*1.64*	*1.38*	*1.12*

[1] Variability listed here is the average of the variability of the three sets of N=10 0.01mile subsections over the entire 0.3 mile section.

Due to the aggregating of the data, the effects of N and subsection length are confounded, except for the 0.01-mile subsections. Nonetheless, the clear effect of both parameters can be seen in the table. However, one would expect the standard deviation to decrease as N increases, and the opposite occurs here. That implies that actual decrease in standard deviation with increase in subsection length would be even greater for the same N.

Minimum subsection lengths are recommended based upon this and similar data, and upon limited additional studies of both real and synthesized data, balanced against the realistic consideration of tradeoffs between subsection length, number of subsections (N), and total length that must be surveyed manually to develop the ground reference. As a simple point of comparison, the recommended subsection lengths are similar to the sample unit areas long required for PAVER inspections. (75)

The recommended minimum subsection lengths for cracking verification are 0.03 mile for asphalt concrete and continuously reinforced concrete and 0.05 mile or 10 slabs for jointed plain concrete pavement.

Number of Subsections per Reference Site

Each reference site is evaluated for statistical equivalence independently using the paired t-test. This means that an adequate value for sample size, N, must be available within each reference site. The sample size is considered as the number of subsections. Higher values of N will result in greater power of the statistical equivalence test. This again, however, will result in a greater total length of ground reference surveys being required. A minimum N of 10 subsections is recommended.

Equivalence Limits

For two ratings to be equivalent, it does not mean that they are identical. It means that the ratings will produce satisfactorily the same outcome. Therefore, the equivalence limits for equivalence testing are not limits based principally or solely on the statistical variability of the data. Rather, the limits are the answer to the question: what difference between the two sets of results is unacceptable?

The selection of an acceptable difference is dependent principally upon the application. However, the statistical variability of the differences between the vendor and reference rating and the sample size may make some desirable equivalence limits difficult to achieve with current technology. These tradeoffs are discussed in following sections on agency risk (alpha) and power.

In a pavement asset management system, the equivalence limits should be based upon the impacts on the decision support system outputs. What difference in ratings would make a significant difference in condition state? What difference would be likely to affect the recommended treatment category? The process to determine the limits will depend substantially upon the specific pavement asset management system and can vary from simple to complex. The limits can vary with an individual distress or with a specific composite index. The equivalence limits can also vary with condition, if that matters in the final use of the data.

Both upper and lower equivalence limits are set; the range of values between the equivalence limits is also called the equivalence interval. Typically, the equivalence limits will be plus or minus the same value, or a range around the mean, although that is not a necessary constraint. To perform a paired t-test for equivalence, considering the upper and lower equivalence limits, two one-sided t-tests (sometimes noted as TOST) must be performed rather than a single two-sided paired t-test. Typically, software packages that include equivalence tests will embed this in the paired t-test for equivalence.

For this example using HPMS Cracking Percent criteria the equivalence limits would be based on the differences that would change the qualitative ratings. Thus, for this example, the selected equivalence limits will be the same as the acceptance limits previously discussed for the alternate approach in the two-tier method, although the limits are applied differently. For reference, the HPMS Cracking Percent criteria are provided in Table 3. [76]

Table 3. HPMS Cracking Percent Criteria for Qualitative Section Condition Rating

Pavement Type	Cracking Percent – Good	Cracking Percent – Fair	Cracking Percent – Poor
Asphalt Concrete Pavement	Less than 5%	Greater than or equal to 5% and less than or equal to 20%	Greater than 20%
Jointed Concrete Pavement	Less than 5%	Greater than or equal to 5% and less than or equal to 15%	Greater than 15%
Continually Reinforced Concrete Pavement	Less than 5%	Greater than or equal to 5% and less than or equal to 10%	Greater than 10%

For this example on the asphalt concrete sections, the selected equivalence limits for HPMS Cracking Percent are:

- For reference ratings between 0-30%, upper and lower equivalence limits between the vendor and reference ratings are ±4%

- For reference ratings greater than 30%, upper and lower equivalence limits between the vendor and reference ratings are ±10%

For this example on the jointed plain concrete pavements, the selected upper and lower equivalence limits for HPMS Cracking Percent are ± 7.5%. For continuously reinforced concrete pavements, the selected upper and lower equivalence limits are ± 10%.

Standard Deviation

The standard deviation to be used when performing the paired t-test for equivalence is the standard deviation of the differences between the ground reference and vendor rating values on the specific reference site. This will be determined from the data itself and will be calculated as part of the statistical cracking verification process.

However, assessing the acceptable balance between sample size (N), agency risk (alpha), equivalence limits, and power of the test, this standard deviation will need to be estimated initially. If historical data is available, it can be used to estimate the standard deviation. In the absence of any other information, the standard deviation can be estimated from the variability of the ground reference subsection ratings within a single section. The estimated standard deviations are only for initial planning purposes; the true standard deviations will be used in the statistical determinations of equivalence.

The reason it may be necessary to estimate the standard deviation is because an agency needs to set the sample size and alpha prior to data collection, such as during the procurement process. When an agency develops their request for proposal, it will be necessary to identify the statistical testing and limits expected of the vendors, ensuring that everyone is aware of the decision-making criteria. In cases where it is not necessary to make those decisions beforehand, this step of estimating the standard deviation would not be required.

Agency Risk of Accepting Non-equivalent Rating as Equivalent (Alpha)

When using hypothesis testing for discerning statistically significant differences, as is common for research studies and publications, it is commonly accepted (although often debated in recent controversies) to set the risk of determining a difference is significant when it is not at 5%. This is the same as setting the value of alpha to 0.05 ($\alpha = 0.05$).

In the paired t-test for equivalence, **alpha is the risk of accepting a rating as equivalent (within the equivalence limits) when it is not**. Considering the variability of current pavement ratings and the relatively small recommended sample size (N) for cracking verification, agencies may need to sometimes accept a greater risk than alpha = 0.05. This is particularly likely for fine cracking.

For a two one-sided paired t-test for equivalence, the hypotheses are formulated as follows:

- Null hypothesis (H_0): The difference between the means is outside your equivalence interval. The means are not equivalent.

- Alternative hypothesis (H_1): The difference between the means is inside your equivalence interval. The means are equivalent.

For the two one-sided paired t-test for equivalence, the higher of the two p-values is used. As for other t-tests, **if the p-value for the test is less than alpha**, the null hypothesis is rejected. However, **in the paired t-test for equivalence, if the p-value for the test is less than alpha, the null hypothesis is rejected, and the conclusion is that the means are equivalent**.

Power of the Equivalence Test

The power of the test is 1-beta (1-β), where **beta is the risk of determining the vendor rating is not equivalent to the ground reference rating when it is in fact equivalent**. In other words, power is the probability of correctly concluding that the vendor rating is within the equivalence limits. An equivalence test with low power is disadvantageous to the vendor.

For cracking verification, it is recommended that power be kept above 0.8 (or 80% chance of correctly determining equivalence). Higher values of power are preferable. Figure 17 shows a power curve as an example in the context of comparing HPMS Cracking Percent reference and vendor ratings. The figure is annotated to illustrate key inputs in the context of the specific example.

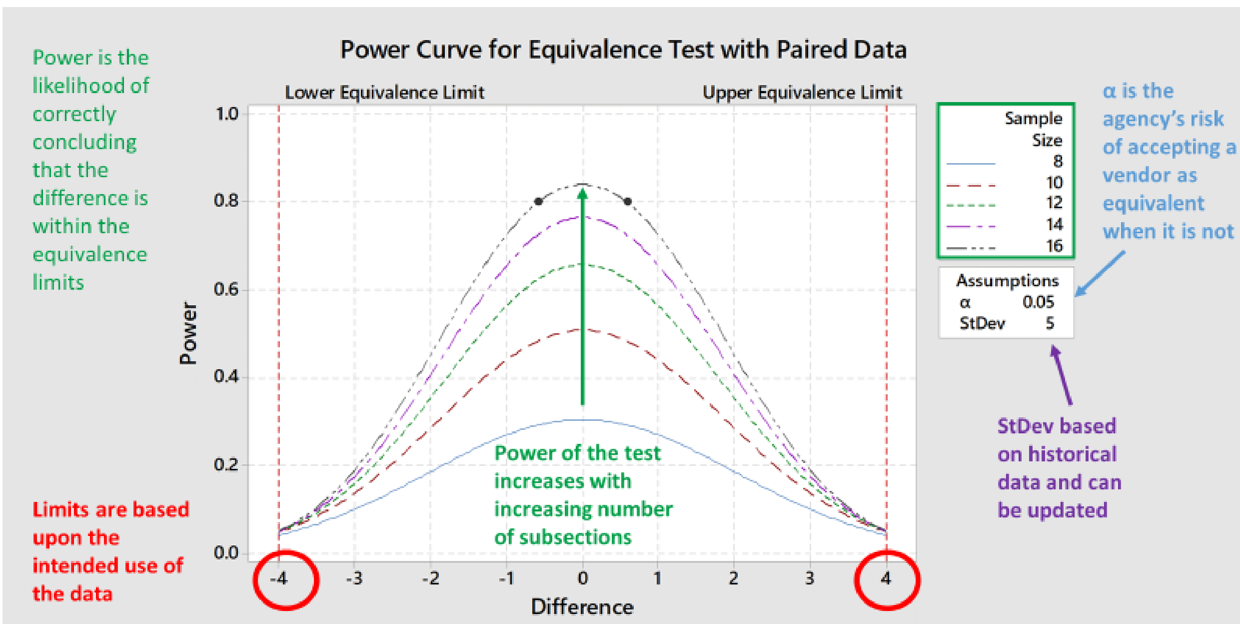


Figure 17. Annotated example power curve for two-sided paired t-test for equivalence.

There is a balance between N, alpha, standard deviation, the equivalence limits and power. Once the vendor ratings have been completed, N has already previously been determined and standard deviation of the differences between the reference and vendor ratings is a function of the rating data. The equivalence limits should have been determined from the use of the data. Therefore, the only parameters left to balance are alpha (agency risk) and power (essentially equivalent to 1- vendor risk). To avoid excessive risk to the vendors, an agency may have to increase their risk acceptance (alpha). Such adjustments can be done on an individual reference site basis, but must be applied equally to all vendors. For example, a higher value of alpha might be accepted on a reference site with predominantly very fine cracking.

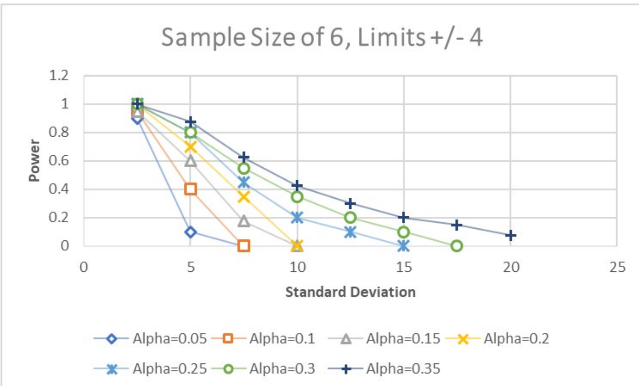
If a satisfactory balance between alpha and power cannot be achieved for particular agency equivalence limits, those limits may need to be reconsidered. The agency's planned use of the data may require a precision not available with current technology, and the pavement asset management system may be providing decision support beyond the capability of the data.

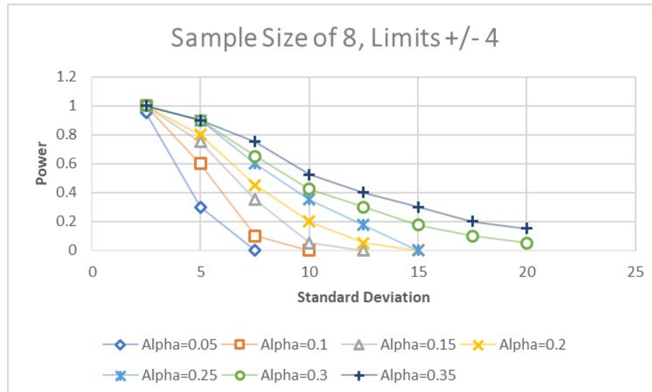
For this example, power curves have been generated for the selected equivalence limits and sample size using the HPMS Cracking Percent criteria for the two one-sided paired t-test for equivalence. The power curves considered most likely to be needed are included in Attachment 1 to this appendix. Summary graphs for the relationships between power, alpha, sample size, and standard deviation for a broader range of values are also provided in Attachment 1.

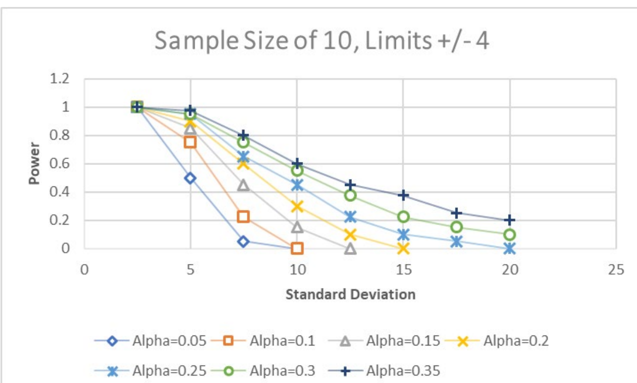
Attachment 1. Power Curves for Two One-Sided Paired T-Test for Equivalence

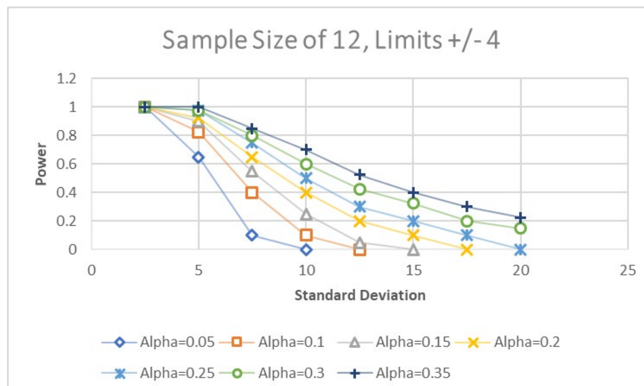
The following sets of Power Curves for the two one-sided paired t-test for equivalence were developed using the Minitab software. [77]

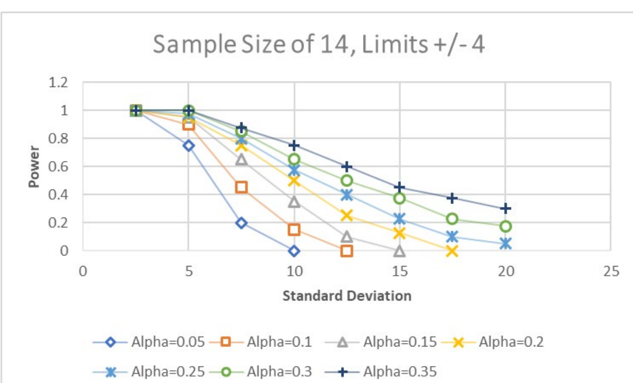
Power Curves for ± 4% Equivalence Limits HPMS Cracking Percent

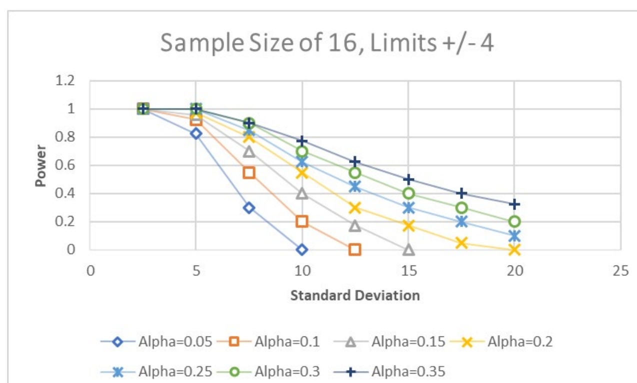


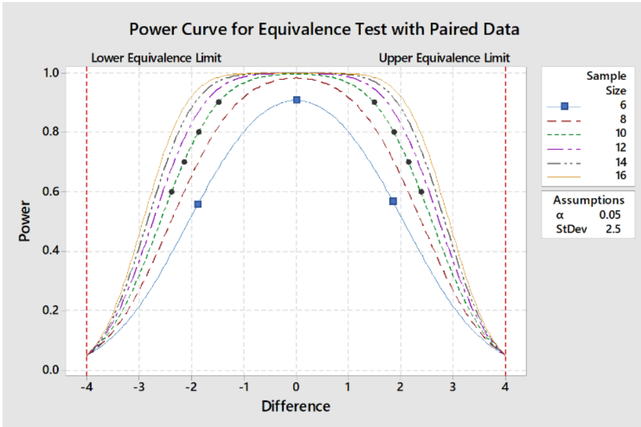


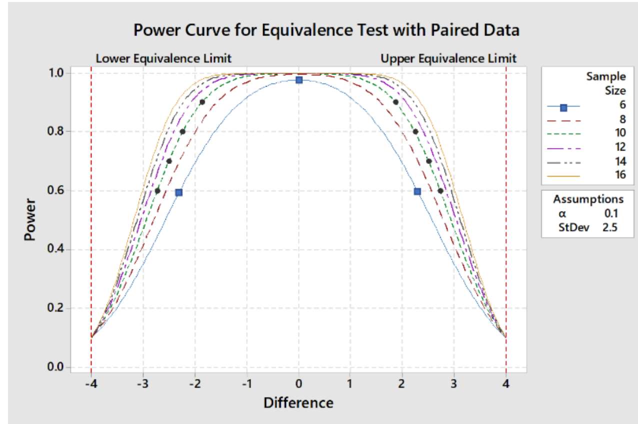


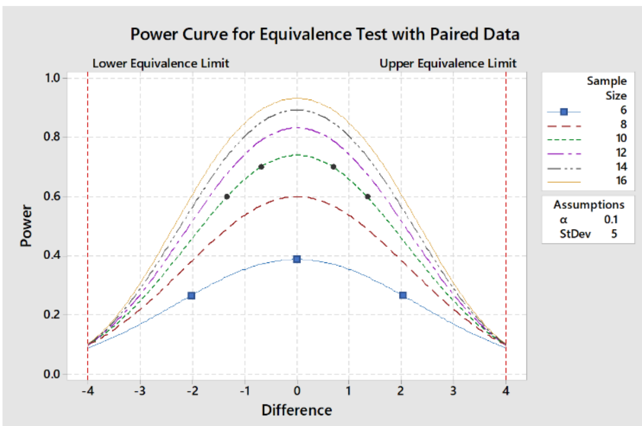


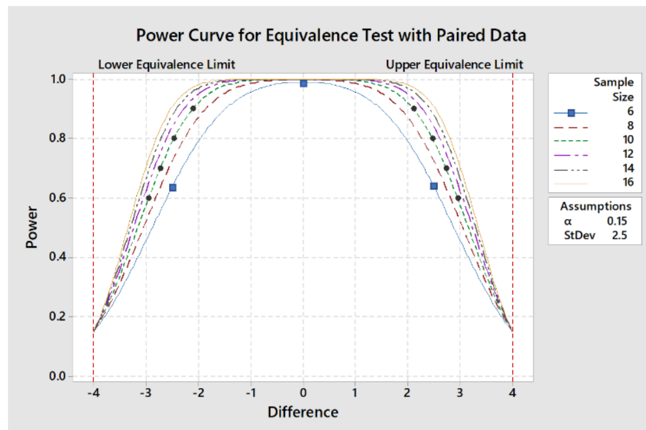


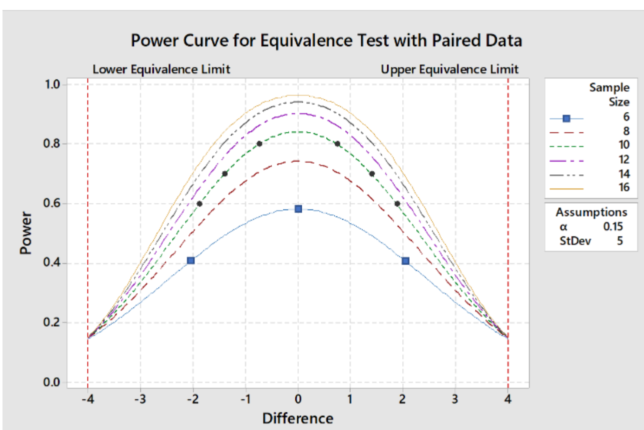


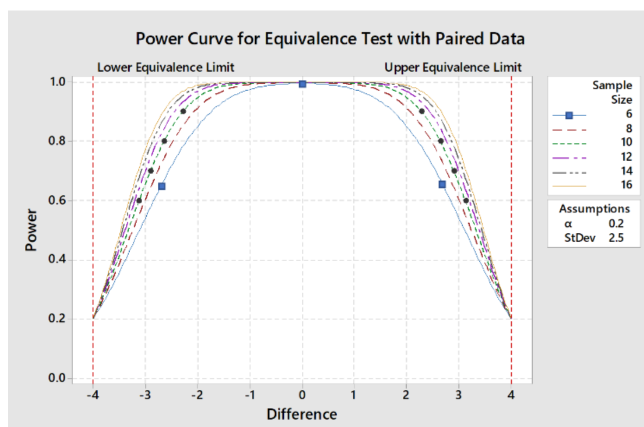

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.05
StDev 2.5

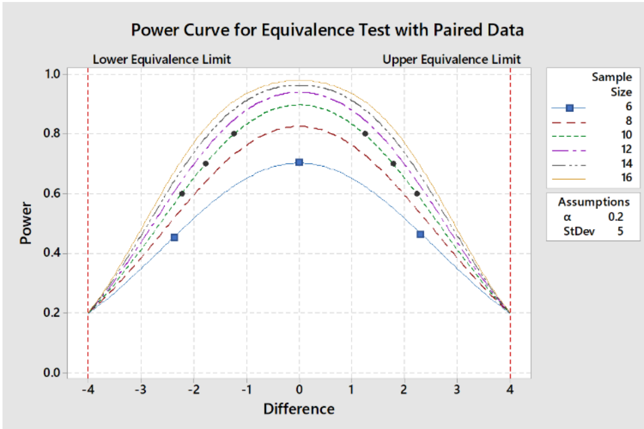

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.1
StDev 2.5

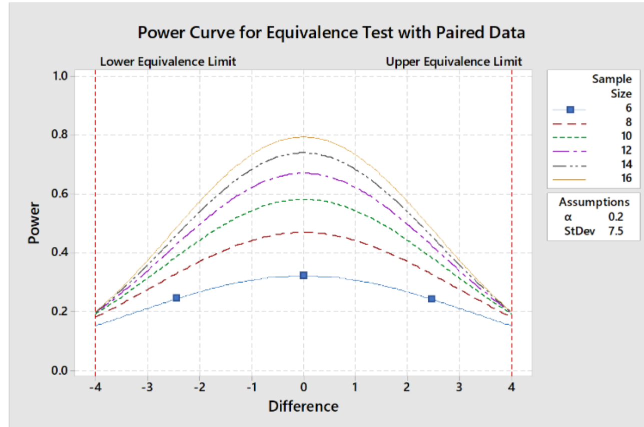

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.1
StDev 5

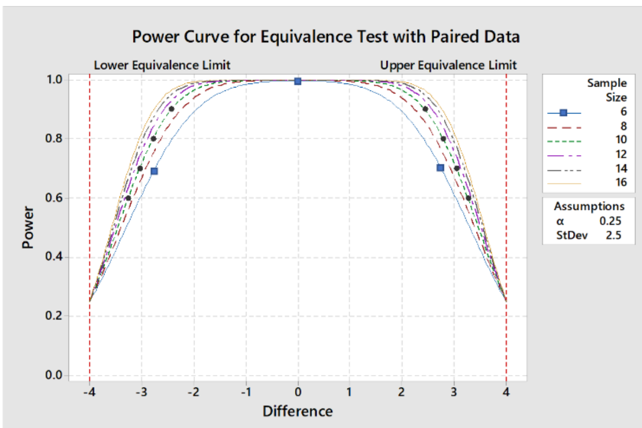

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.15
StDev 2.5

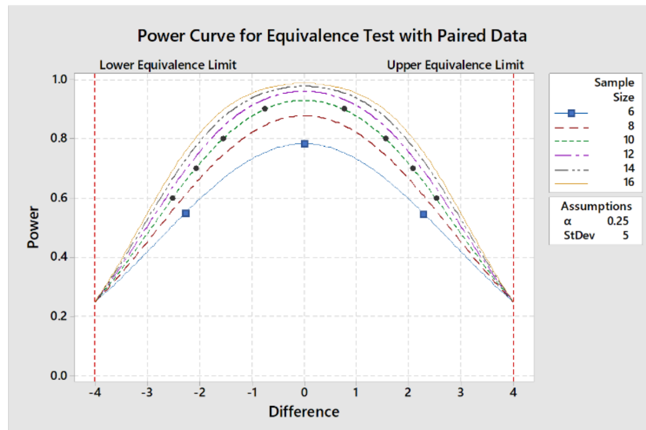

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.15
StDev 5

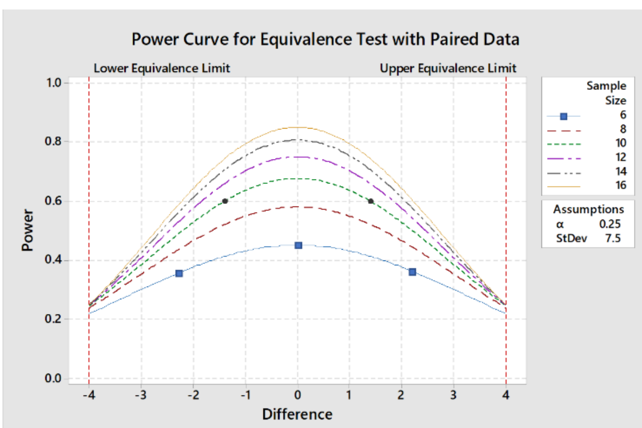

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 2.5

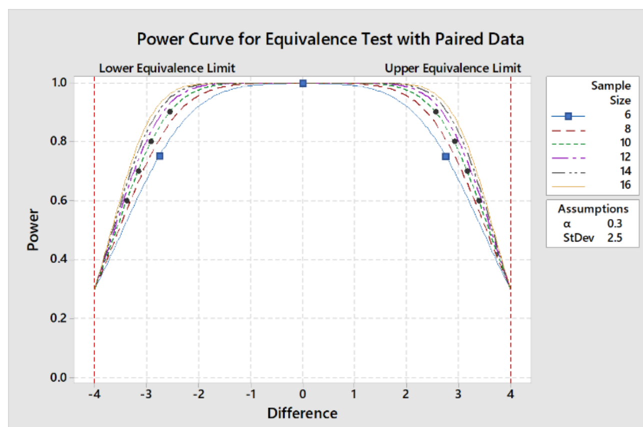

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 5

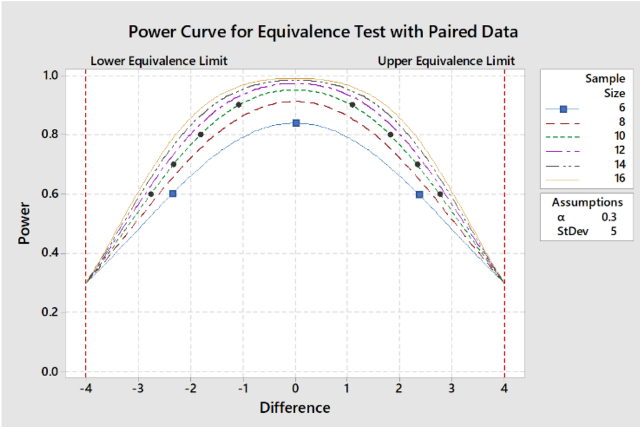

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 7.5

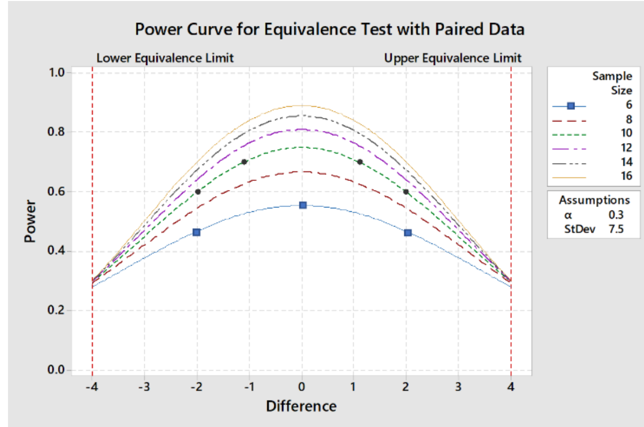

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 2.5

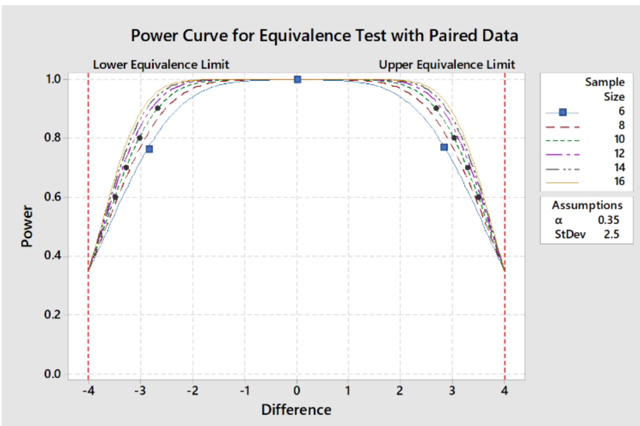

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 5

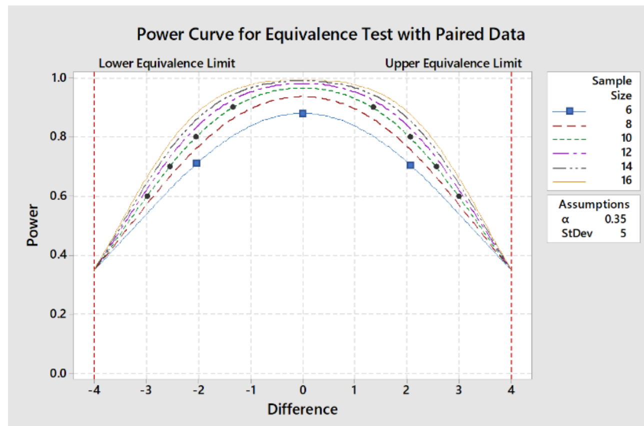

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 7.5

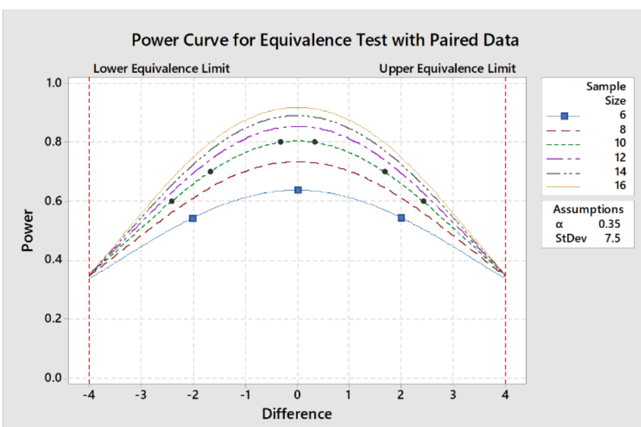

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 2.5

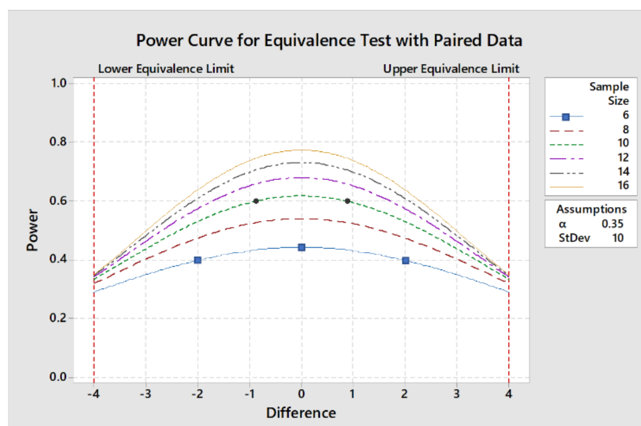

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 5


Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 7.5

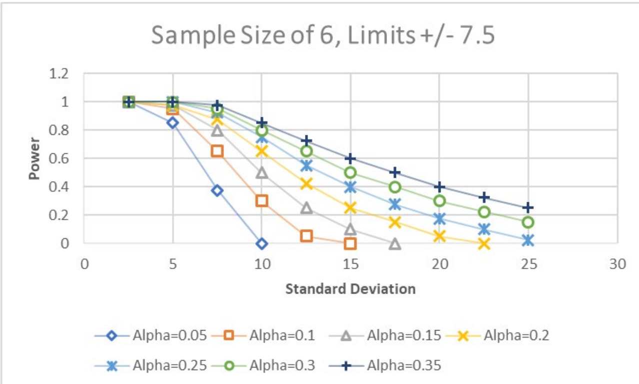

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 2.5

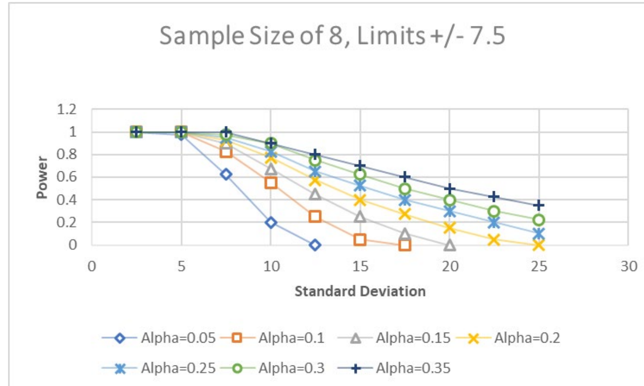

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 5

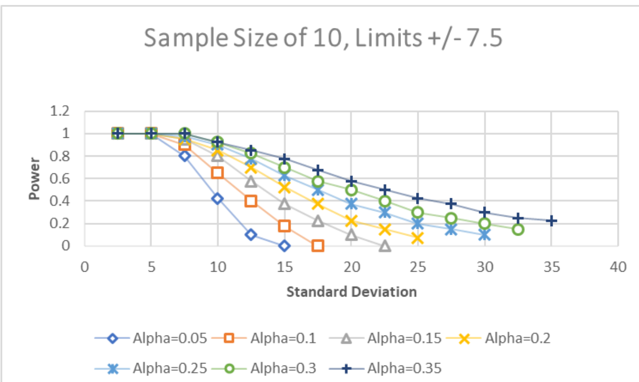

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 7.5

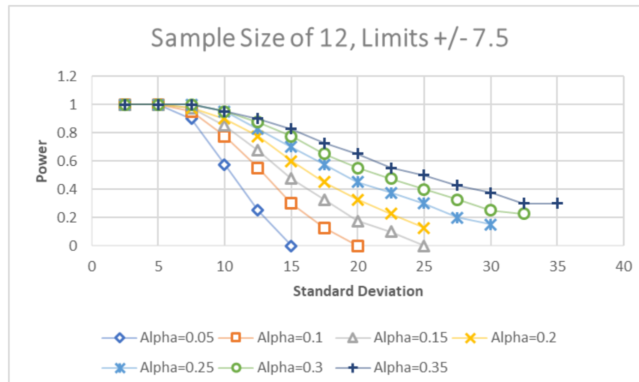

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 10

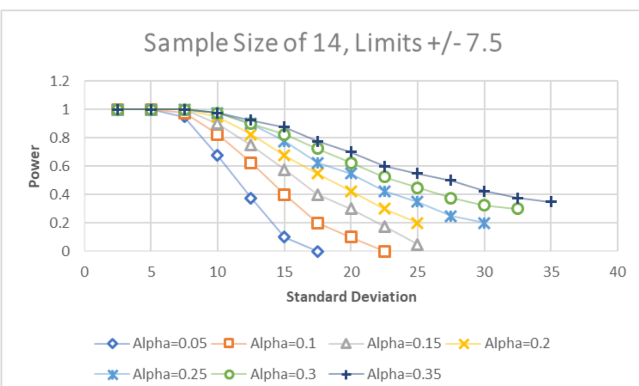
Power Curves for ± 7.5% Equivalence Limits HPMS Cracking Percent

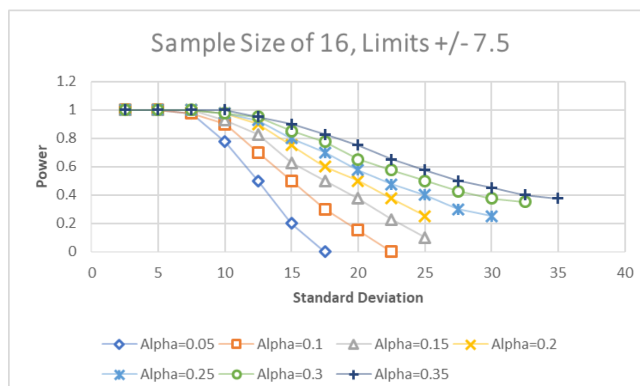


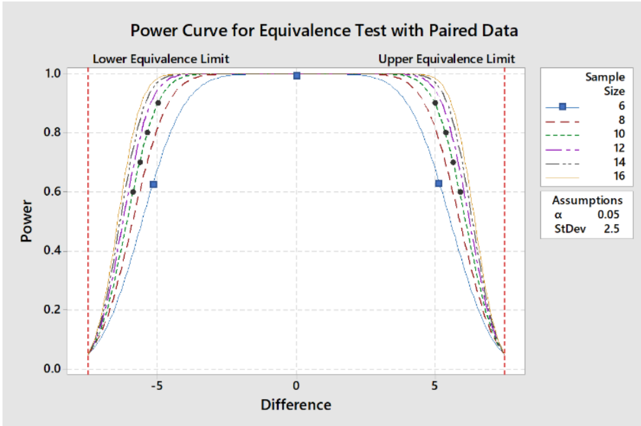


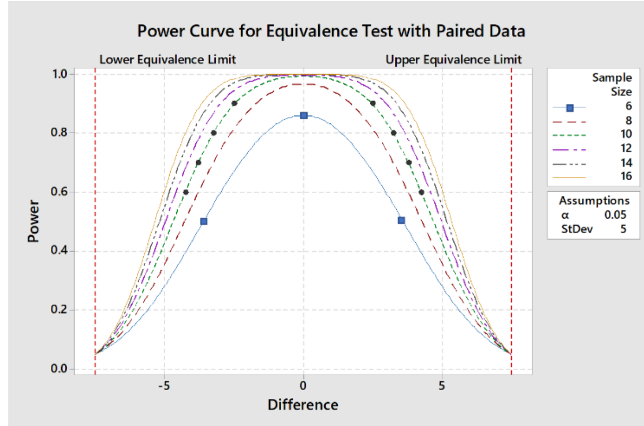


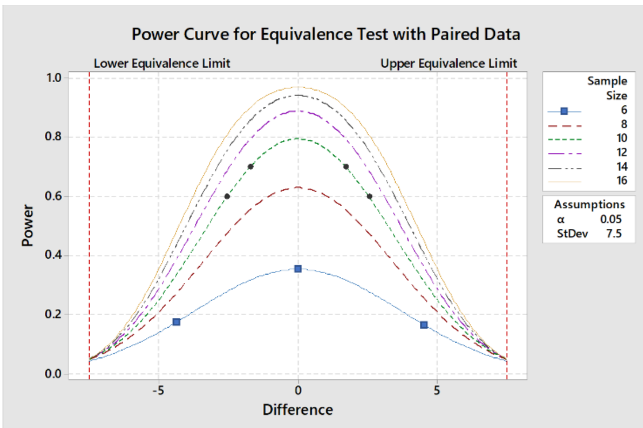


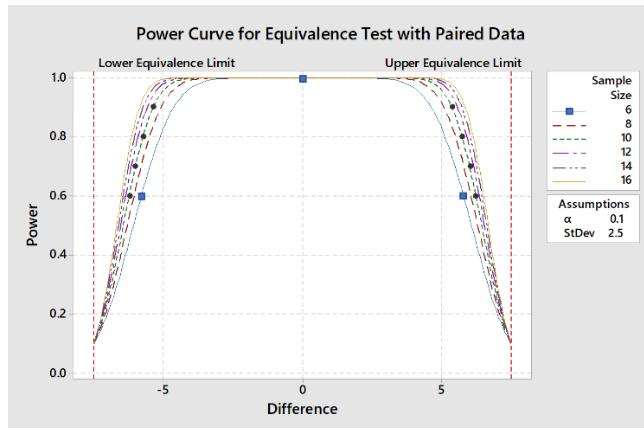


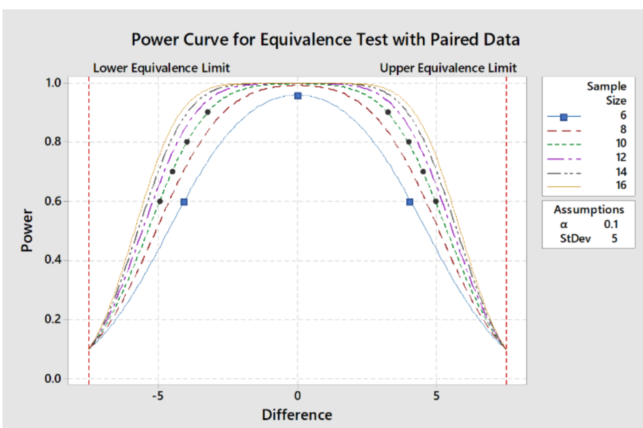


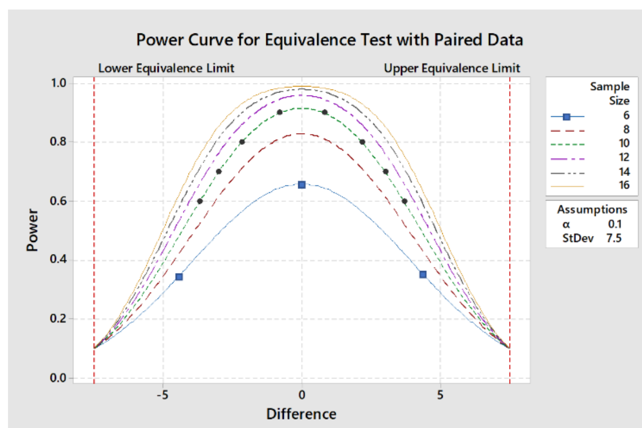

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-5
0
5
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.05
StDev 2.5

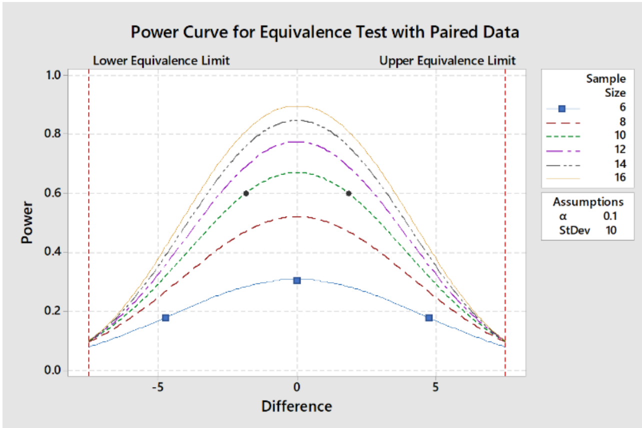

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-5
0
5
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.05
StDev 5

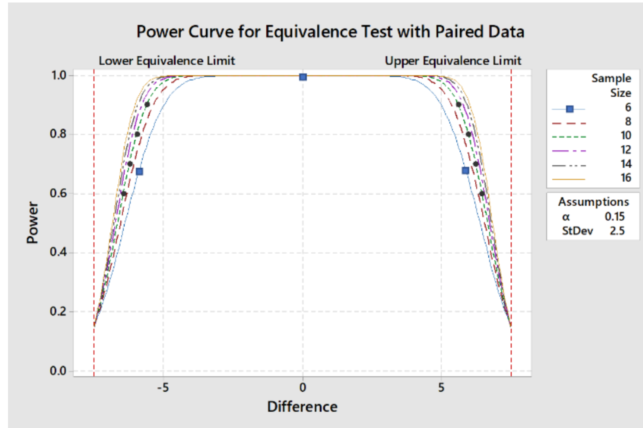

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-5
0
5
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.05
StDev 7.5

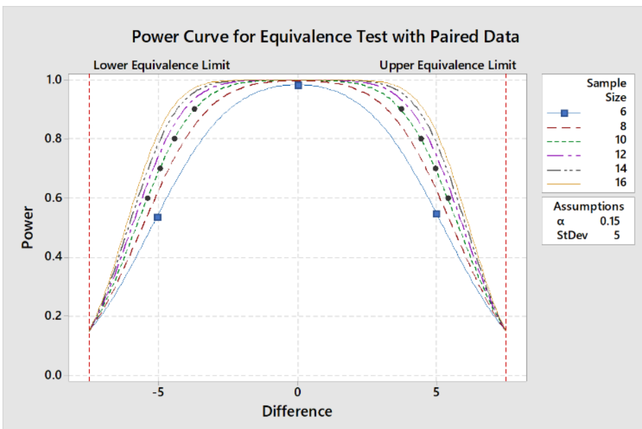

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-5
0
5
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.1
StDev 2.5

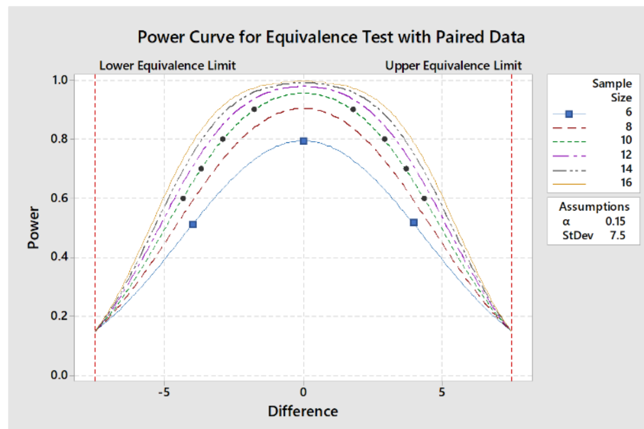

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-5
0
5
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.1
StDev 5

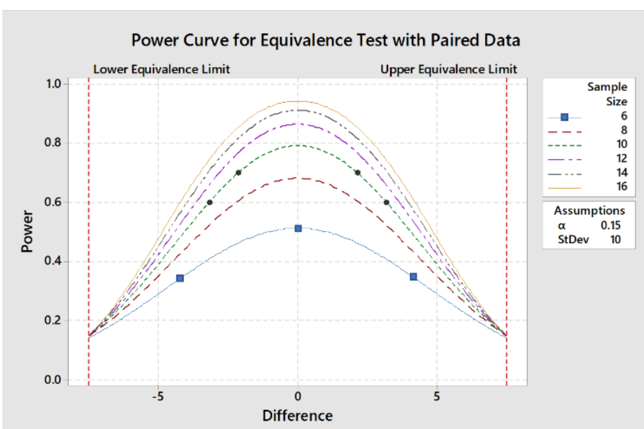

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-5
0
5
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.1
StDev 7.5

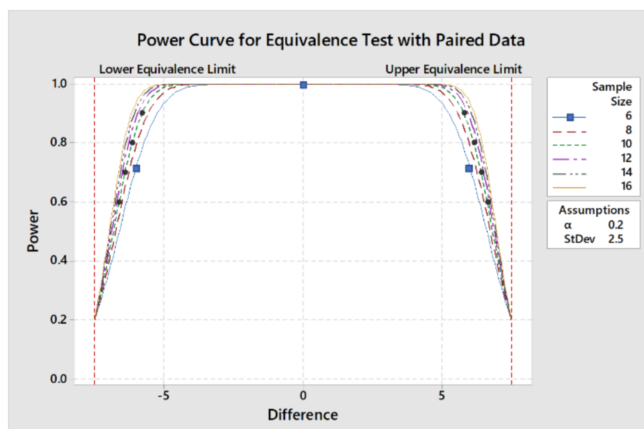

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.1
StDev 10

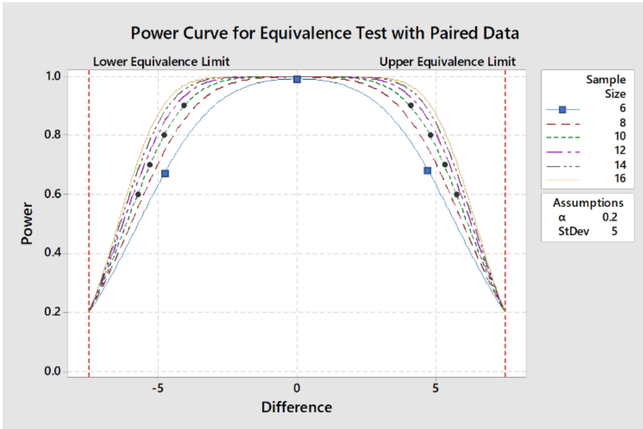

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.15
StDev 2.5

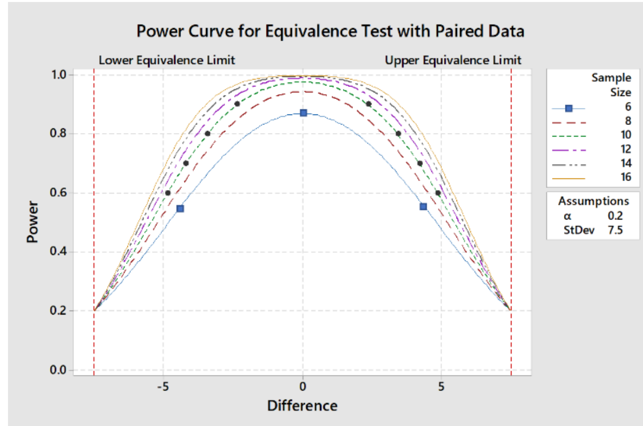

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.15
StDev 5

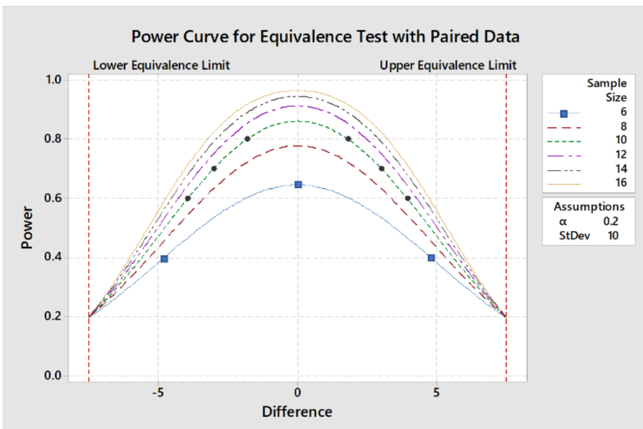

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.15
StDev 7.5

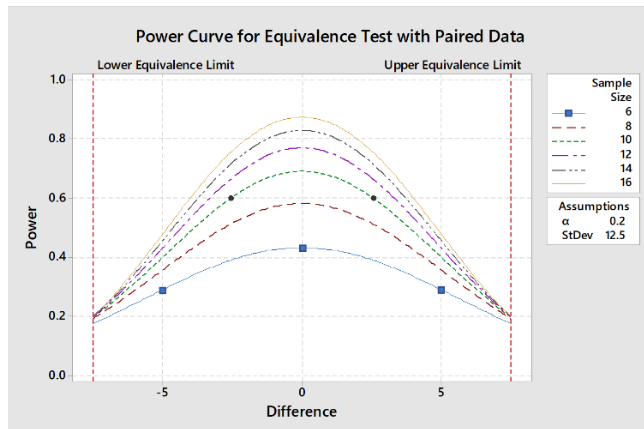

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.15
StDev 10

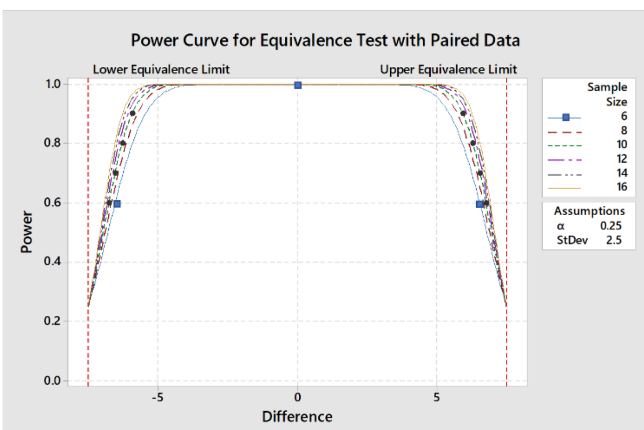

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 2.5

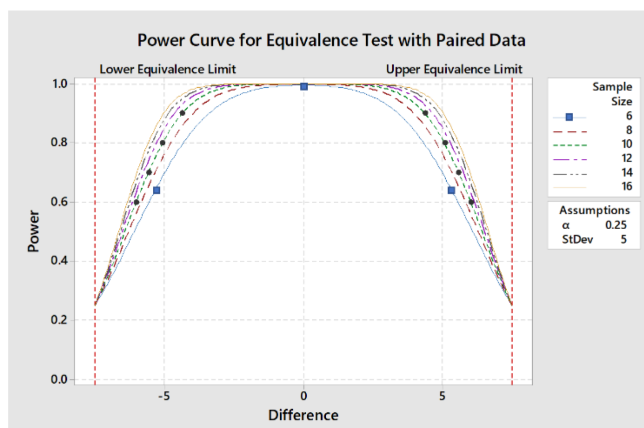

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 5

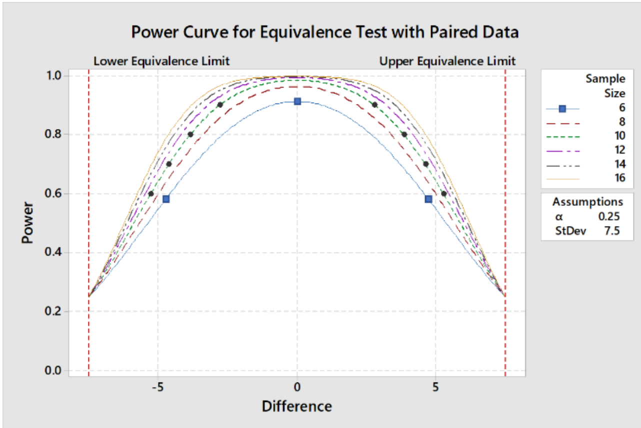

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 7.5

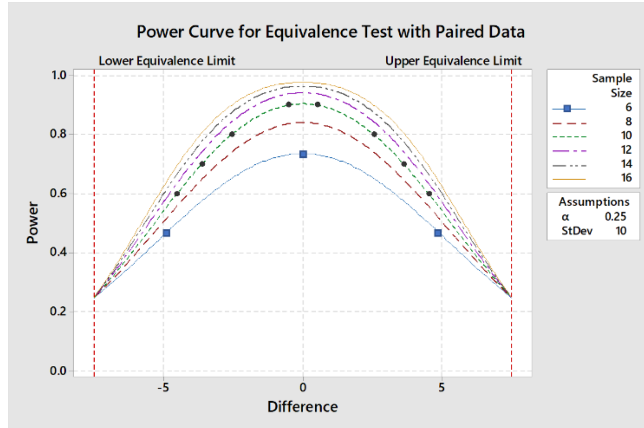

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 10

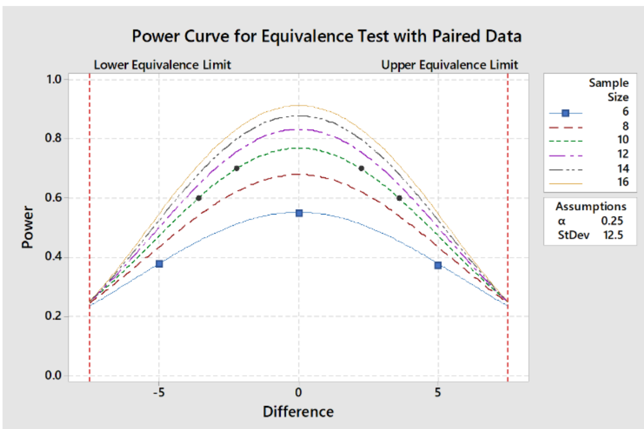

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 12.5

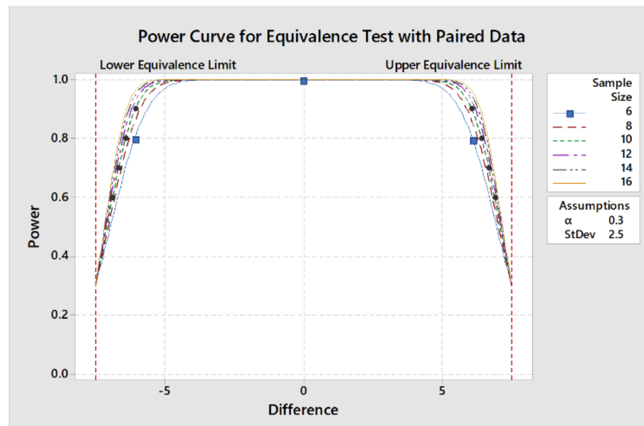

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 2.5

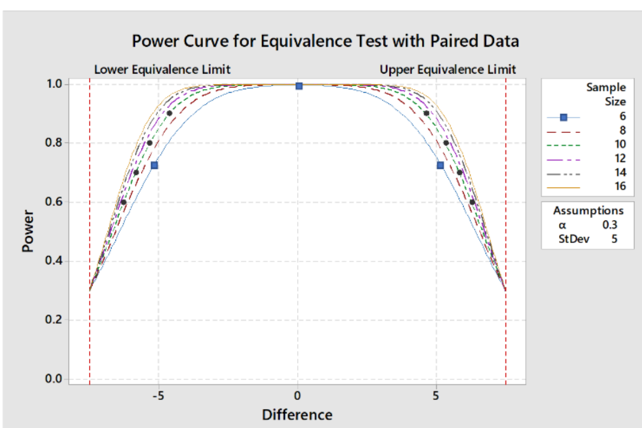

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 5

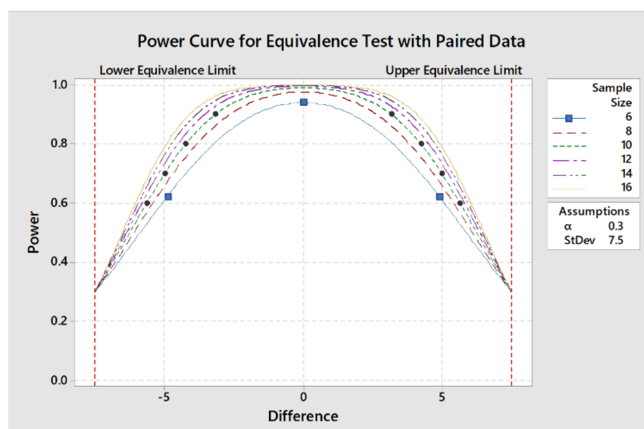

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 7.5

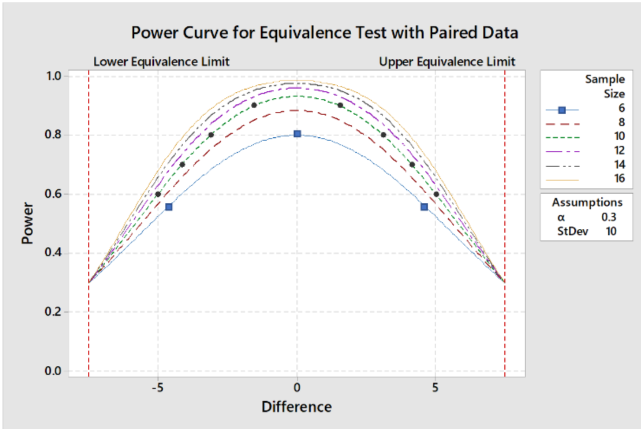

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 10

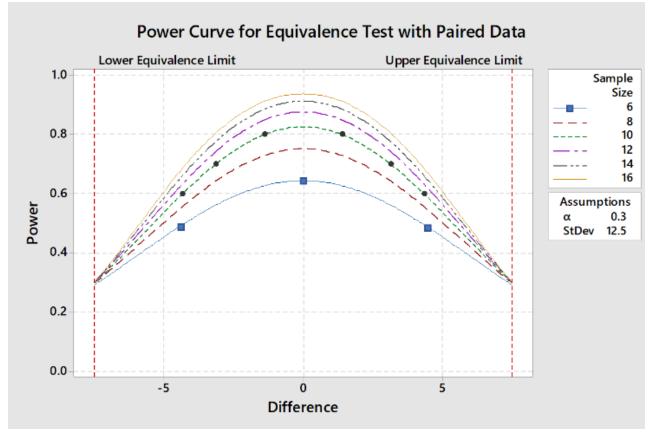

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 12.5

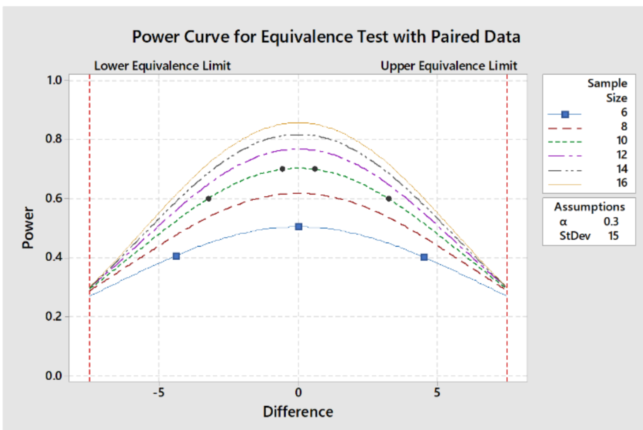

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 2.5

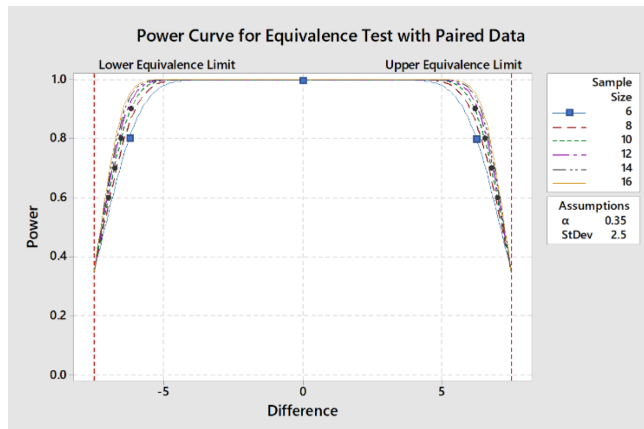

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 5

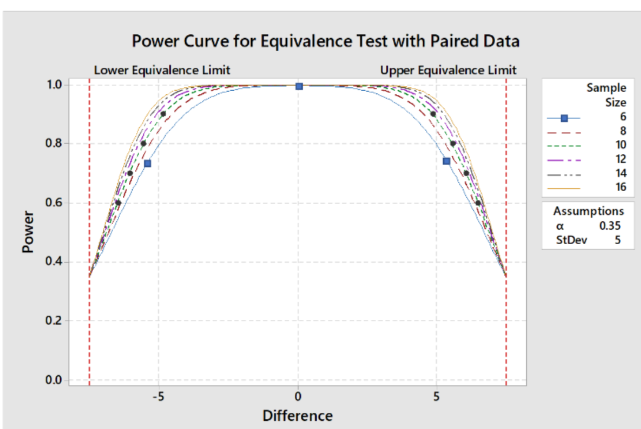

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 7.5

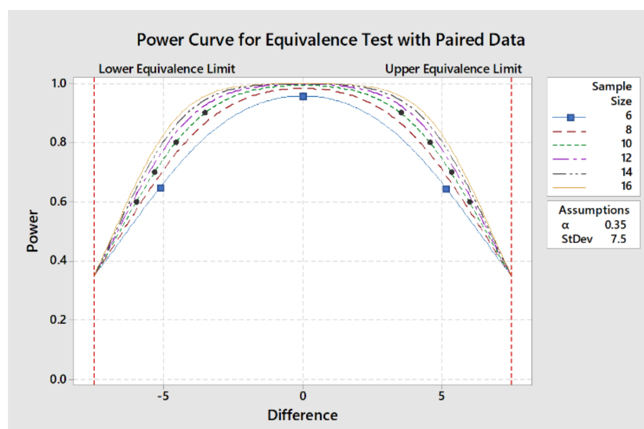

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-5
0
5
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 10

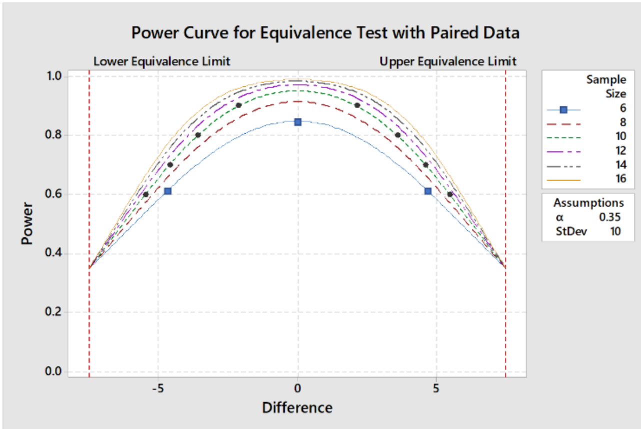

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-5
0
5
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 12.5

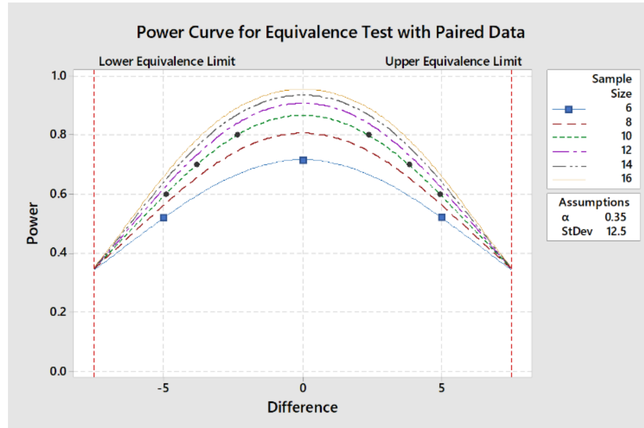

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-5
0
5
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 15

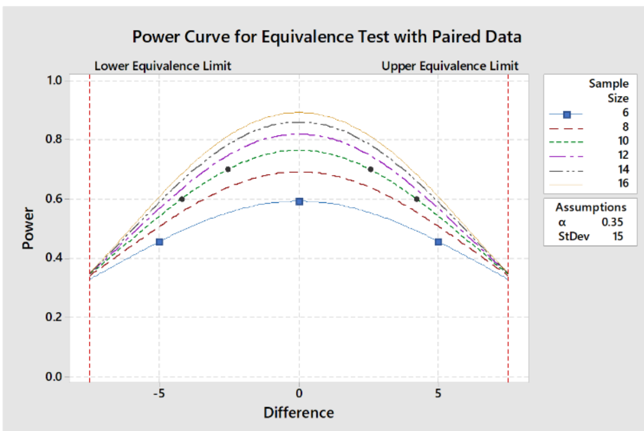

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-5
0
5
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 2.5

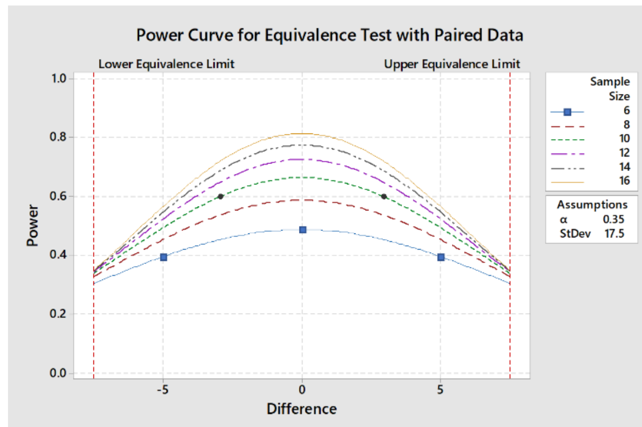

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-5
0
5
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 5


Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-5
0
5
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 7.5

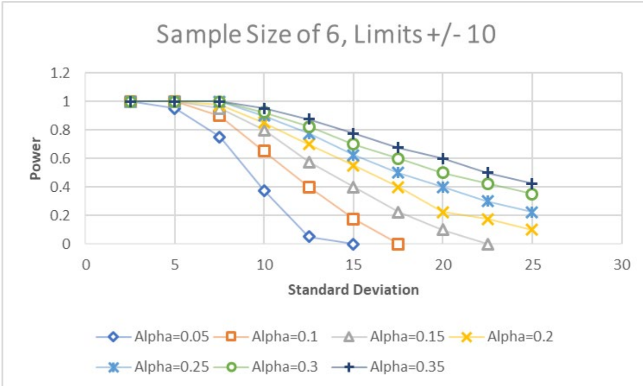

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 10

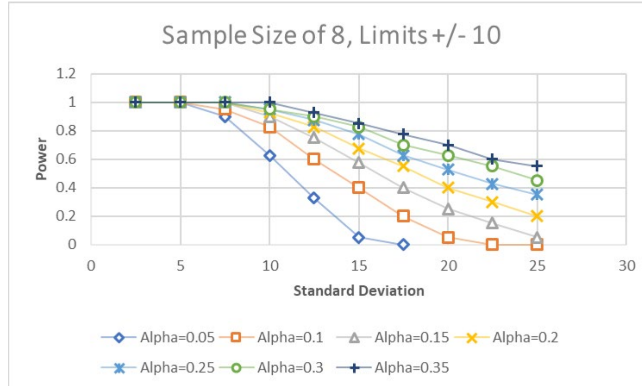

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 12.5

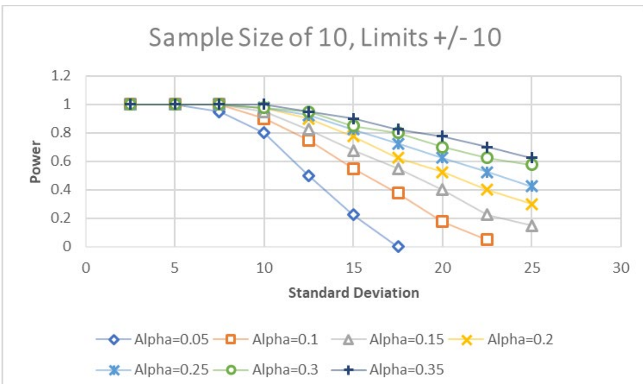

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 15

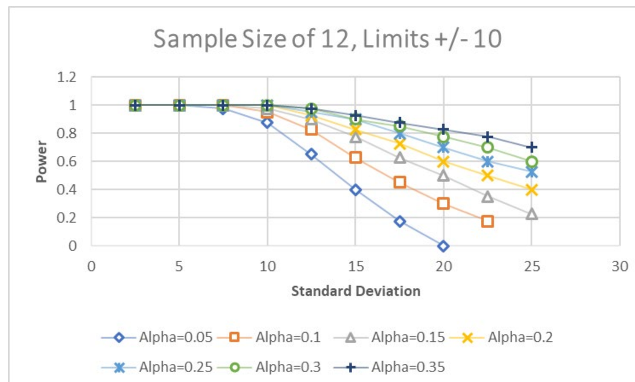

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 17.5

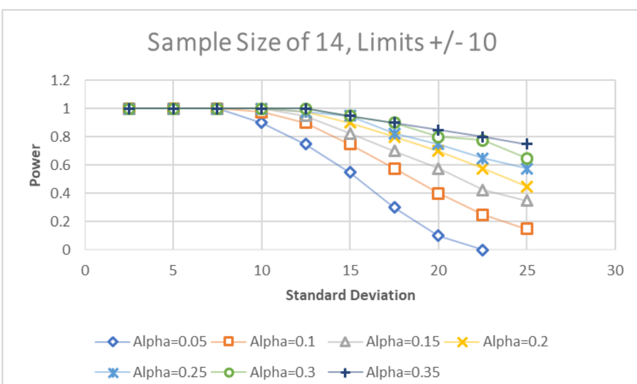
Power Curves for ± 10% Equivalence Limits HPMS Cracking Percent

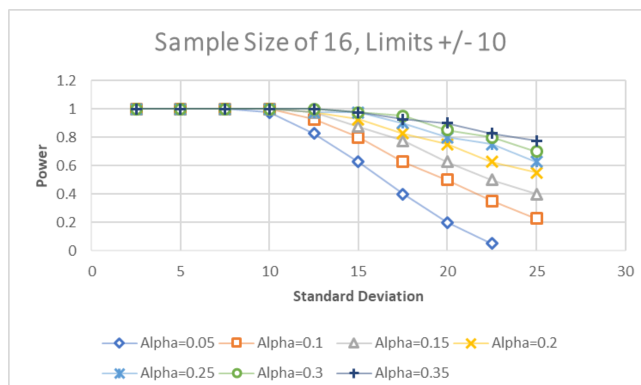


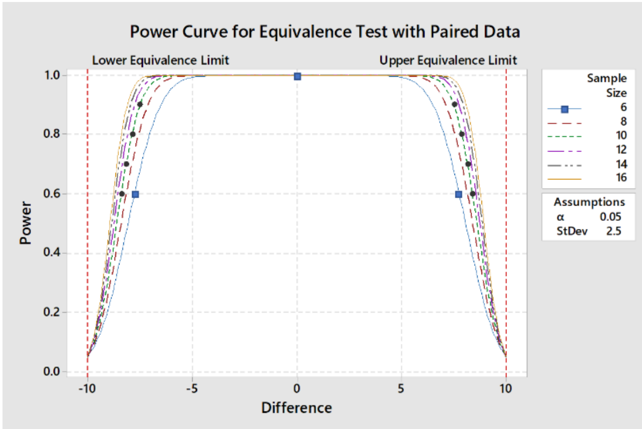


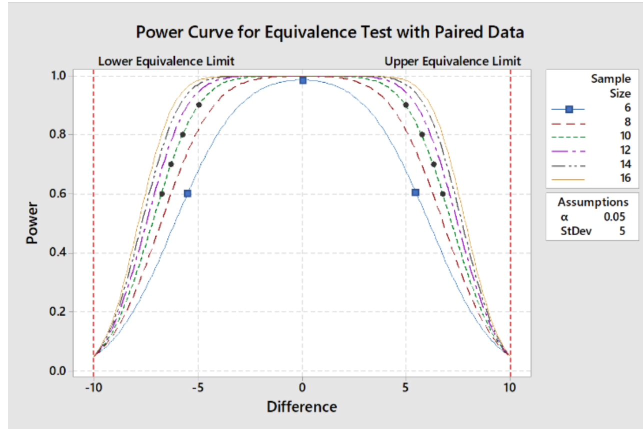


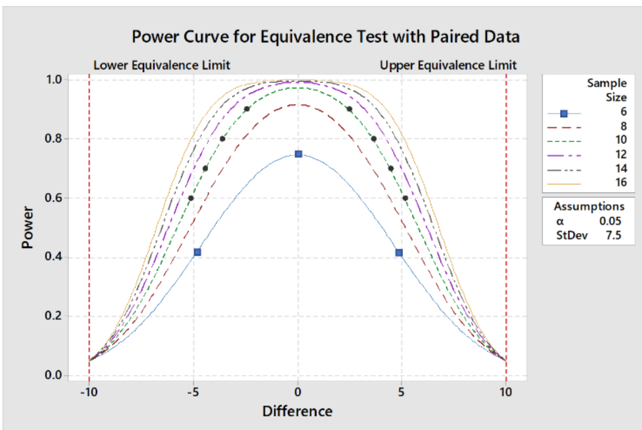


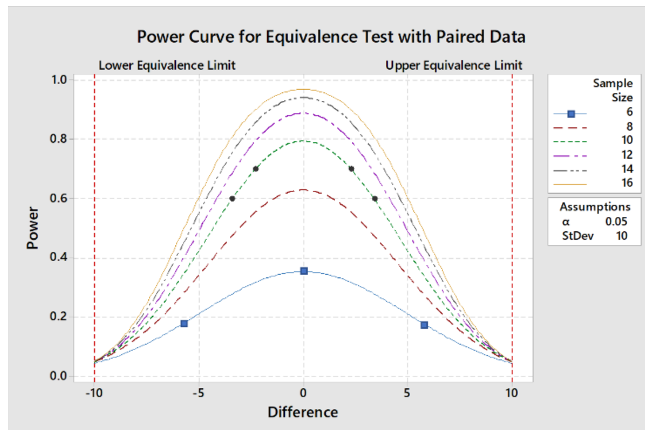


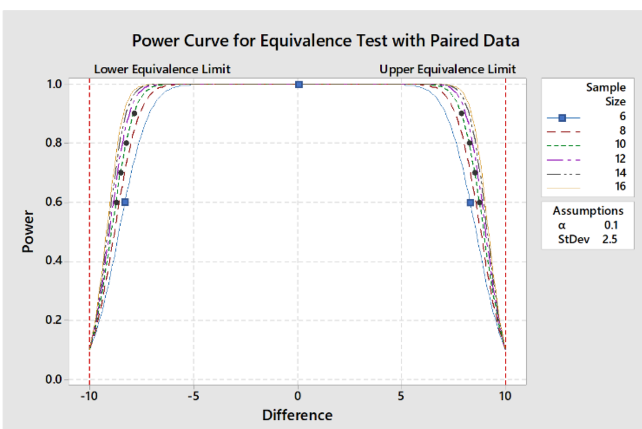


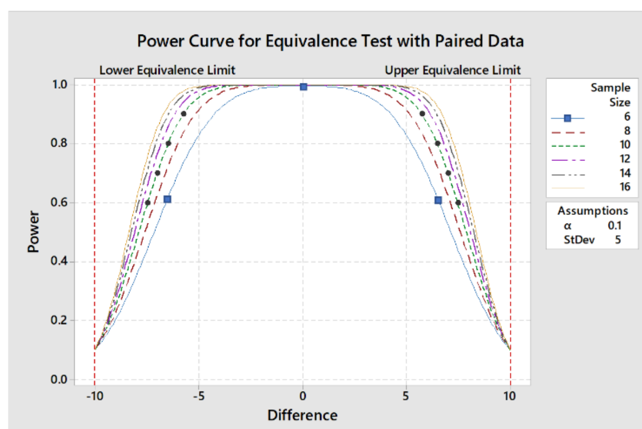

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.05
StDev 2.5

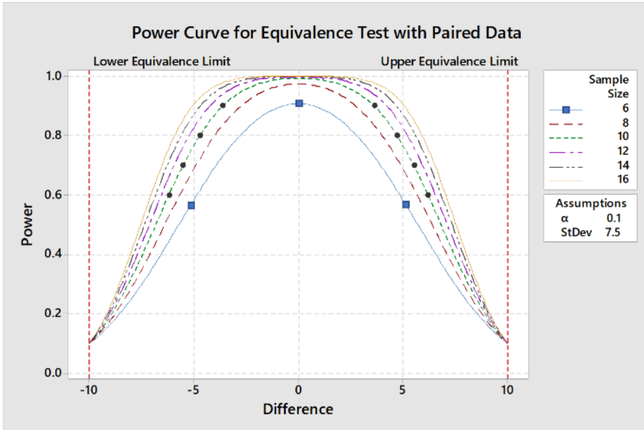

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.05
StDev 5

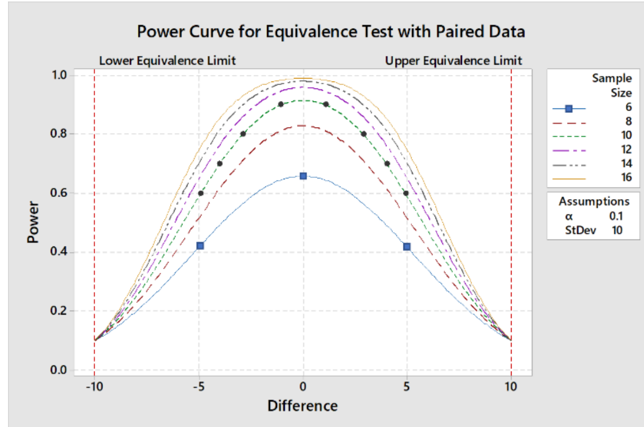

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.05
StDev 7.5

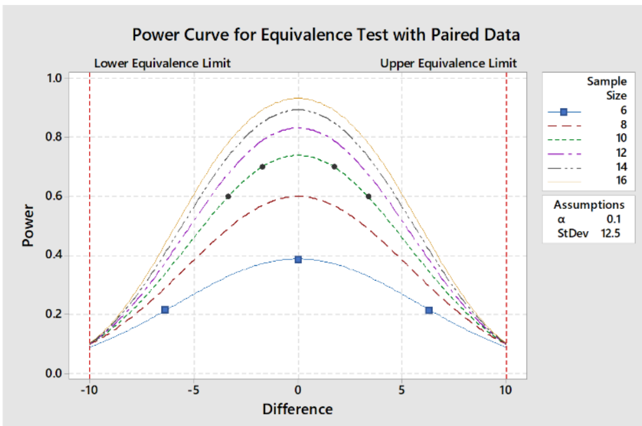

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.05
StDev 10

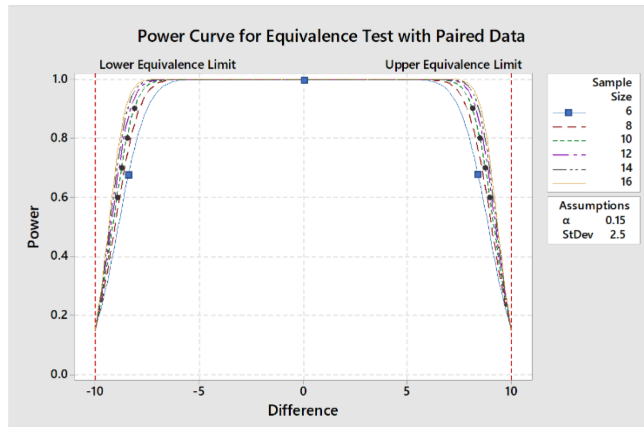

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.1
StDev 2.5

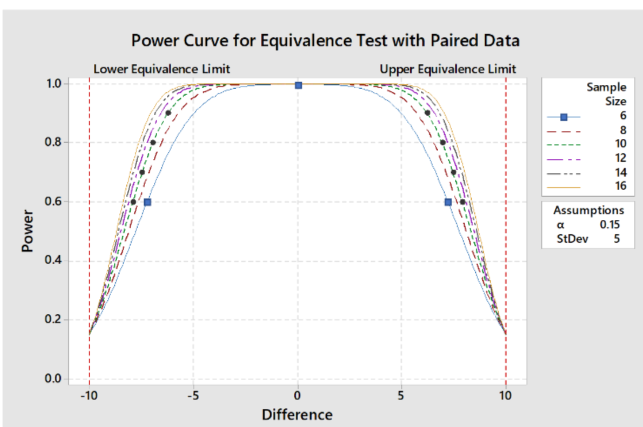

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.1
StDev 5

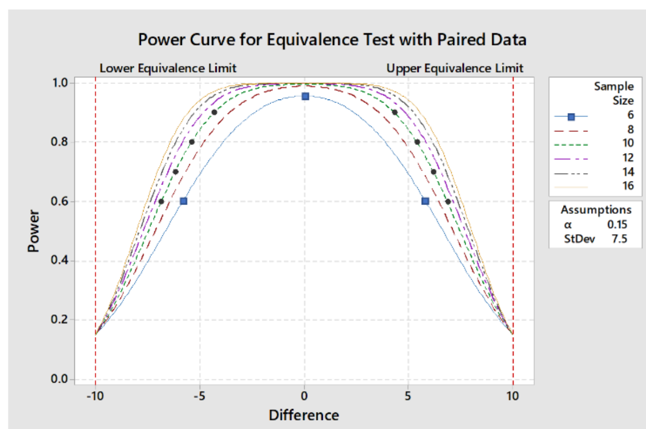

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Sample Size
6
8
10
12
14
16
Assumptions
α 0.1
StDev 7.5

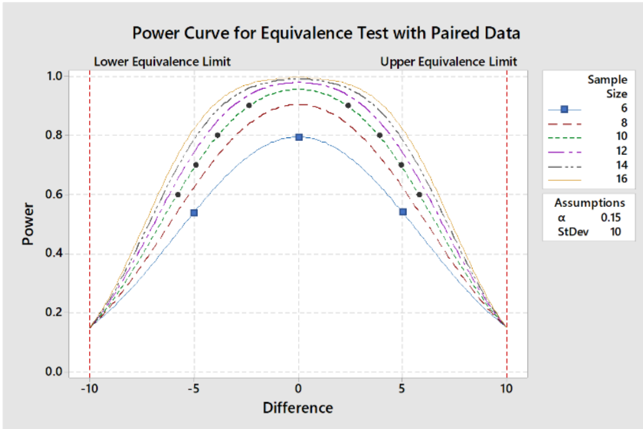

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Sample Size
6
8
10
12
14
16
Assumptions
α 0.1
StDev 10

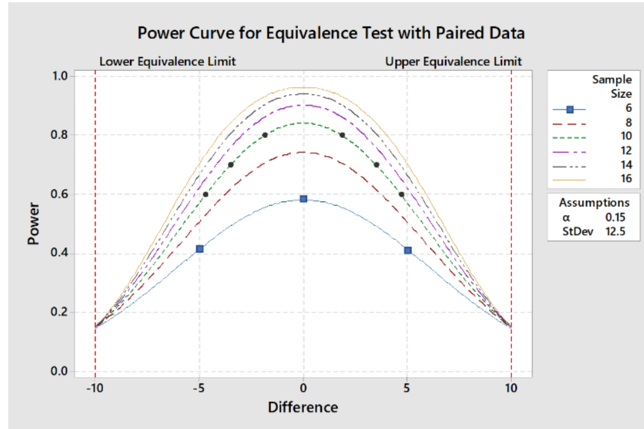

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Sample Size
6
8
10
12
14
16
Assumptions
α 0.1
StDev 12.5

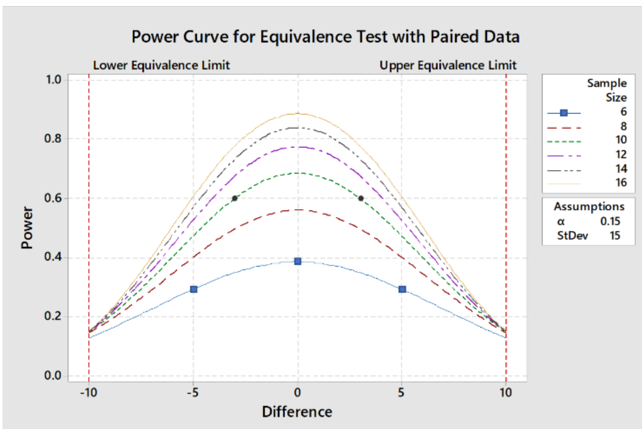

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Sample Size
6
8
10
12
14
16
Assumptions
α 0.15
StDev 2.5

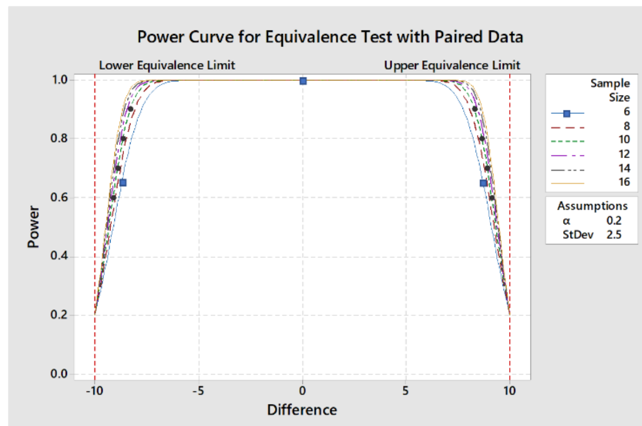

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Sample Size
6
8
10
12
14
16
Assumptions
α 0.15
StDev 5

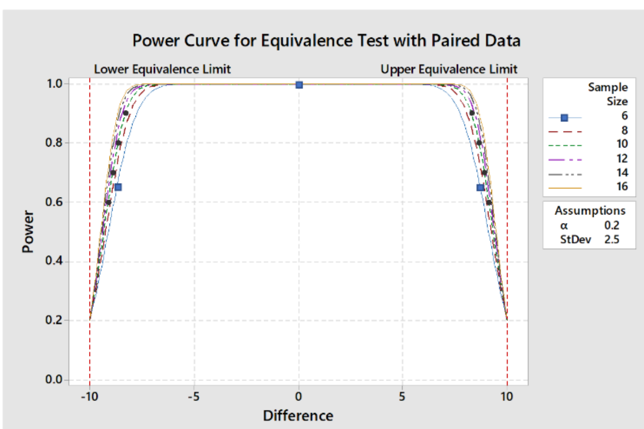

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Sample Size
6
8
10
12
14
16
Assumptions
α 0.15
StDev 7.5

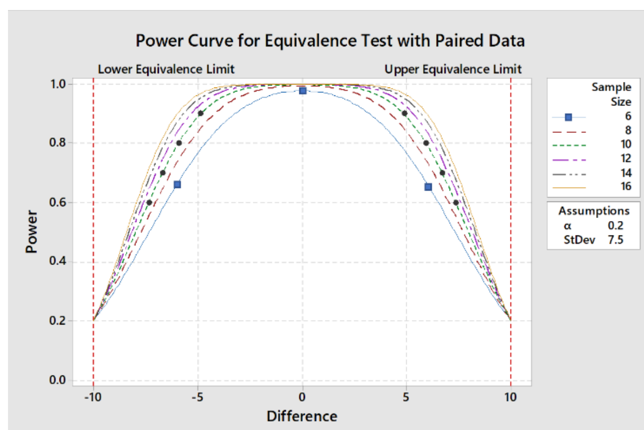

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Sample Size
6
8
10
12
14
16
Assumptions
α 0.15
StDev 10

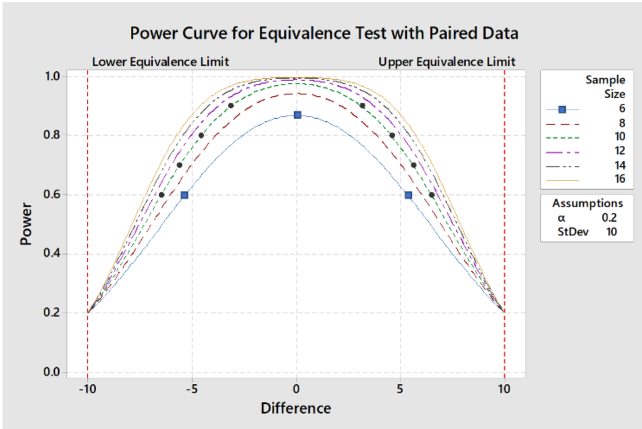

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Sample Size
6
8
10
12
14
16
Assumptions
α 0.15
StDev 12.5

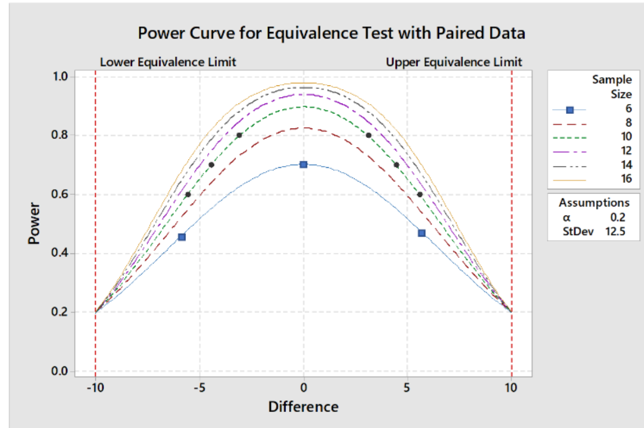

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Sample Size
6
8
10
12
14
16
Assumptions
α 0.15
StDev 15

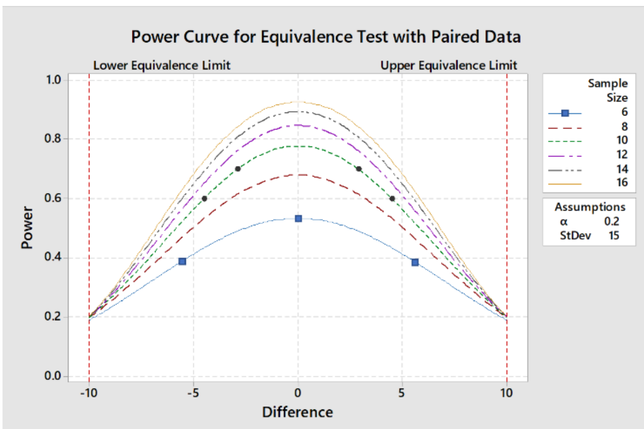

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 2.5

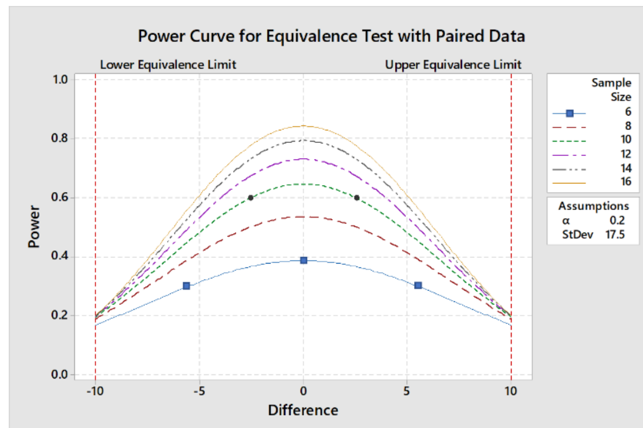

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 2.5

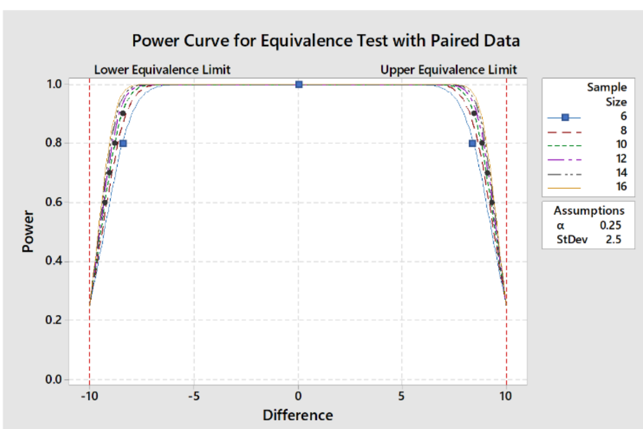

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 7.5

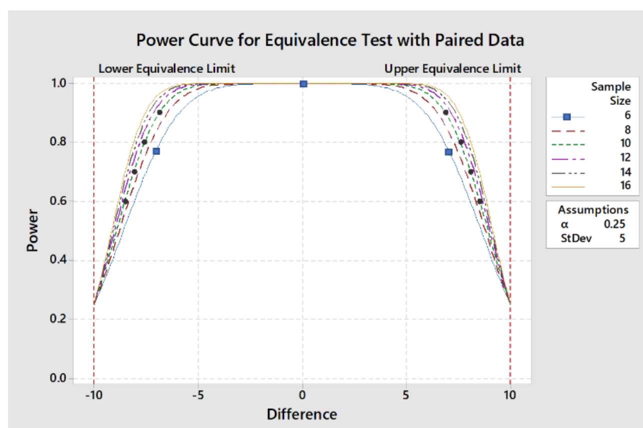

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 10

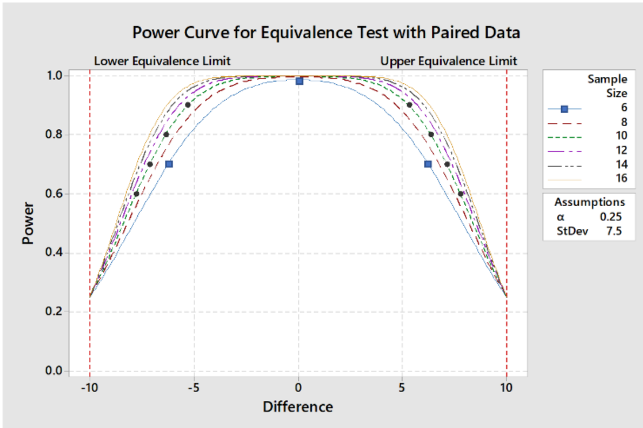

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 12.5

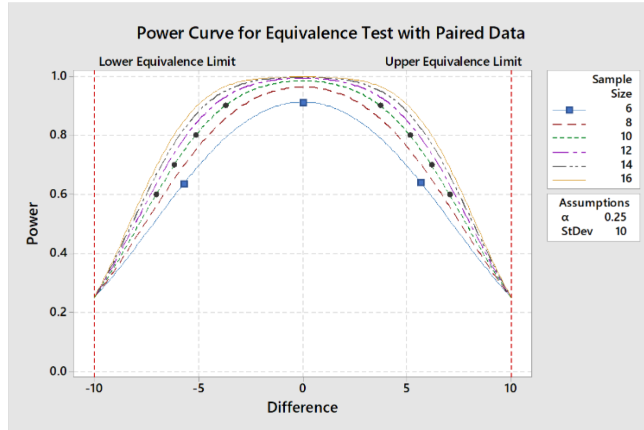

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 15

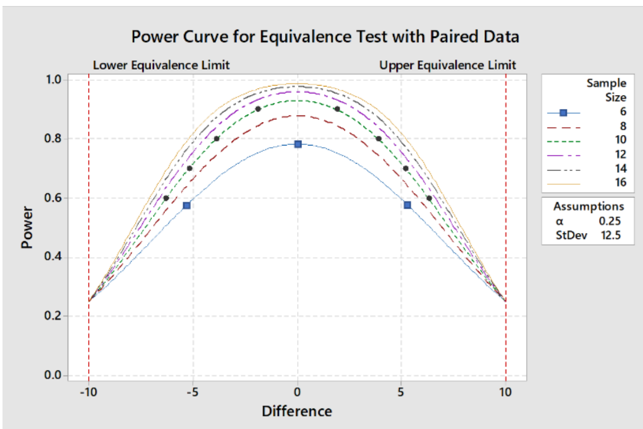

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.2
StDev 17.5

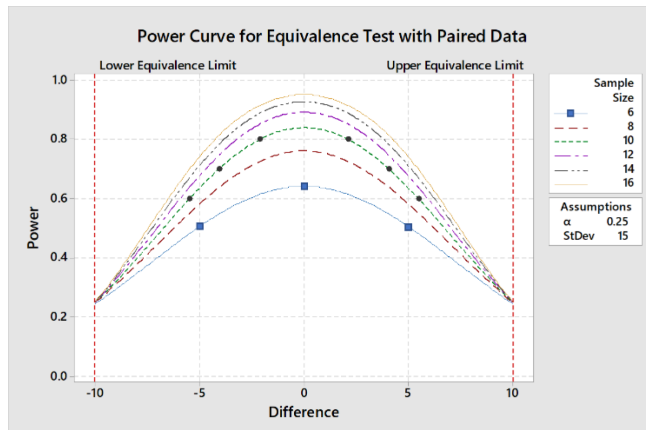

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 2.5

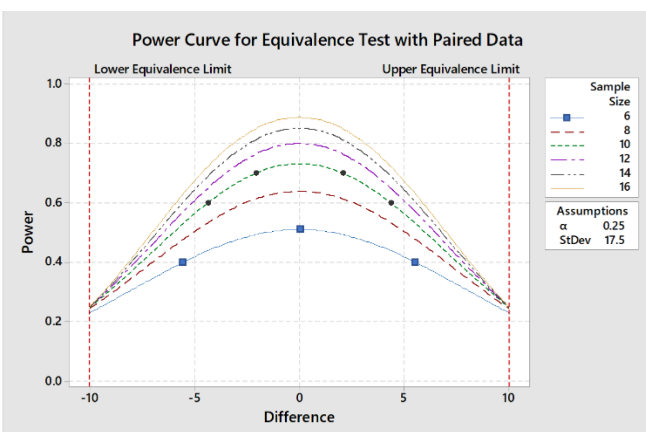

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 5

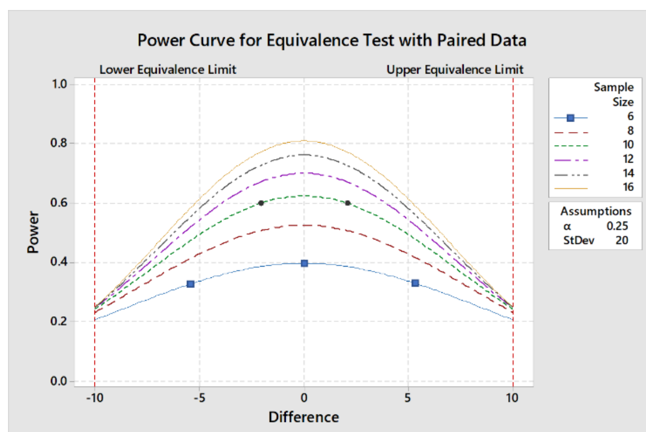

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 7.5
Power
Difference

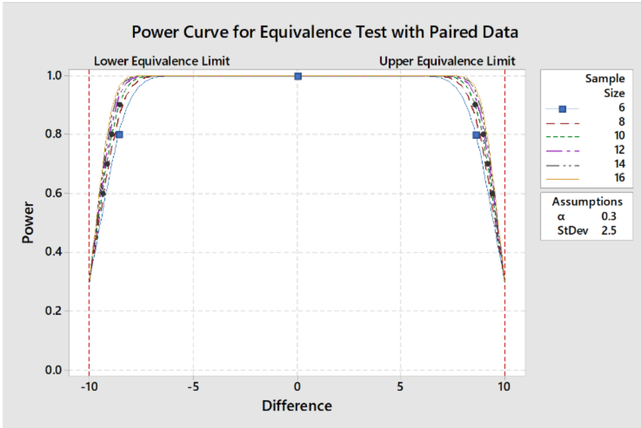

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 10
Power
Difference

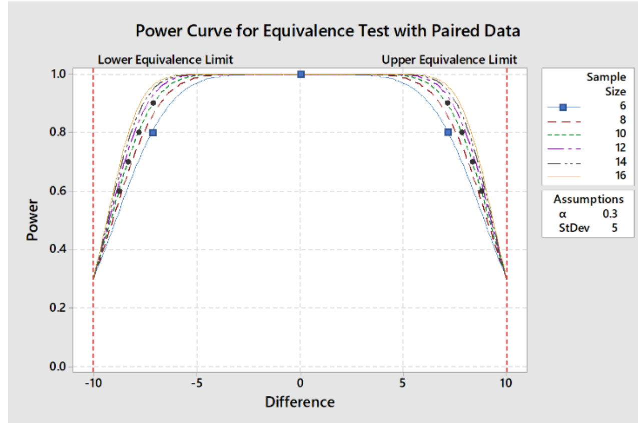

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 12.5
Power
Difference

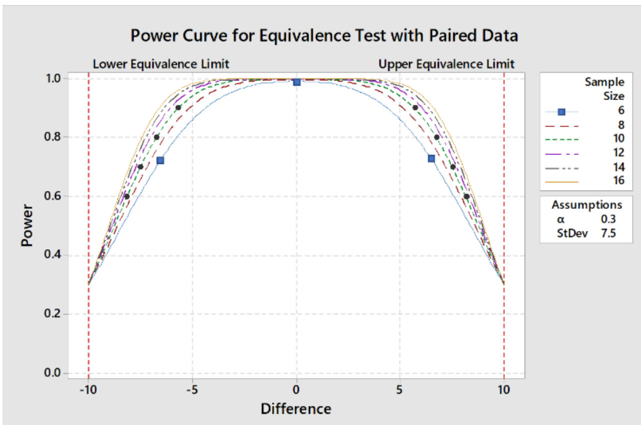

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 15
Power
Difference

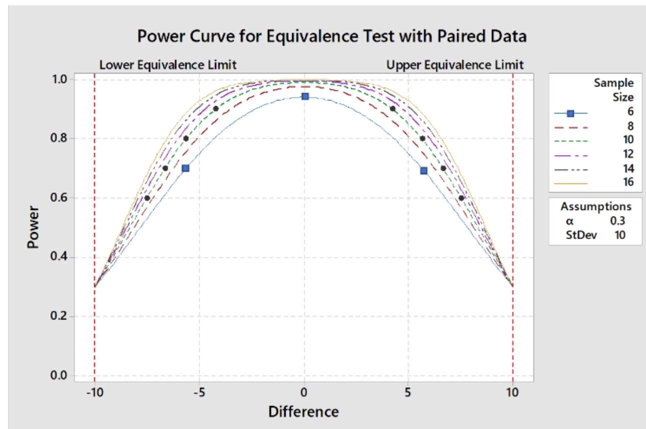

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 17.5
Power
Difference

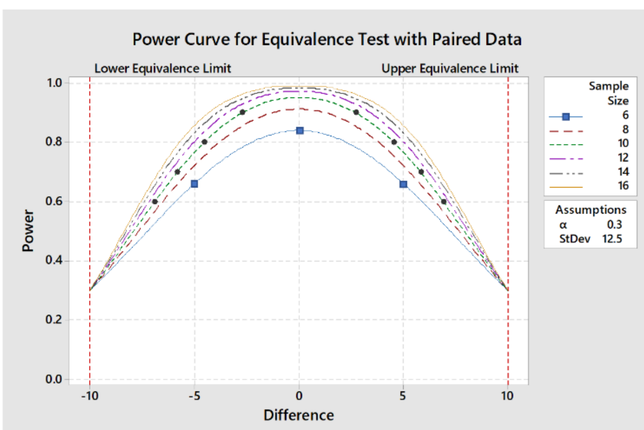

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Sample Size
6
8
10
12
14
16
Assumptions
α 0.25
StDev 20
Power
Difference

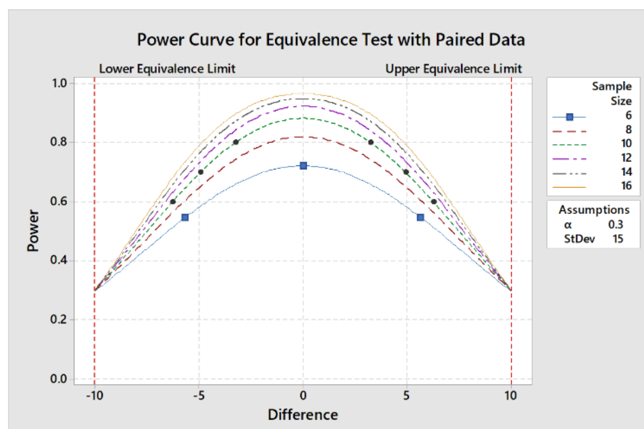

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 2.5

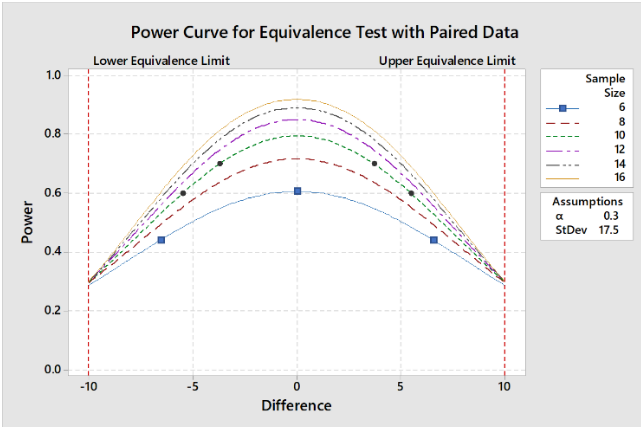

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 5

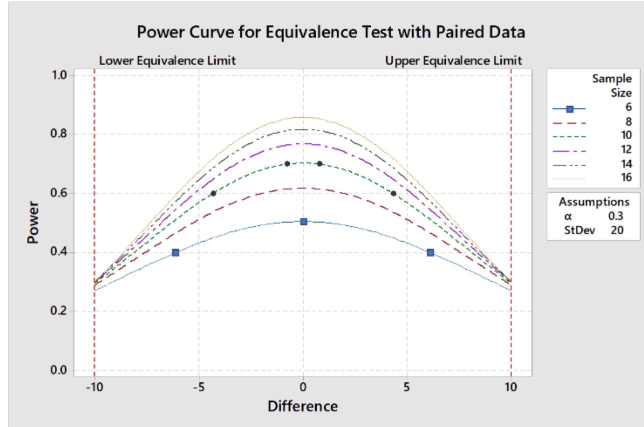

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 7.5

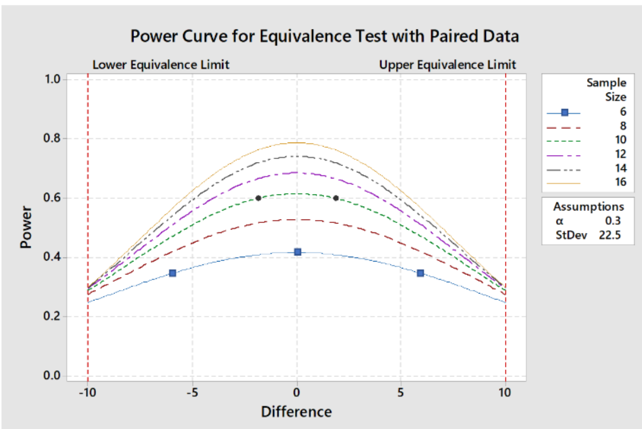

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 10

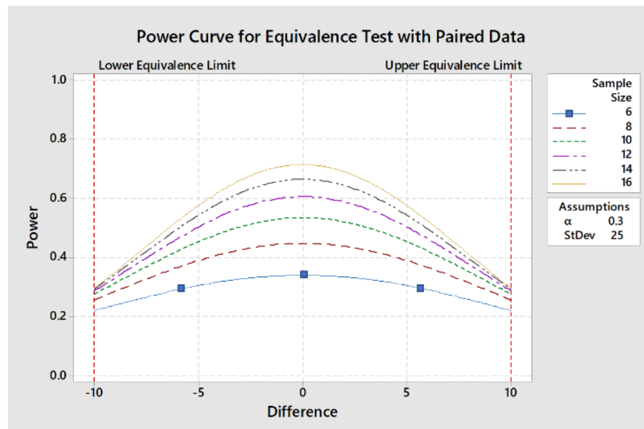

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 12.5

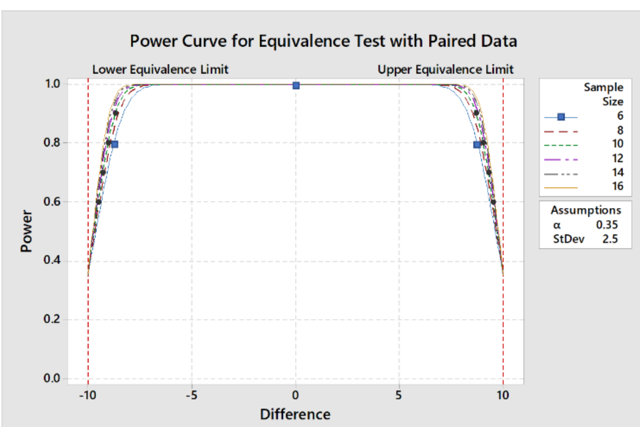

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 15

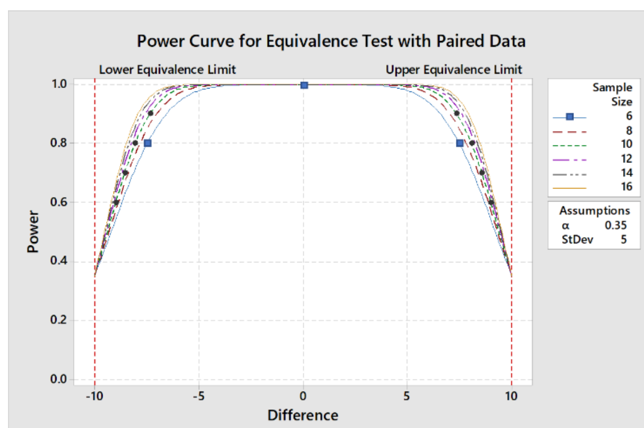

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 17.5

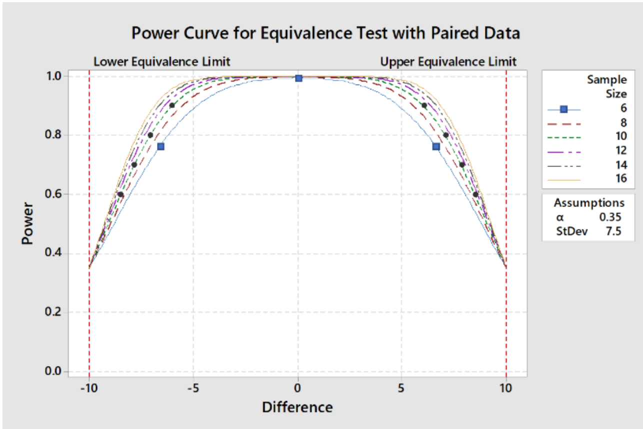

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 20

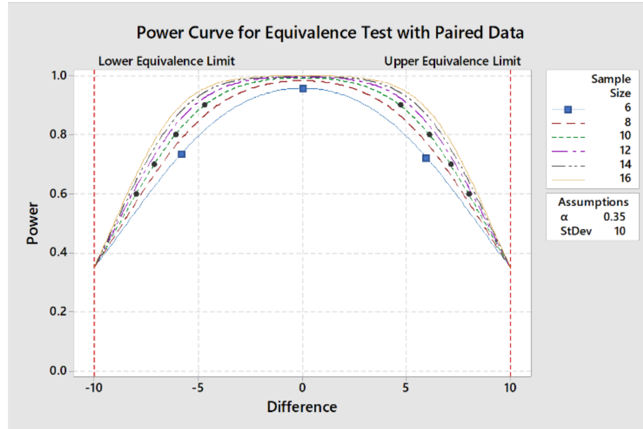

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 22.5

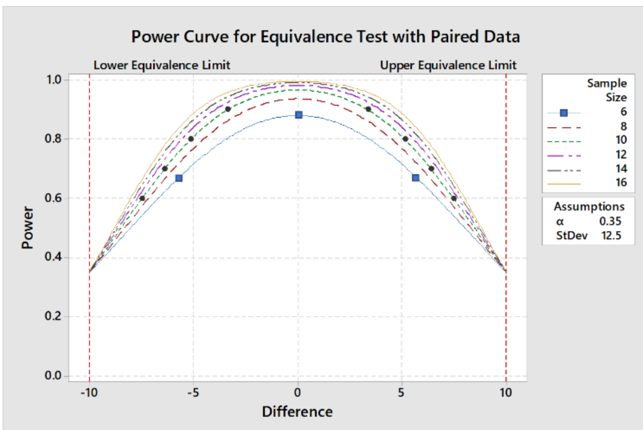

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.3
StDev 25

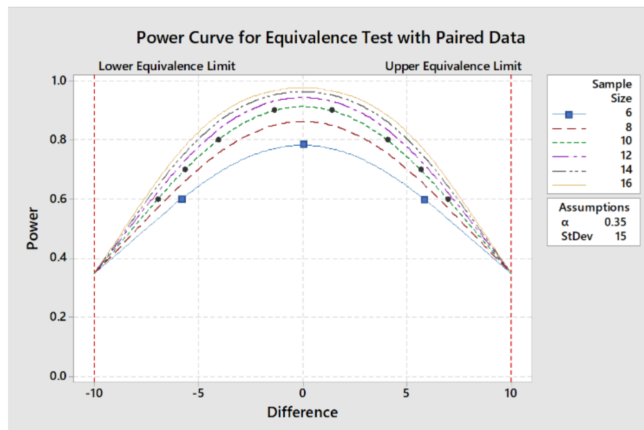

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 2.5

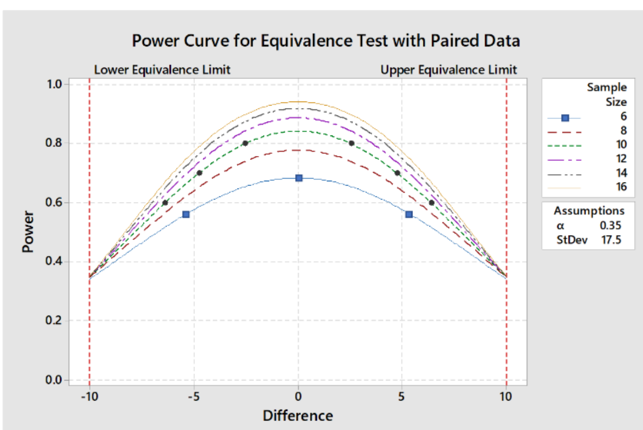

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 5

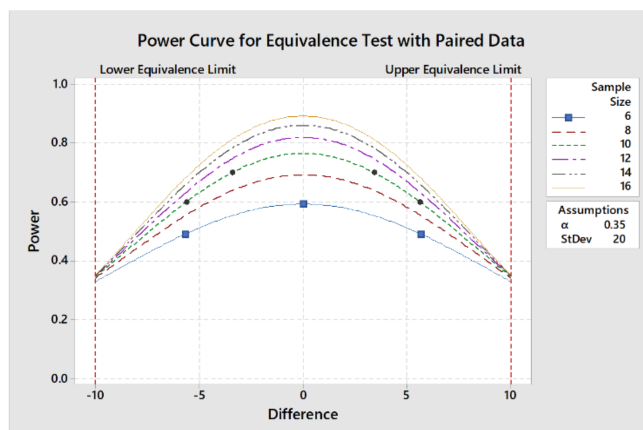

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 7.5

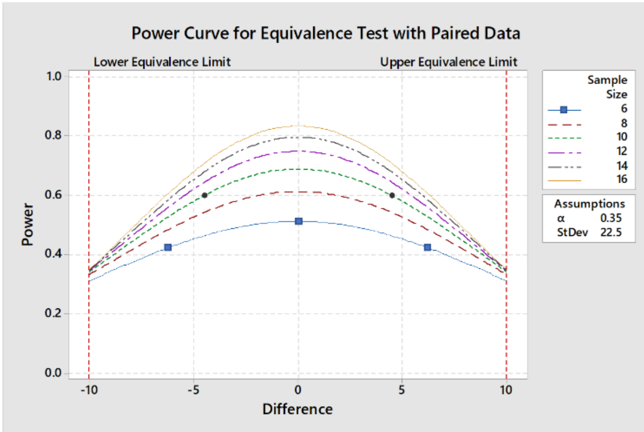

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 10

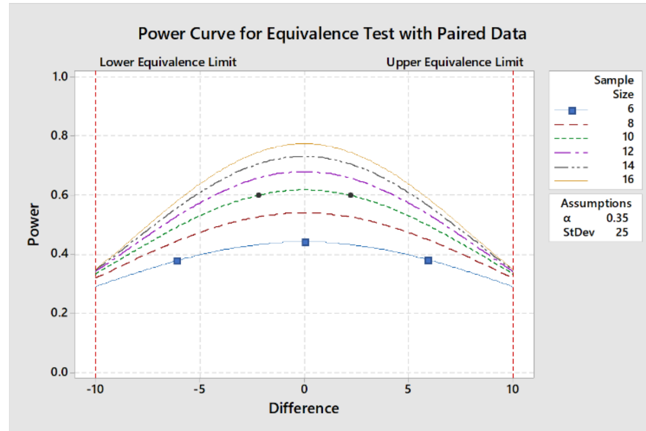

Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 12.5


Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 15


Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 17.5


Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
1.0
0.8
0.6
0.4
0.2
0.0
-10
-5
0
5
10
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 20


Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 22.5


Power Curve for Equivalence Test with Paired Data
Lower Equivalence Limit
Upper Equivalence Limit
Power
Difference
Sample Size
6
8
10
12
14
16
Assumptions
α 0.35
StDev 25

APPENDIX B – EXAMPLE OF PAIRED T-TESTING FOR EQUIVALENCE ON REFERENCE SITES

A fictional agency is conducting cracking equivalence verifications for two vendors. The agency's procedures and decisions are described in the following steps.

Step 1—Identification of Cracking Distress Types to be Verified

For this example, the agency is conducting the cracking equivalence verification at the reference sites based upon HPMS Cracking Percent for asphalt concrete and jointed plain concrete pavements.

Step 2—Reference Site Selection and Weighting

Prior year condition and inventory data from the agency's roadway management system and HPMS reporting is examined with regard to pavement types, surface textures, and frequently-occurring distress conditions. The agency's treatment decision process is also considered to note distresses of high concern and impact. In this case, the emphasis is on the HPMS Cracking Percent, but the number of sections or combination of distresses on the reference sites can be configured to also consider the agency's internal pavement asset management processes. Based upon these considerations, the fictional agency decides to use six asphalt concrete and three portland cement concrete reference sites, with the prior HPMS Cracking Percent qualitative ratings shown in Table 4. When selecting the specific sites, they also select sites with differing crack widths and patterns; including some approximately 1mm cracks.

Table 4. Reference Site Selection and Weighting for Fictional Agency

Reference Site ID	Pavement Type	HPMS Cracking Percent	Site Weight
A	AC	Good	2
B	AC	Fair	2
C	AC	Fair	1
D	AC	Fair	1
E	AC	Poor	1
F	AC	Poor	1
G	JPC	Good	1
H	JPC	Fair	2
I	JPC	Poor	1

The agency also weighs the relative importance of each of these sites, considering the distribution of pavement types and conditions in their network, as well as the importance of the cracking patterns to their own pavement asset management decision-making process. The reference sites and weighting factors selected by the fictional agency are shown in Table 4.

The minimum recommended subsection length for asphalt concrete sites is 0.03 miles to avoid increased variability due to small-value errors. For portland cement concrete sites, since the HPMS Cracking Percent is the percentage of slabs, enough slabs are needed to make the comparisons meaningful. Therefore, for JPC, the minimum recommended subsection length is

0.05 miles (or 10 slabs). The agency must determine how many subsections to use on each reference site. The greater the number of subsections, the greater the power of the statistical equivalence test. For this initial effort, the agency decides to use the recommended minimum of 10 subsections per control site.

Step 3—Ground Reference Ratings

The fictional agency performs ground cracking reference ratings under closed traffic lanes, following the previously described procedures. After the reference ratings are complete, the actual conditions are compared to those planned to be sure the selected reference sites will meet the intended condition distribution and weights. The ground reference ratings for two of the asphalt concrete sites are shown in Table 5, as illustrative examples of the necessary ratings and statistical calculations. Similarly, ratings for two of the portland cement concrete sites are shown in Table 6. Only HPMS Cracking Percent ratings are used for equivalence testing in this example.

Table 5. Example AC Ground Reference Ratings and Vendor Ratings

Reference Site ID	Subsection Length (mi)	Reference: HPMS Cracking Percent	Vendor P: HPMS Cracking Percent	Vendor Q: HPMS Cracking Percent
B	0.03	0.0	0.0	0.0
B	0.03	0.5	0.8	0.1
B	0.03	6.6	7.6	0.5
B	0.03	13.0	16.7	7.5
B	0.03	5.6	9.3	15.6
B	0.03	2.6	5.5	8.8
B	0.03	5.7	15.5	4.75
B	0.03	12.9	16.7	11.5
B	0.03	4.5	6.7	15.4
B	0.03	12.8	8.3	6.0
Mean/SD		**6.4/5.0**	**8.7/6.0**	**7.0/5.9**
E	0.03	21.0	14.5	20.0
E	0.03	13.7	19.7	6.2
E	0.03	24.2	16.0	17.0
E	0.03	23.9	23.5	16.9
E	0.03	26.3	18.4	18.9
E	0.03	11.3	14.9	8.6
E	0.03	21.5	22.9	13.7
E	0.03	22.3	22.5	21.5
E	0.03	24.9	33.2	17.4
E	0.03	33.3	32.4	27.5
Mean/SD		**22.2/6.2**	**21.8/6.6**	**16.7/6.1**

Table 6. Example JCP Ground Reference Ratings and Vendor Ratings

Reference Site ID	Subsection Length (mi)	Reference: HPMS Cracking Percent	Vendor P: HPMS Cracking Percent	Vendor Q: HPMS Cracking Percent
G	0.05	0.0	0.0	0.0
G	0.05	0.0	0.0	6.7
G	0.05	6.7	6.7	13.4
G	0.05	0.0	0.0	6.7
G	0.05	6.7	0.0	6.7
G	0.05	6.7	0.0	6.7
G	0.05	0.0	0.0	0.0
G	0.05	6.7	0.0	6.7
G	0.05	0.0	0.0	6.7
G	0.05	0.0	0.0	6.7
Mean/SD		**2.7/3.5**	**0.7/2.1**	**6.0/3.8**
I	0.05	20.0	13.3	13.3
I	0.05	26.7	26.7	20.0
I	0.05	13.3	13.3	13.3
I	0.05	40.0	40.0	33.3
I	0.05	20.0	13.3	13.3
I	0.05	26.7	13.3	26.7
I	0.05	33.3	33.3	26.7
I	0.05	33.3	26.7	33.3
I	0.05	13.3	13.3	6.7
I	0.05	26.7	33.3	20.0
Mean/SD		**25.3/8.7**	**22.7/10.5**	**20.7/9.1**

Step 4—Vendor Ratings

The two vendors, Vendor P and Vendor Q, collect images over all reference sites. The vendors submit cracking distress summaries for the same subsections as marked and rated by the ground reference ratings (0.03 mile subsections for AC sites and 0.05 mile subsections for JCP sites). The vendor ratings are included in Table 5 and Table 6.

Step 5—Statistical Equivalence Testing

Reference Site B

For asphalt concrete reference site B, the overall ground reference rating of HPMS Cracking Percent is 6% (6.4%). Therefore, the equivalence limits of ± 4% apply.

The standard deviation that is used in paired t-testing is the standard deviation of the differences between the pairs. However, in the absence of other information, the standard deviation of the reference rating was approximately 5%, so that was used while setting up the tests as an initial rough estimate of the variability expected. After the vendor ratings were submitted, the standard deviation of the differences between the reference and Vendor P is found to be 3.5% and the standard deviation of the differences between the reference and Vendor Q is 6.4%.

Examining the power curves with limits of ± 4%, N=10, and standard deviation of 5, it is estimated that in order to have a power of the equivalence test of approximately 0.8, an alpha value of 0.10 to 0.15 would be needed. (The two relevant power curves bracketing these values are shown in Figures 18 and 19.) This can also be seen in Figure 20. This means that the fictional agency would be accepting a risk of between 10% and 15% of accepting a rating as equivalent when it is not (again, the interpretation of the alpha-value is different for equivalence testing than for hypothesis testing looking to verify statistically significant differences). The agency finds this balance between power and alpha to be acceptable and proceeds with the equivalence testing.

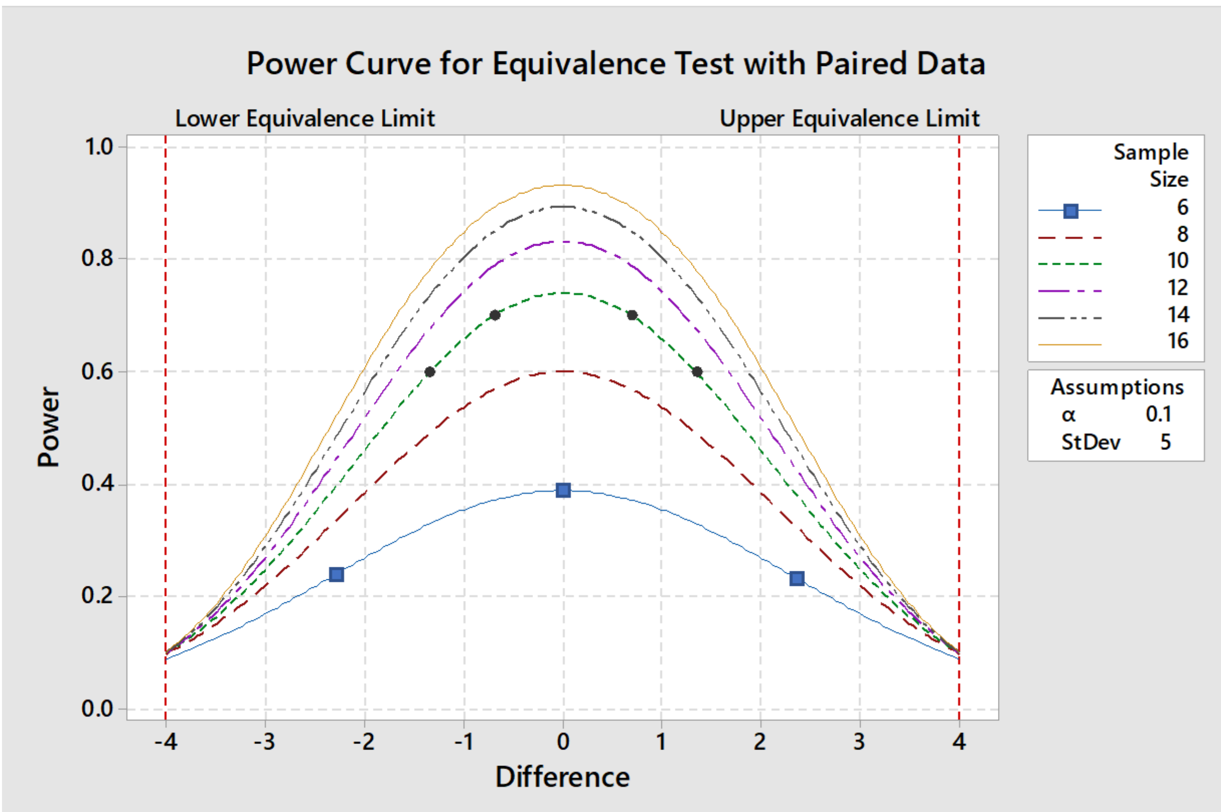


Figure 18. Power curve for equivalence test with paired data, limits ± 4%, standard deviation 5% and alpha = 0.10.

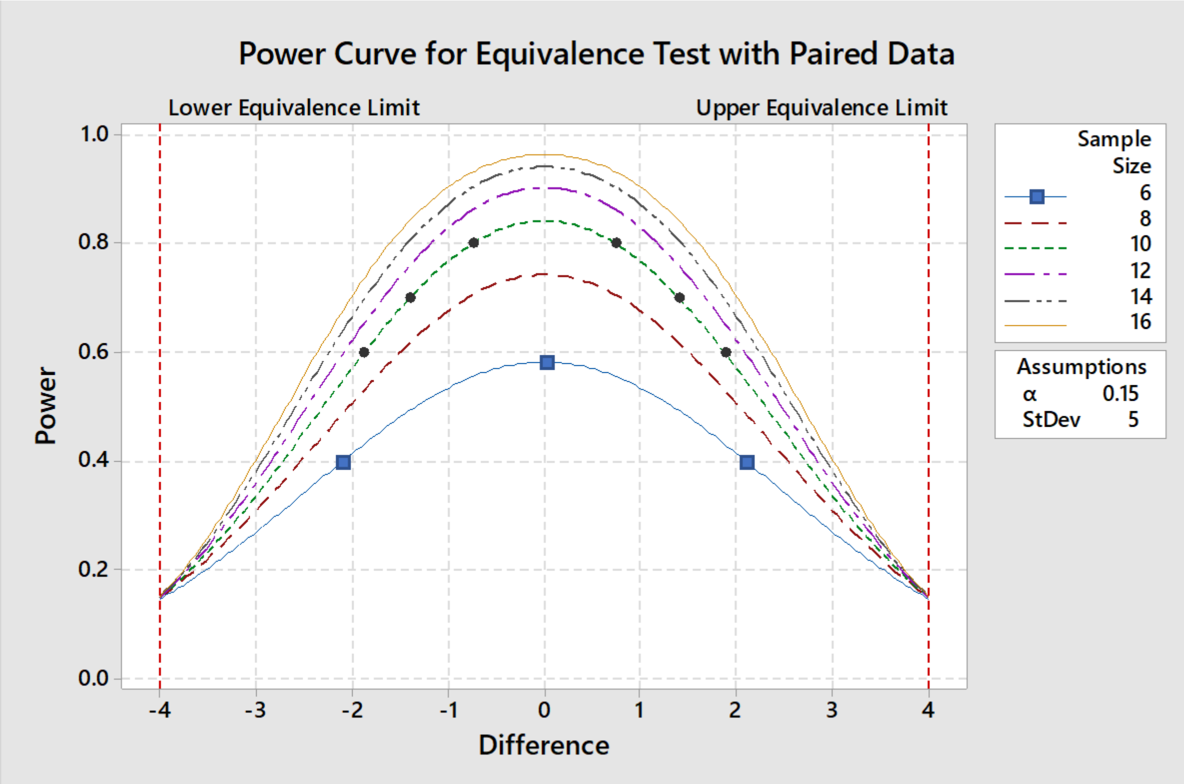


Figure 19. Power curve for equivalence test with paired data, limits ± 4%, standard deviation 5% and alpha = 0.15.

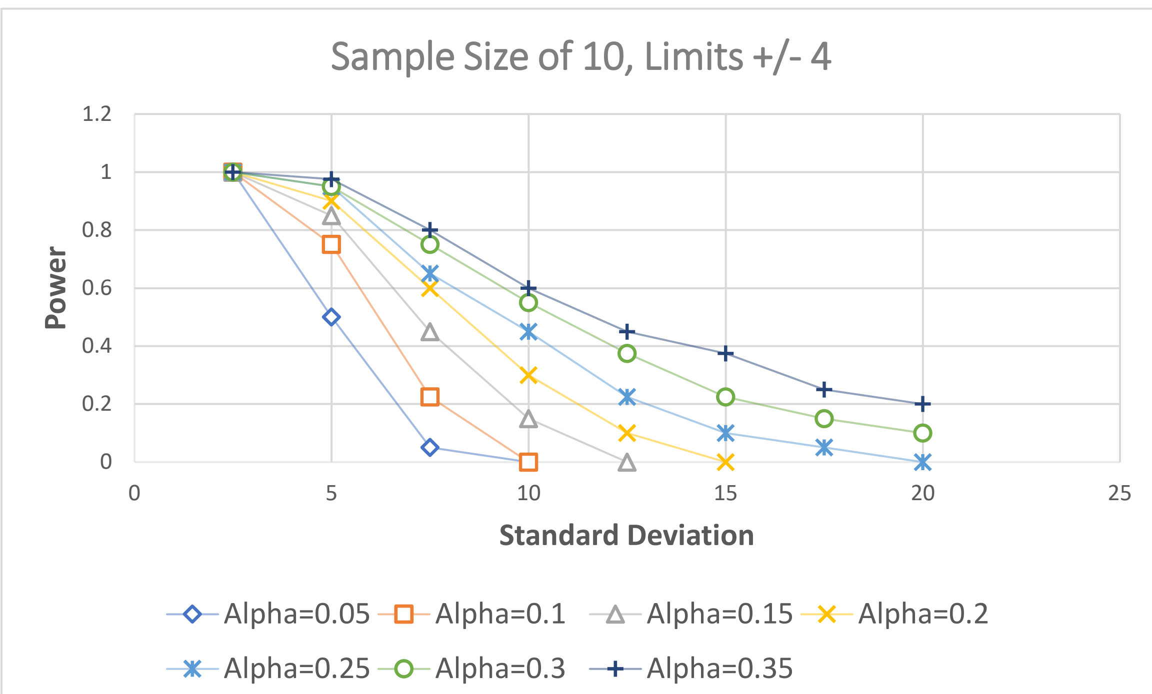


Figure 20. Relationship between standard deviation of differences, power, and alpha for 10 subsections when acceptance limits are ± 4% HPMS cracking.

Since the agency and vendors are rating exactly the same 0.03-mile subsections, paired t-testing formulated for equivalence can be used. For Vendor P, the results are shown in Figure 21. For Vendor Q, the results are shown in Figure 22. For Reference Site B, the ratings of both vendors are found to be equivalent to the reference rating.

Reference Site B: Equivalence Test with Paired Data: Vendor P, Reference Method

Test mean = mean of Vendor P
Reference mean = mean of Reference

Descriptive Statistics

Variable	N	Mean	StDev	SE Mean
Vendor P	10	8.7100	6.0398	1.9099
Reference	10	6.4200	4.9638	1.5697

Difference: Mean(Vendor P)-Mean(Reference) Diff=2.29 StDev=3.65

Null hypothesis: Difference≤-4 or Difference ≥4
Alternative hypothesis: -4 < Difference < 4
α level: 0.1

Null Hypothesis	DF	T-Value	P-Value
Difference ≤ -4	9	5.4471	0.000
Difference ≥ 4	9	-1.4808	0.086

The greater of the two P-Values is 0.086. ***Can claim equivalence***

Figure 21. Reference Site B equivalence testing for Vendor P.

Reference Site B: Equivalence Test with Paired Data: Vendor Q, Reference Method

Test mean = mean of Vendor Q
Reference mean = mean of Reference

Descriptive Statistics

Variable	N	Mean	StDev	SE Mean
Vendor P	10	7.0150	5.8990	1.8654
Reference	10	6.4200	4.9638	1.5697

Difference: Mean(Vendor Q)-Mean(Reference) Diff=.595 StDev=6.41

Null hypothesis: Difference≤-4 or Difference ≥4
Alternative hypothesis: -4 < Difference < 4
α level: 0.1

Null Hypothesis	DF	T-Value	P-Value
Difference ≤ -4	9	2.2649	0.025
Difference ≥ 4	9	-1.6784	0.064

The greater of the two P-Values is 0.064. ***Can claim equivalence***

Figure 22. Reference Site B equivalence testing for Vendor Q.

Reference Site E

For asphalt concrete reference site E, the overall ground reference rating of HPMS Cracking Percent is 22.2% (22%). Therefore, the equivalence limits of ± 4% again apply.

The standard deviation of the reference rating was approximately 6%, so that was used as an initial estimation of the variability. After the vendor ratings were submitted, the standard deviation of the differences between the reference and Vendor P is found to be 5.7% and the standard deviation of the differences between the reference and Vendor Q is 2.8%.

Examining the power curves (Figures 18, 19, and 20) with limits of ± 4%, N=10, and standard deviation of approximately 5%, it is estimated that in order to have a power of the equivalence test of approximately 80%, an alpha value of 0.10 would again be needed. For Vendor P, the results of the paired t-testing for equivalence are shown in Figure 23. For Vendor Q, the results are shown in Figure 24. Equivalence to the ground reference ratings can be claimed for Vendor P for reference site E, but not for Vendor Q.

Equivalence Test with Paired Data: Vendor P, Reference Method

Test mean = mean of Vendor P
Reference mean = mean of Reference

Descriptive Statistics

Variable	N	Mean	StDev	SE Mean
Vendor P	10	21.800	6.6451	2.1014
Reference	10	22.240	6.2110	1.9641

Difference: Mean(Vendor P)-Mean(Reference) Diff=-0.44 StDev=5.68

Null hypothesis: Difference≤-4 or Difference ≥4
Alternative hypothesis: -4 < Difference < 4
α level: 0.1

Null Hypothesis	DF	T-Value	P-Value
Difference ≤ -4	9	1.9806	0.039
Difference ≥ 4	9	-2.4701	0.018

The greater of the two P-Values is 0.039. ***Can claim equivalence***

Figure 23. Reference Site E Equivalence Testing for Vendor P.

Equivalence Test with Paired Data: Vendor Q, Reference Method

Test mean = mean of Vendor Q
Reference mean = mean of Reference

Descriptive Statistics

Variable	N	Mean	StDev	SE Mean
Vendor Q	10	16.770	6.1540	1.9461
Reference	10	22.240	6.2110	1.9641

Difference: Mean(Vendor Q)-Mean(Reference) Diff=-5.47 StDev=2.83

Null hypothesis: Difference≤-4 or Difference ≥4
Alternative hypothesis: -4 < Difference < 4
α level: 0.1

Null Hypothesis	DF	T-Value	P-Value
Difference ≤ -4	9	-1.6401	0.932
Difference ≥ 4	9	-10.566	0.000

The greater of the two P-Values is 0.932. ***Cannot claim equivalence***

Reference Site G

For concrete (JCP) reference site G, the overall ground reference rating of HPMS Cracking Percent is 2.6% (3%). Therefore, the equivalence limits of ± 4% apply. The same power curves used for the asphalt concrete ± 4% limits can be used (Figures 18, 19, and 20).

The standard deviation of the reference rating was approximately 3.5%, so that was used as an initial estimation of the variability. After the vendor ratings were submitted, the standard deviation of the differences between the reference and Vendor P is found to be 3.2% and the standard deviation of the differences between the reference and Vendor Q is 3.5%.

Again, examining the power curves with limits of ± 4, N=10, and standard deviation of approximately 5%, it is estimated that in order to have a power of the equivalence test of approximately 80%, an alpha value of 0.10 would again be needed. The actual standard deviations of the differences were slightly lower, so the true power of the test would be higher. For Vendor P, the results of the paired t-testing for equivalence are shown in Figure 25. For Vendor Q, the results are shown in Figure 26. Equivalence to the ground reference ratings can be claimed for Vendor P for reference site G, but not for Vendor Q.

Equivalence Test with Paired Data: Vendor P, Reference Method
Test mean = mean of Vendor P
Reference mean = mean of Reference

Descriptive Statistics

Variable	N	Mean	StDev	SE Mean
Vendor P	10	0.67000	2.1187	0.67000
Reference	10	2.6800	3.4599	1.0941

Difference: Mean(Vendor P)-Mean(Reference) Diff=-2.01 StDev=3.24

Null hypothesis: Difference≤-4 or Difference ≥4
Alternative hypothesis: -4 < Difference < 4
α level: 0.1

Null Hypothesis	DF	T-Value	P-Value
Difference ≤ -4	9	1.9440	0.042
Difference ≥ 4	9	-5.8723	0.000

The greater of the two P-Values is 0.042. ***Can claim equivalence***

Figure 25. Reference Site G Equivalence Testing for Vendor P.

Equivalence Test with Paired Data: Vendor Q, Reference Method
Test mean = mean of Vendor Q
Reference mean = mean of Reference

Descriptive Statistics

Variable	N	Mean	StDev	SE Mean
Vendor Q	10	6.0300	3.8032	1.2027
Reference	10	2.6800	3.4599	1.0941

Difference: Mean(Vendor Q)-Mean(Reference) Diff= 3.35 StDev= 3.53

Null hypothesis: Difference≤-4 or Difference ≥4
Alternative hypothesis: -4 < Difference < 4
α level: 0.1

Null Hypothesis	DF	T-Value	P-Value
Difference ≤ -4	9	6.5821	0.000
Difference ≥ 4	9	-0.58209	0.287

The greater of the two P-Values is 0.287. ***Cannot claim equivalence***

Figure 26. Reference Site G Equivalence Testing for Vendor Q.

Reference Site I

For concrete (JCP) reference site I, the overall ground reference rating of HPMS Cracking Percent is 25.3% (25%). Therefore, the equivalence limits of ± 7.5% apply.

The standard deviation of the reference rating was approximately 9%, so that was used as an initial estimation of the variability. That would have required the agency to accept an alpha value of approximately 0.15 for a power of 0.80, and N=10 as shown in Figure 27. After the vendor ratings were submitted, however, the standard deviation of the differences between the reference and Vendor P is found to be 5.6% and the standard deviation of the differences between the reference and vendor Q is 3.2%.

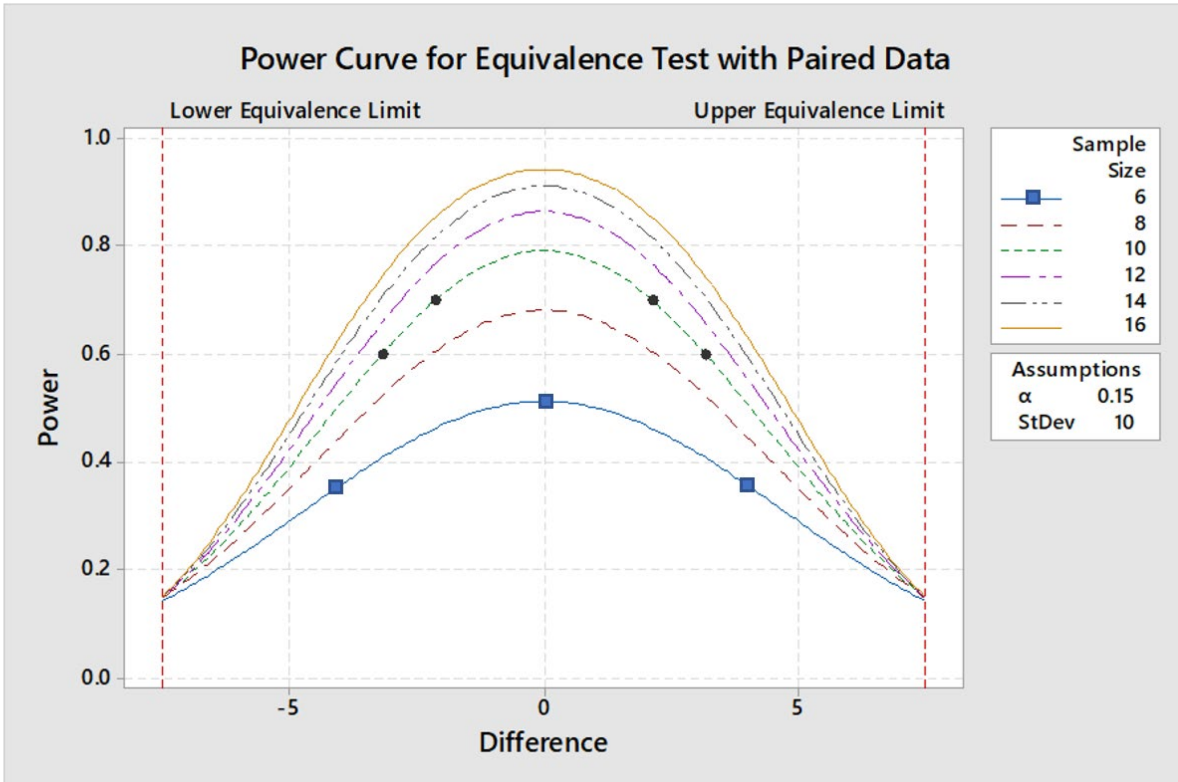


Figure 27. Power curve for equivalence test with paired data, limits ± 7.5%, standard deviation 10% and alpha = 0.15.

Examining the power curves with limits of ± 7.5, N=10, and standard deviation of approximately 5%, it is estimated that in order to have a power of the equivalence test of approximately 80%, an alpha value of 0.05 would suffice. The relevant power curve and summary figure are shown in Figures 28 and 29. The results of the paired t-testing for equivalence for reference site I are shown in Figure 30 and Figure 31, for Vendor P and Vendor Q, respectively. Both were found to be equivalent to the reference rating.

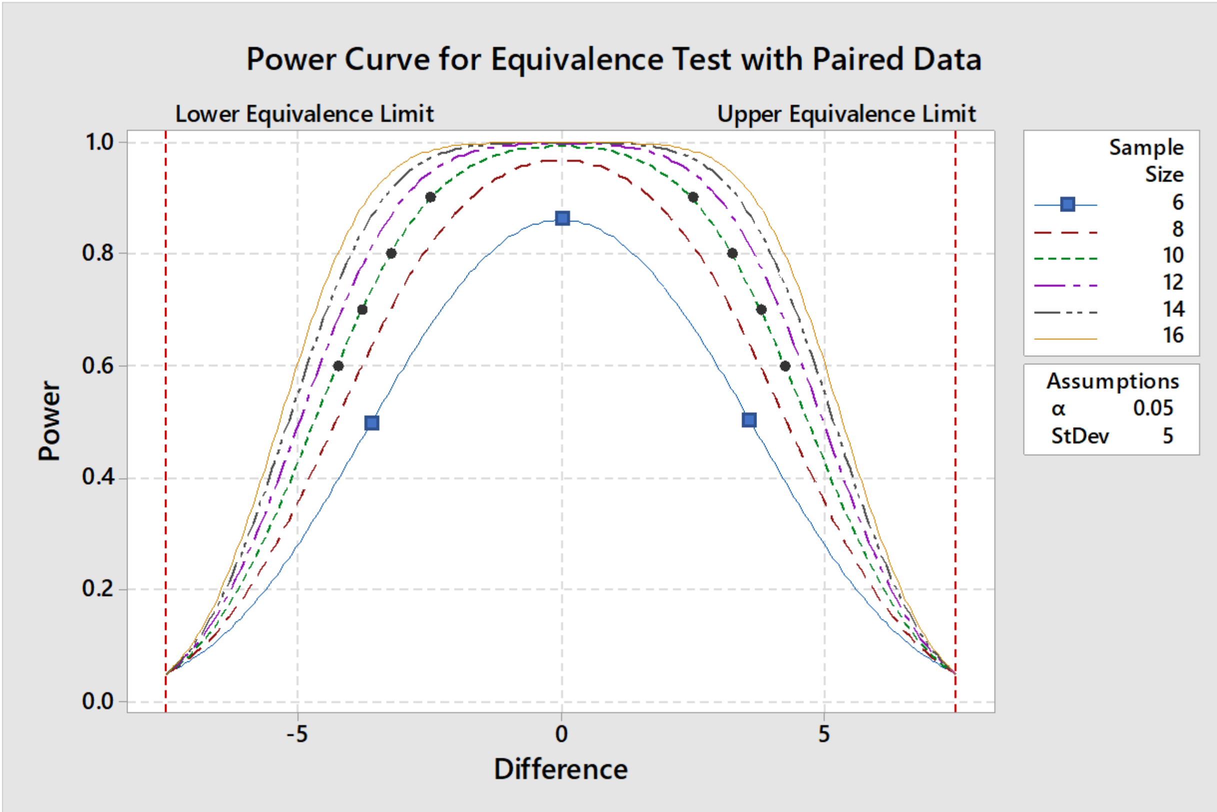


Figure 28. Power curve for equivalence test with paired data, limits ± 7.5%, standard deviation 5% and alpha = 0.05.

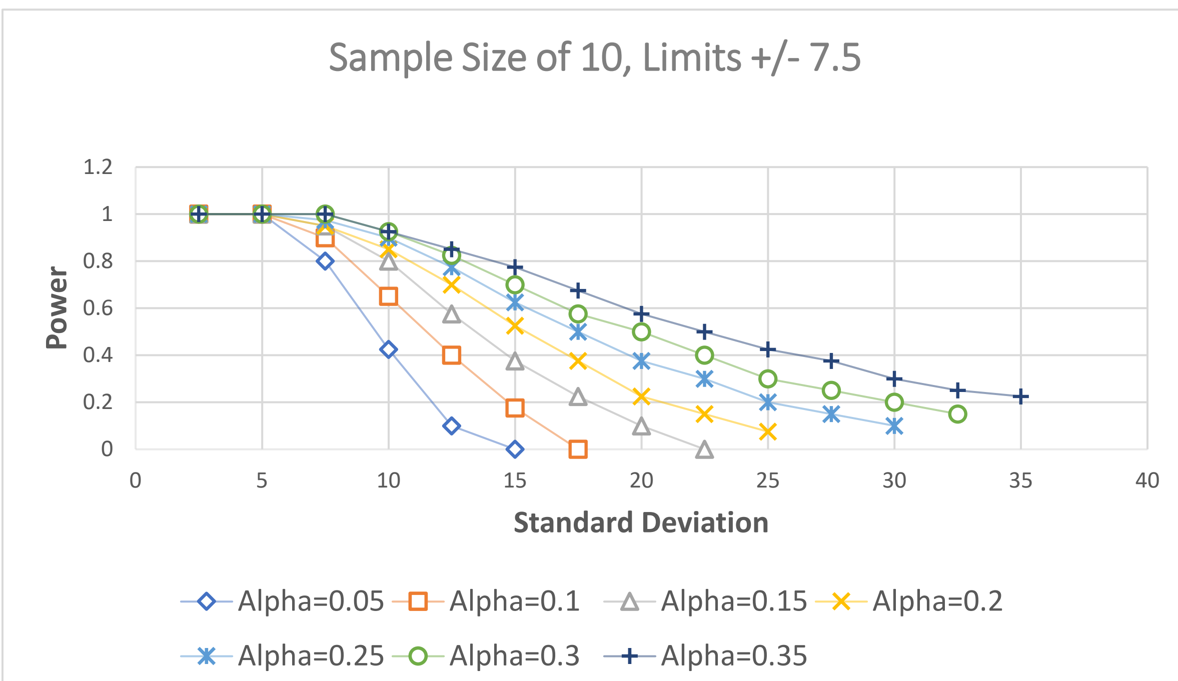


Figure 29. Relationship between standard deviation of differences, number of subsections and alpha to achieve a power of 0.8 when acceptance limits are ± 7.5% HPMS cracking.

Equivalence Test with Paired Data: Vendor P, Reference Method

Test mean = mean of Vendor P
Reference mean = mean of Reference

Descriptive Statistics

Variable	N	Mean	StDev	SE Mean
Vendor P	10	22.667	10.517	3.3259
Reference	10	25.333	8.7771	2.7756

Difference: Mean(Vendor P)-Mean(Reference) Diff=-2.67 StDev= 5.62

Null hypothesis: Difference≤-7.5 or Difference ≥7.5
Alternative hypothesis: -7.5 < Difference < 7.5
α level: 0.1

Null Hypothesis	DF	T-Value	P-Value
Difference ≤ -7.5	9	2.7188	0.012
Difference ≥ 7.5	9	-5.7188	0.000

The greater of the two P-Values is 0.012. ***Can claim equivalence***

Equivalence Test with Paired Data: Vendor Q, Reference Method

Test mean = mean of Vendor Q
Reference mean = mean of Reference

Descriptive Statistics

Variable	N	Mean	StDev	SE Mean
Vendor Q	10	20.667	9.1355	2.8889
Reference	10	25.333	8.7771	2.7756

Difference: Mean(Vendor Q)-Mean(Reference) Diff=-4.67 StDev= 3.22

Null hypothesis: Difference≤-7.5 or Difference ≥7.5
Alternative hypothesis: -7.5 < Difference < 7.5
α level: 0.1

Null Hypothesis	DF	T-Value	P-Value
Difference ≤ -7.5	9	2.7823	0.011
Difference ≥ 7.5	9	-11.947	0.000

The greater of the two P-Values is 0.011. ***Can claim equivalence***

Figure 31. Reference Site I Equivalence Testing for Vendor Q.

Similarly, the fictional agency conducts the paired t-tests for equivalence for the remaining five reference sites. The results are shown in Table 7.

Table 7. Reference Site Selection and Weighting for Fictional Agency

Reference Site ID	Pavement Type	HPMS Cracking Percent	Site Weight	Vendor P Equivalence Results	Vendor P Weighted Score	Vendor Q Equivalence Results	Vendor Q Weighted Score
A	AC	Good	2	EQ	2	Not EQ	0
B	AC	Fair	2	EQ	2	EQ	2
C	AC	Fair	1	EQ	1	EQ	1
D	AC	Fair	1	Not EQ	0	EQ	1
E	AC	Poor	1	EQ	1	Not EQ	0
F	AC	Poor	1	Not EQ	0	EQ	1
G	JPC	Good	1	EQ	1	Not EQ	0
H	JPC	Fair	2	EQ	2	EQ	2
I	JPC	Poor	1	EQ	1	EQ	1
TOTAL SCORE					**10**		**8**

Step 6—Interpreting the Results

As shown in Table 7, considering the site weights, a total score is determined for each vendor. In this example, Vendor P has a total weighted score equal to 10 compared to Vendor Q's weighted score of 8. Therefore, Vendor P is found to be statistically more likely to produce equivalent ratings for the fictional agency's network.

APPENDIX C – GUIDELINES FOR VENDOR SELECTION

For illustrative purposes, the process is demonstrated using the FHWA HPMS [78] cracking reporting standard. An agency may choose to use this standard or to use the illustration to apply to their own cracking definitions and needs.

Equipment Verification Requirements

The vendor shall provide an imaging system capable of collecting clear, high resolution images from which the cracking can be identified to a specified level. The system shall also be capable of quantifying the cracking distresses to the FHWA HPMS reporting standard and/or to meet other reporting standards suitable to the agency's needs.

In the procurement documents, an agency should request the vendor provide the following information, or alternatively require the vendor to meet or exceed the following criteria:

1. Imaging System Clarity
   Vendor shall provide agency information regarding the number of pixels per image and the associated fill factor. Currently recommended limits for 3D imaging are 16-bit images with a minimum of 4,096 pixels and a minimum fill factor of 90%.

2. Image Resolution
   Imaging system must be capable of producing images in which a 1mm crack is visible when confirmed at slow speed and/or a 3mm crack at speeds exceeding 60 mph.

3. Image capture width
   The width of image shall cover a minimum of 13.5 ft. wide.

4. System capabilities
   a. Illumination Source – The Downward Perspective Image shall be collected with a uniform and consistent form of illumination applied to the pavement surface. The illumination shall be regulated to provide sufficient contrast and crack-shadows for the clear discernment of cracking and patching. Images bearing ambient and/or vehicle shadows that obscure pavement features will not be accepted.

   b. Data Collection Speed – Data should be collected at or near prevailing highway speeds, typically 25 to 65 miles per hour.

   c. Required System Storage – The data storage file shall be self-describing and self-contained. Self-description here means that the data can be interpreted by different systems at different points in time. Self-contained means that all data needed for interpretation is included within the data "container." For example, metadata will not only provide the identifying information such as location and date, but also a string of information linking the header to all the accessible data stored for that specific location. FHWA is in the final publication phase of a report titled "Evaluation of Proposed Standard Data Format and Compression Algorithms for 2D/3D Pavement Surface Image" [79] which proposes a standard data format for the 2D/3D image systems. An agency may consider requiring the 2D and 3D images provided by the vendor comply with the proposed standard data format.

5. Define a Quality Image
   Image quality should provide the following characteristics: pavement surface is visible without shadows, reflection from wet surface, or other conditions of the imaging process resulting in images which cannot be clearly viewed.

6. Validation of Image Quality
   Image quality will be assessed in section 7 below. Ongoing NCHRP Project 1-60, Measuring the Characteristics of Pavement Surface Images and Developing Standard Practices for Calibration, Certification, and Verification of Imaging Systems has the following objectives: (1) identify and develop methods for measuring the characteristics of surface images used for pavement evaluation and analysis and (2) develop recommended standard practices for the calibration, certification, and verification of such images, for consideration and adoption by AASHTO. It is expected the results from this project will provide objective image quality validation protocols.

7. Cracking Distress Interpretation Type (An agency needs to choose which of the three cracking distress interpretations they desire to specify.)

   Options:
   - Manually rating (reviewing and classifying the cracks by viewing the digital images)
   - Semi-automated rating (rating is performed by computer processing of raw distress data, algorithms or other, and verified by manual reviews of the processed distress)
   - Fully automated rating (distress classifications are produced by computer processing using algorithms or other methods, with little to no human intervention)

Cracking Distress Data Verification Steps

The vendor should be requested to collect and report cracking data on the reference sites identified in Table 8. For illustration purposes, the HPMS Percent Cracking section length, subsection length, and number of subsections are identified.

Table 8. List of Reference Sections and Corresponding Details.

Reference Site Number	Location Description	Pavement Type	Section Length	Subsection Length	Number of Subsections	Weighting Factor
ACP-1	TBD	ACP	0.3 mile	0.03 mile	10	TBD
ACP-2	TBD	ACP	0.3 mile	0.03 mile	10	TBD
ACP-3	TBD	ACP	0.3 mile	0.03 mile	10	TBD
ACP-…	TBD	ACP	0.3 mile	0.03 mile	10	TBD
JCP-1	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
JCP-2	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
JCP-3	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
JCP-…	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
CRCP-1	TBD	CRCP	0.3 mile	0.03 mile	10	TBD
CRCP-2	TBD	CRCP	0.3 mile	0.03 mile	10	TBD
CRCP-3	TBD	CRCP	0.3 mile	0.03 mile	10	TBD
CRCP-…	TBD	CRCP	0.3 mile	0.03 mile	10	TBD

The agency has identified (number) reference sites that shall be collected and reported. Each reference site has been given a weighting factor that will be used for analysis purposes. Sections are selected to represent typical pavement types and conditions throughout the agency. The weighting criteria are based upon the distribution of pavement conditions throughout the agency and upon the criticality of identified distresses to the pavement management system.

The ground Reference Rating will be conducted by the agency. All cracking equal to or greater than 1mm in width shall be reported. The sections and subsections will be clearly marked with Start, End, and Intermediate Points by the agency.

The vendor shall collect and report HPMS Cracking Percent data, to include the detailed measures as well as the HPMS Cracking Percentage on all designated reference sections. Data shall be reported at the specified subsection lengths. An agency may choose to require cracking data other than HPMS Cracking Percent. This alternative cracking data must be defined in the solicitation.

The HPMS Percent Cracking for asphalt concrete pavement (ACP), jointed concrete pavement (JCP), and continuously reinforced concrete pavement (CRCP) are defined as:

- For ACP, Cracking Percent is the percentage of the total area exhibiting visible fatigue type cracking (any cracking in the wheel path) for all severity levels in the wheel path in each section.
    - Cracking shall be measured and reported for both wheel paths. Measuring and reporting cracking outside of the wheel path areas is not required for HPMS reporting. Report the lineal length of cracking for each wheel path.
    - Any and all severity levels (sealed and unsealed) shall be reported.

- Multiply the lineal length of wheel path with cracking by 39 in. to get the total area of wheel path cracking.
- Cracking Percent is reported as the area of wheel path cracking divided by the total area of the section, multiplied by 100. To limit the potential variation in reported lane width, for the purposes of this cracking comparison the lane width is set at 12 ft. wide.
- Report Cracking Percent to the nearest one percent on each subsection.

- For JCP, Cracking Percent is the percentage of slabs within the section that exhibit transverse cracking. Partial slabs shall contribute to the section that contains the majority of the slab length.
  - Reported cracking for jointed concrete pavements excludes longitudinal cracks, corner breaks, D-cracking, and ASR cracking that may occur on a slab.
  - Report the actual total number of slabs in each subsection as well as the total number of slabs for the entire reference section.
  - Report the total number of slabs containing one or more transverse cracks extending for at least one-half the lane width. This is the number of cracked slabs.
  - The percentage of cracking reported is calculated as the number of cracked slabs, divided by the total number of slabs in the section, multiplied by 100.
  - Report Cracking Percent to the nearest one percent on each subsection.

- For CRCP, the Cracking Percent is the percentage of the area of the section exhibiting longitudinal cracking, punchouts, and/or patching. Transverse cracking shall not be considered in the Cracking Percent for CRCP.
  - Cracking and distresses may occur anywhere on the pavement. Transverse cracks that are at or near right angles to the direction of travel in the lane should not be included in the calculation.
  - Distresses to be included are longitudinal cracking (any severity), punchouts, and patched areas.
  - For longitudinal cracking, the cracked area is determined as the length of the crack multiplied by 1 ft. width.
  - For punchouts, the area is determined by the two transverse cracks and the edge of the pavement or longitudinal joint.
  - Report the area of all patches.
  - Percentage of Cracking for CRCP pavements is determined as the area of pavement where longitudinal cracking, punchouts, and patches are detected divided by the total area of the section, multiplied by 100. To limit the potential variation in reported lane width, for the purposes of this cracking comparison the lane width is set at 12 ft. wide.
  - Report Cracking Percent to the nearest one percent on each subsection.

HPMS Cracking Percent data shall be delivered to the agency in a data table following the formats illustrated in Tables 9, 10, and 11.

Table 9. Data Delivery Table for ACP.

Reference Site Number	Subsection Number	Subsection Station (mile)	Length of Wheel Path Cracking	HPMS Cracking Percent
ACP-1	1	0 to 0.03		
ACP-1	2	0.03 to 0.06		
ACP-1	…	…		
ACP-1	10	0.27 to 0.30		
ACP-2	1	0 to 0.03		
…	…	…		
ACP-3	10	0.27 to 0.30		
ACP-…				

Table 10. Data Delivery Table for JCP.

Reference Site Number	Subsection Number	Subsection Station (mile)	Total Number of Slabs	Number of Cracked Slabs	HPMS Cracking Percent
JCP-1	1	0 to 0.05			
JCP-1	2	0.05 to 0.10			
JCP-1	…	…			
JCP-1	10	0.45 to 0.50			
JCP-2	1	0 to 0.05			
…	…	…			
JCP-3	10	0.45 to 0.50			
JCP-…					

Table 11. Data Delivery Table for CRCP.

Reference Site Number	Subsection Number	Subsection Station (mile)	Length of Longitudinal Cracking	Area of Punchout	Area of Patching	Total Distress Area	HPMS Cracking Percent
CRCP-1	1	0 to 0.3					
CRCP-1	2	0.03 to 0.06					
CRCP-1	…	…					
CRCP-1	10	0.27 to 0.30					
CRCP-2	1	0 to 0.03					
…	…	…					
CRCP-3	10	0.27 to 0.30					
CRCP-…	…	…					

The agency will perform the statistical equivalence evaluation using the paired t-test for equivalence method. Each reference site will be evaluated on an independent basis. The pass/fail criteria are then applied to the section weighting factor and an overall vendor score is achieved.

Alternative Cracking Distress Data Verification Option – Proof of Verification

An agency may elect to allow a vendor to provide results of verification tests from another agency as a means of proving both their equipment capabilities and their ability to produce acceptable cracking distresses from the images. The vendor should be requested to submit documentation, similar to the process described above, from a reputable agency that is in a similar climatic region and has similar pavement types.

The vendor shall provide an imaging system capable of collecting clear, high resolution images from which the cracking can be identified to a specified level. The system shall also be capable of quantifying the cracking distresses to the FHWA HPMS reporting standard and/or to meet the agency's needs. An agency may choose to require cracking data other than HPMS Cracking Percent. This alternative cracking data must be defined in the solicitation.

In the procurement documents, an agency should request the vendor provide the following information, or alternatively require the vendor to meet or exceed the following criteria:

1. Imaging System Clarity
   Vendor should provide agency information regarding the number of pixels per image and the associated fill factor. Current recommended limits for 3D imaging are 16-bit images with a minimum of 4,096 pixels and a minimum fill factor of 90%.

2. Image Resolution
   Imaging system must be capable of producing images in which a 1mm crack is visible when confirmed at slow speed and/or a 3mm crack at speeds exceeding 60 mph.

3. Image capture width
   The width of image shall cover a minimum of 13.5 ft. wide.

4. System capabilities
   a. Illumination source – The Downward Perspective Image shall be collected with a uniform and consistent form of illumination applied to the pavement surface. The illumination shall be regulated to provide sufficient contrast and crack-shadows for the clear discernment of cracking and patching. Images bearing ambient and/or vehicle shadows that obscure pavement features will not be accepted.

   b. Data collection speed – Data should be collected at or near prevailing highway speeds, typically 25 to 65 miles per hour.

   c. Required system storage – The data storage file shall be self-describing and self-contained. Self-description here means that the data can be interpreted by different systems at different points in time. Self-contained means that all data needed for interpretation is included within the data "container." For example, metadata will not only provide the identifying information such as location and date, but also a string of information linking the header to all the accessible data stored for that specific location. FHWA is in the final publication phase of a report titled "Evaluation of Proposed Standard Data Format and Compression Algorithms for 2D/3D Pavement

Surface Image” [79] which proposes a standard data format for the 2D/3D image systems. An agency may consider requiring the 2D and 3D images provided by the vendor comply with the proposed standard data format.

5. Define a quality image
   Image quality should provide the following characteristics: pavement surface is visible without shadows, reflection from wet surface, or other conditions of the imaging process resulting in images which cannot be clearly viewed.

6. Validation of Image Quality
   Ongoing NCHRP project 1-60, Measuring the Characteristics of Pavement Surface Images and Developing Standard Practices for Calibration, Certification, and Verification of Imaging Systems has the following objectives: (1) identify and develop methods for measuring the characteristics of surface images used for pavement evaluation and analysis and (2) develop recommended standard practices for the calibration, certification, and verification of such images, for consideration and adoption by AASHTO. It is expected the results from this project will provide objective image quality validation protocols.

7. Cracking distress interpretation type (An agency needs to choose which of the three cracking distress interpretations they desire to specify.)

   Options:
   - Manually rating (reviewing and classifying the cracks by viewing the digital images)
   - Semi-automated rating (rating is performed by computer processing of raw distress data, algorithms or other, and verified by manual reviews of the processed distress)
   - Fully automated rating (distress classifications are produced by computer processing using algorithms or other methods, with little to no human intervention)

8. The vendor should be requested to submit documentation from a reputable agency that is in a similar climatic region and has similar pavement types as the requesting agency. The documentation must include an analysis of the vendors cracking distress evaluations as compared to reference sites for ACP, JCP and CRCP (agency to select pavement types). The documentation must include the details of the statistical equivalence evaluation, including the pass/fail criteria for each specific reference site.

9. The agency will review this documentation to determine acceptability of the vendor’s ability to successfully determine cracking distresses.

Alternative Cracking Distress Data Verification Option – Post Award Reference Sites

In an effort to reduce the pre-selection time and efforts required of the vendor, an agency may desire to select a vendor based upon that vendor’s proposed equipment and experience. In this case, all vendor equipment considered should address the criteria below. The selected vendor would then be required to pass a number of agency established reference sites prior to beginning production level data collection.

1. Imaging System Clarity
   Vendor should provide agency information regarding the number of pixels per image and the associated fill factor. Current recommended limits for 3D imaging are 16-bit images with a minimum of 4,096 pixels and a minimum fill factor of 90%.

2. Image Resolution
   Imaging system must be capable of producing images in which a 1mm crack is visible when confirmed at slow speed and/or a 3mm crack at speeds exceeding 60 mph.

3. Image capture width
   The width of image shall cover a minimum of 13.5 ft. wide.

4. System capabilities
   a. Illumination source – The Downward Perspective Image shall be collected with a uniform and consistent form of illumination applied to the pavement surface. The illumination shall be regulated to provide sufficient contrast and crack-shadows for the clear discernment of cracking and patching. Images bearing ambient and/or vehicle shadows that obscure pavement features will not be accepted.

   b. Data collection speed – Data should be collected at or near prevailing highway speeds, typically 25 to 65 miles per hour.

   c. Required system storage – The data storage file shall be self-describing and self-contained. Self-description here means that the data can be interpreted by different systems at different points in time. Self-contained means that all data needed for interpretation is included within the data "container." For example, metadata will not only provide the identifying information such as location and date, but also a string of information linking the header to all the accessible data stored for that specific location. FHWA is in the final publication phase of a report titled "Evaluation of Proposed Standard Data Format and Compression Algorithms for 2D/3D Pavement Surface Image" [(2)] which proposes a standard data format for the 2D/3D image systems. An agency may consider requiring the 2D and 3D images provided by the vendor comply with the proposed standard data format.

5. Define a quality image
   Image quality should provide the following characteristics: pavement surface is visible without shadows, reflection from wet surface, or other conditions of the imaging process resulting in images which cannot be clearly viewed.

6. Validation of Image Quality
   Image quality will be assessed in section ***Cracking Distress Data Verification Steps*** below. Ongoing NCHRP project 1-60, Measuring the Characteristics of Pavement Surface Images and Developing Standard Practices for Calibration, Certification, and Verification of Imaging Systems has the following objectives: (1) identify and develop methods for measuring the

characteristics of surface images used for pavement evaluation and analysis and (2) develop recommended standard practices for the calibration, certification, and verification of such images, for consideration and adoption by AASHTO. It is expected the results from this project will provide objective image quality validation protocols.

7. Cracking distress interpretation type (An agency needs to choose which of the three cracking distress interpretations they desire to specify.)

   Options:
   - Manually rating (reviewing and classifying the cracks by viewing the digital images)
   - Semi-automated rating (rating is performed by computer processing of raw distress data, algorithms or other, and verified by manual reviews of the processed distress)
   - Fully automated rating (distress classifications are produced by computer processing using algorithms or other methods, with little to no human intervention)

Cracking Distress Data Verification Steps

The vendor should be requested to collect and report cracking data on the reference sites identified in Table 12. For illustration purposes, the HPMS Percent Cracking section length, subsection length and number of subsections are identified.

The agency has identified (number) of reference sites that shall be collected and reported. Sections are selected to represent typical pavement types and conditions throughout the agency.

Table 12. List of Reference Sections and Corresponding Details.

Reference Site Number	Location Description	Pavement Type	Section Length	Subsection Length	Number of Subsections	Weighting Factor
ACP-1	TBD	ACP	0.3 mile	0.03 mile	10	TBD
ACP-2	TBD	ACP	0.3 mile	0.03 mile	10	TBD
ACP-3	TBD	ACP	0.3 mile	0.03 mile	10	TBD
ACP-…	TBD	ACP	0.3 mile	0.03 mile	10	TBD
JCP-1	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
JCP-2	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
JCP-3	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
JCP-…	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
CRCP-1	TBD	CRCP	0.3 mile	0.03 mile	10	TBD
CRCP-2	TBD	CRCP	0.3 mile	0.03 mile	10	TBD
CRCP-3	TBD	CRCP	0.3 mile	0.03 mile	10	TBD
CRCP-…	TBD	CRCP	0.3 mile	0.03 mile	10	TBD

The ground Reference Rating will be conducted by the agency, with the option of including the vendor personnel. All cracking equal to or greater than 1mm in width shall be reported. The sections and subsections will be clearly marked: Start, End, and Intermediate Points by the agency.

The vendor shall collect and report HPMS Cracking Percent data, to include the detailed measures as well as the HPMS Cracking Percentage on all designated reference sections. Data shall be reported at the specified subsection lengths. An agency may choose to require cracking data other than HPMS Cracking Percent. This alternative cracking data must be defined in the solicitation.

The HPMS Percent Cracking for asphalt concrete pavement (ACP), jointed concrete pavement (JCP) and continuously reinforced concrete pavement (CRCP) are defined as:

- For ACP, Cracking Percent is the percentage of the total area exhibiting visible fatigue type cracking (any cracking in the wheel path) for all severity levels in the wheel path in each section.
    - Cracking shall be measured and reported for both wheel paths. Measuring and reporting cracking outside of the wheel path areas is not required for HPMS reporting. Report the lineal length of cracking for each wheel path.
    - Any and all severity levels (sealed and unsealed) shall be reported.
    - Multiply the lineal length of wheel path with cracking by 39 in. to get the total area of wheel path cracking.
    - Cracking Percent is reported as the area of wheel path cracking divided by the total area of the section, multiplied by 100. To limit the potential variation in

reported lane width, for the purposes of this cracking comparison the lane width is set at 12 ft. wide.
    - Report Cracking Percent to the nearest one percent on each subsection.

- For JCP, Cracking Percent is the percentage of slabs within the section that exhibit transverse cracking. Partial slabs shall contribute to the section that contains the majority of the slab length.
    - Reported cracking for jointed concrete pavements excludes longitudinal cracks, corner breaks, D-cracking, and ASR cracking that may occur on a slab.
    - Report the actual total number of slabs in each subsection as well as the total number of slabs for the entire reference section.
    - Report the total number of slabs containing one or more transverse cracks extending for at least one-half the lane width. This is the number of cracked slabs.
    - The percentage of cracking reported is calculated as the number of cracked slabs, divided by the total number of slabs in the section, multiplied by 100.
    - Report Cracking Percent to the nearest one percent on each subsection.

- For CRCP, the Cracking Percent is the percentage of the area of the section exhibiting longitudinal cracking, punchouts, and/or patching. Transverse cracking shall not be considered in the Cracking Percent for CRCP.
    - Cracking and distresses may occur anywhere on the pavement. Transverse cracks that are at or near right angles to the direction of travel in the lane should not be included in the calculation.
    - Distresses to be included are longitudinal cracking (any severity), punchouts, and patched areas.
    - For longitudinal cracking, the cracked area is determined as the length of the crack multiplied by 1-ft. width.
    - For punchouts, the area is determined by the two transverse cracks and the edge of the pavement or longitudinal joint.
    - Report the area of all patches.
    - Percentage of Cracking for CRCP pavements is determined as the area of pavement where longitudinal cracking, punchouts, and patches are detected divided by the total area of the section, multiplied by 100. To limit the potential variation in reported lane width, for the purposes of this cracking comparison the lane width is set at 12 ft. wide.
    - Report Cracking Percent to the nearest one percent on each subsection.

HPMS Cracking Percent data shall be delivered to the agency in a data table following the formats illustrated in Tables 13, 14, and 15.

Table 13. Data Delivery Table for ACP.

Reference Site Number	Subsection Number	Subsection Station (mile)	Length of Wheel Path Cracking	HPMS Cracking Percent
ACP-1	1	0 to 0.03		
ACP-1	2	0.03 to 0.06		
ACP-1	…	…		
ACP-1	10	0.27 to 0.30		
ACP-2	1	0 to 0.03		
…	…	…		
ACP-3	10	0.27 to 0.30		
ACP-…				

Table 14. Data Delivery Table for JCP.

Reference Site Number	Subsection Number	Subsection Station (mile)	Total Number of Slabs	Number of Cracked Slabs	HPMS Cracking Percent
JCP-1	1	0 to 0.05			
JCP-1	2	0.05 to 0.10			
JCP-1	…	…			
JCP-1	10	0.45 to 0.50			
JCP-2	1	0 to 0.05			
…	…	…			
JCP-3	10	0.45 to 0.50			
JCP-…					

Table 15. Data Delivery Table for CRCP.

Reference Site Number	Subsection Number	Subsection Station (mile)	Length of Longitudinal Cracking	Area of Punchout	Area of Patching	Total Distress Area	HPMS Cracking Percent
CRCP-1	1	0 to 0.3					
CRCP-1	2	0.03 to 0.06					
CRCP-1	…	…					
CRCP-1	10	0.27 to 0.30					
CRCP-2	1	0 to 0.03					
…	…	…					
CRCP-3	10	0.27 to 0.30					
CRCP-…	…	…					

The agency will perform the statistical equivalence evaluation using the paired t-test for equivalence method. Each reference site will be evaluated on an independent basis. The risk will need to be balanced between the agency and vendor by establishing the statistical parameters (alpha, N, limits, power) used for the test in advance of the vendor's ratings. The agency and vendor can then determine, on a section by section basis, whether the vendor is producing cracking distress data that is equivalent to the reference ratings. For sections found not to be equivalent, the vendor will need to modify their equipment and/or crack evaluation processes until equivalent results are achieved. To accomplish the calibration, multiple attempts may be required by the vendor. If equivalency cannot be achieved with the predefined statistical parameters of alpha, power and limits, these can be modified until equivalency is achieved. The modified statistical parameters then indicate the increased risk to the agency associated with accepting vendor data on similarly distressed roadways.

APPENDIX D – GUIDELINES FOR EQUIPMENT AND SOFTWARE PURCHASING

For illustrative purposes, the process is demonstrated using the FHWA HPMS [(78)] cracking reporting standard. An agency may choose to use this standard or to use the illustration to apply to their own cracking definitions and needs.

When an agency is purchasing data collection equipment and software for in-house use, the following set of guidelines can be used in the procurement process. These guidelines are intended to provide guidance in accomplishing the agencies review of vendor data collection systems. This standard provides the appropriate format for collecting and preparing data reports under the current HPMS Field manual reporting requirements. The HPMS cracking percent is used for illustrative purposes. An agency may choose to require cracking data other than HPMS Cracking Percent. This alternative cracking data must be defined in the solicitation.

Equipment

The vendor shall provide an imaging system capable of collecting clear, high resolution images from which the cracking can be identified to a specified level. The system shall also be capable of quantifying the cracking distresses to the FHWA HPMS reporting standard and/or to meet the agency's needs.

In the procurement documents, an agency should request the vendor provide the following information, or alternatively require the vendor to meet or exceed the following criteria:

1. Imaging System Clarity
   Vendor should provide agency information regarding the number of pixels per image and the associated fill factor. Current recommended limits for 3D imaging are 16-bit images with a minimum of 4,096 pixels and a minimum fill factor of 90%

2. Image Resolution
   Imaging system must be capable of producing images in which a 1mm crack is visible when confirmed at slow speed and/or a 3mm crack at speeds exceeding 60 mph.

3. Image capture width
   The width of image shall cover a minimum of 13.5 ft. wide.

4. System capabilities
   a. Illumination source – The Downward Perspective Image shall be collected with a uniform and consistent form of illumination applied to the pavement surface. The illumination shall be regulated to provide sufficient contrast and crack-shadows for the clear discernment of cracking and patching. Images bearing ambient and/or vehicle shadows that obscure pavement features will not be accepted.

   b. Data collection speed – Data should be collected at or near prevailing highway speeds, typically 25 to 65 miles per hour.

c. Required system storage – The data storage file shall be self-describing and self-contained. Self-description here means that the data can be interpreted by different systems at different points in time. Self-contained means that all data needed for interpretation is included within the data "container." For example, metadata will not only provide the identifying information such as location and date, but also a string of information linking the header to all the accessible data stored for that specific location. FHWA is in the final publication phase of a report titled "Evaluation of Proposed Standard Data Format and Compression Algorithms for 2D/3D Pavement Surface Image" (78) which proposes a standard data format for the 2D/3D image systems. An agency may consider requiring the 2D and 3D images provided by the vendor comply with the proposed standard data format.

5. Define a quality image
   Image quality should provide the following characteristics: pavement surface is visible without shadows, reflection from wet surface, or other conditions of the imaging process resulting in images which cannot be clearly viewed.

6. Validation of Image Quality
   As part of the section below, ***Cracking Distress Data Verification Steps***, the image quality will be assessed. Ongoing NCHRP project 1-60, Measuring the Characteristics of Pavement Surface Images and Developing Standard Practices for Calibration, Certification, and Verification of Imaging Systems has the following objectives: (1) identify and develop methods for measuring the characteristics of surface images used for pavement evaluation and analysis; and (2) develop recommended standard practices for the calibration, certification, and verification of such images, for consideration and adoption by AASHTO. It is expected the results from this project will provide objective image quality validation protocols.

7. Proof of equipment hardware and software durability and utility
   Vendor shall provide references from agencies using the proposed equipment and software on systems of a similar size or larger. If the vendor proposes equipment or software that is new or has not been demonstrated, the vendor must address these concerns regarding the use of the equipment in their proposal. This could take the form of a warranty to perform, or support if equipment problems do arise.

8. Equipment configuration to provide the required data
   Vendor shall demonstrate that the system can provide the distress data in the HPMS reporting format and/or a format required by the agency. This could include the thoroughness and adaptability of software delivery to provide the agency distress types and severity levels. It should also include the ability to adjust any software tools used to identify cracking and/or interpret cracking patterns. This item may take the form of vendor support, at least for an initial period while agency personnel learn to perform this task alone.

The vendor equipment shall be able to collect and report HPMS Cracking Percent data, to include the detailed measures as well as the HPMS Cracking Percentage. An agency may choose to require cracking data other than HPMS Cracking Percent. This alternative cracking data must be defined in the solicitation.

The HPMS Percent Cracking for asphalt concrete pavement (ACP), jointed concrete pavement (JCP) and continuously reinforced concrete pavement (CRCP) are defined as:

- For ACP, Cracking Percent is the percentage of the total area exhibiting visible fatigue type cracking (any cracking in the wheel path) for all severity levels in the wheel path in each section.
    - Cracking shall be measured and reported for both wheel paths. Measuring and reporting cracking outside of the wheel path areas is not required for HPMS reporting. Report the lineal length of cracking for each wheel path.
    - Any and all severity levels (sealed and unsealed) shall be reported.
    - Multiply the lineal length of wheel path with cracking by 39 in. to get the total area of wheel path cracking.
    - Cracking Percent is reported as the area of wheel path cracking divided by the total area of the section, multiplied by 100. To limit the potential variation in reported lane width, for the purposes of this cracking comparison the lane width is set at 12 ft. wide.
    - Report Cracking Percent to the nearest one percent on each subsection.

- For JCP, Cracking Percent is the percentage of slabs within the section that exhibit transverse cracking. Partial slabs shall contribute to the section that contains the majority of the slab length.
    - Reported cracking for jointed concrete pavements excludes longitudinal cracks, corner breaks, D-cracking, and ASR cracking that may occur on a slab.
    - Report the actual total number of slabs in each subsection as well as the total number of slabs for the entire reference section.
    - Report the total number of slabs containing one or more transverse cracks extending for at least one-half the lane width. This is the number of cracked slabs.
    - The percentage of cracking reported is calculated as the number of cracked slabs, divided by the total number of slabs in the section, multiplied by 100.
    - Report Cracking Percent to the nearest one percent on each subsection.

- For CRCP, the Cracking Percent is the percentage of the area of the section exhibiting longitudinal cracking, punchouts, and/or patching. Transverse cracking shall not be considered in the Cracking Percent for CRCP.
    - Cracking and distresses may occur anywhere on the pavement. Transverse cracks that are at or near right angles to the direction of travel in the lane should not be included in the calculation.
    - Distresses to be included are longitudinal cracking (any severity), punchouts, and patched areas.
    - For longitudinal cracking, the cracked area is determined as the length of the crack multiplied by 1-ft. width.

- For punchouts, the area is determined by the two transverse cracks and the edge of the pavement or longitudinal joint.
- Report the area of all patches.
- Percentage of Cracking for CRCP pavements is determined as the area of pavement where longitudinal cracking, punchouts, and patches are detected divided by the total area of the section, multiplied by 100. To limit the potential variation in reported lane width, for the purposes of this cracking comparison the lane width is set at 12 ft. wide.
- Report Cracking Percent to the nearest one percent on each subsection.

9. Ensure the equipment provide the degree of automation which meets the agency's needs. The agency should specify their expectations for automated distress interpretation. The two types of crack detections and rating are:

    a. Semi-automated rating (rating is performed by computer processing of raw distress data, algorithms or other, and verified by manual reviews of the processed distress)

    b. Fully automated rating (distress classifications are produced by computer processing using algorithms or other methods, with little to no human intervention)

No matter the system selected, it is important that the agency validate the crack identification and classification process. This should be accomplished using reference sites where the agency has completed a detailed distress survey. The data should be compared using the statistical equivalence evaluation using the paired t-test for equivalence method.

Cracking Distress Data Verification Steps

The vendor should be requested to collect and report cracking data on the reference sites identified in Table 16. For illustration purposes, the HPMS Percent Cracking section length, subsection length and number of subsections are identified.

The agency has identified (number) of reference sites that shall be collected and reported. Sections are selected to represent typical pavement types and conditions throughout the agency.

Table 16. List of Reference Sections and Corresponding Details.

Reference Site Number	Location Description	Pavement Type	Section Length	Subsection Length	Number of Subsections	Weighting Factor
ACP-1	TBD	ACP	0.3 mile	0.03 mile	10	TBD
ACP-2	TBD	ACP	0.3 mile	0.03 mile	10	TBD
ACP-3	TBD	ACP	0.3 mile	0.03 mile	10	TBD
ACP-…	TBD	ACP	0.3 mile	0.03 mile	10	TBD
JCP-1	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
JCP-2	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
JCP-3	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
JCP-…	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
CRCP-1	TBD	CRCP	0.3 mile	0.03 mile	10	TBD
CRCP-2	TBD	CRCP	0.3 mile	0.03 mile	10	TBD
CRCP-3	TBD	CRCP	0.3 mile	0.03 mile	10	TBD
CRCP-…	TBD	CRCP	0.3 mile	0.03 mile	10	TBD

The ground Reference Rating will be conducted by the agency, with the option of including the vendor personnel. All cracking equal to or greater than 1mm in width shall be reported. The sections and subsections will be clearly marked with Start, End, and Intermediate Points by the agency.

The vendor shall collect and report HPMS Cracking Percent data, to include the detailed measures as well as the HPMS Cracking Percentage on all designated reference sections. Data shall be reported at the specified subsection lengths. An agency may choose to require cracking data other than HPMS Cracking Percent. This alternative cracking data must be defined in the solicitation.

HPMS Cracking Percent data shall be delivered to the agency in a data table following the formats illustrated in Tables 17, 18, and 19.

Table 17. Data Delivery Table for ACP.

Reference Site Number	Subsection Number	Subsection Station (mile)	Length of Wheel Path Cracking	HPMS Cracking Percent
ACP-1	1	0 to 0.03		
ACP-1	2	0.03 to 0.06		
ACP-1	…	…		
ACP-1	10	0.27 to 0.30		
ACP-2	1	0 to 0.03		
…	…	…		
ACP-3	10	0.27 to 0.30		
ACP-…				

Table 18. Data Delivery Table for JCP.

Reference Site Number	Subsection Number	Subsection Station (mile)	Total Number of Slabs	Number of Cracked Slabs	HPMS Cracking Percent
JCP-1	1	0 to 0.05			
JCP-1	2	0.05 to 0.10			
JCP-1	…	…			
JCP-1	10	0.45 to 0.50			
JCP-2	1	0 to 0.05			
…	…	…			
JCP-3	10	0.45 to 0.50			
JCP-…					

Table 19. Data Delivery Table for CRCP.

Reference Site Number	Subsection Number	Subsection Station (mile)	Length of Longitudinal Cracking	Area of Punchout	Area of Patching	Total Distress Area	HPMS Cracking Percent
CRCP-1	1	0 to 0.3					
CRCP-1	2	0.03 to 0.06					
CRCP-1	…	…					
CRCP-1	10	0.27 to 0.30					
CRCP-2	1	0 to 0.03					
…	…	…					
CRCP-3	10	0.27 to 0.30					
CRCP-…	…	…					

The agency will perform the statistical equivalence evaluation using the paired t-test for equivalence method. Each reference site will be evaluated on an independent basis. The risk will need to be balanced between the agency and vendor by establishing the statistical parameters (alpha, N, limits, power) used for the test in advance of the vendor's ratings. The agency and vendor can then determine, on a section by section basis, whether the vendor is producing cracking distress data that is equivalent to the reference ratings. For sections found not to be equivalent, the vendor will need to modify their equipment and/or crack evaluation processes until equivalent results are achieved. To accomplish the calibration, multiple attempts may be required by the vendor. If equivalency cannot be achieved with the predefined statistical parameters of alpha, power and limits, these can be modified until equivalency is achieved. The modified statistical parameters then indicate the increased risk to the agency associated with accepting vendor data on similarly distressed roadways.

Alternative Cracking Distress Data Verification Option – Proof of Verification

An agency may elect to allow a vendor to provide results of verification tests from another agency as a means of proving both their equipment capabilities and their ability to produce acceptable cracking distresses from the images. The vendor should be requested to submit documentation, similar to the process described above, from a reputable agency that is in a similar climatic region and has similar pavement types.

The vendor shall provide an imaging system capable of collecting clear, high resolution images from which the cracking can be identified to a specified level. The system shall also be capable of quantifying the cracking distresses to the FHWA HPMS reporting standard and/or to meet the agency's needs. An agency may choose to require cracking data other than HPMS Cracking Percent. This alternative cracking data must be defined in the solicitation.

In the procurement documents, an agency should request the vendor provide the following information, or alternatively require the vendor to meet or exceed the following criteria:

1. Imaging System Clarity
   Vendor should provide agency information regarding the number of pixels per image and the associated fill factor. Current recommended limits for 3D imaging are 16-bit images with a minimum of 4,096 pixels and a minimum fill factor of 90%.

2. Image Resolution
   Imaging system must be capable of producing images in which a 1mm crack is visible when confirmed at slow speed and/or a 3mm crack at speeds exceeding 60 mph.

3. Image capture width
   The width of image shall cover a minimum of 13.5 ft. wide.

4. System capabilities
   d. Illumination source – The Downward Perspective Image shall be collected with a uniform and consistent form of illumination applied to the pavement surface. The illumination shall be regulated to provide sufficient contrast and crack-shadows for the clear discernment of cracking and patching. Images bearing ambient and/or vehicle shadows that obscure pavement features will not be accepted.

   e. Data collection speed – Data should be collected at or near prevailing highway speed, typically 25 to 65 miles per hour.

   f. Required system storage – The data storage file shall be self-describing and self-contained. Self-description here means that the data can be interpreted by different systems at different points in time. Self-contained means that all data needed for interpretation is included within the data "container." For example, metadata will not only provide the identifying information such as location and date, but also a string of information linking the header to all the accessible data stored for that specific location. FHWA is in the final publication phase of a report titled "Evaluation of Proposed Standard Data Format and Compression Algorithms for 2D/3D Pavement Surface Image" [(2)] which proposes a standard data format for the 2D/3D image systems. An agency may consider requiring the 2D and 3D images provided by the vendor comply with the proposed standard data format.

5. Define a quality image
   Image quality should provide the following characteristics: pavement surface is visible without shadows, reflection from wet surface, or other conditions of the imaging process resulting in images which cannot be clearly viewed.

6. Validation of Image Quality
   Ongoing NCHRP project 1-60, Measuring the Characteristics of Pavement Surface Images and Developing Standard Practices for Calibration, Certification, and Verification of Imaging Systems has the following objectives: (1) identify and develop methods for measuring the characteristics of surface images used for pavement evaluation and analysis; and (2) develop recommended standard practices for the calibration, certification, and verification of such images, for consideration and adoption by AASHTO. It is expected the results from this project will provide objective image quality validation protocols.

7. Cracking distress interpretation type (An agency needs to choose which of the three cracking distress interpretations they desire to specify.)

   Options:
   - Manually rating (reviewing and classifying the cracks by viewing the digital images)
   - Semi-automated rating (rating is performed by computer processing of raw distress data, algorithms or other, and verified by manual reviews of the processed distress)
   - Fully automated rating (distress classifications are produced by computer processing using algorithms or other methods, with little to no human intervention)

8. The vendor should be requested to submit documentation from a reputable agency that is in a similar climatic region and has similar pavement types as the requesting agency. The documentation must include an analysis of the vendors cracking distress evaluations as compared to reference sites for ACP, JCP, and CRCP (agency to select pavement types). The documentation must include the details of the statistical equivalence evaluation, including the pass/fail criteria for each specific reference site.

9. The agency will review this documentation to determine acceptability of the vendor's ability to successfully determine cracking distresses.

Cracking Distress Data Verification Steps

The vendor should be requested to collect and report cracking data on the reference sites identified in Table 20. For illustration purposes, the HPMS Percent Cracking section length, subsection length and number of subsections are identified.

The agency has identified (number) of reference sites that shall be collected and reported. Sections are selected to represent typical pavement types and conditions throughout the agency.

Table 20. List of Reference Sections and Corresponding Details.

Reference Site Number	Location Description	Pavement Type	Section Length	Subsection Length	Number of Subsections	Weighting Factor
ACP-1	TBD	ACP	0.3 mile	0.03 mile	10	TBD
ACP-2	TBD	ACP	0.3 mile	0.03 mile	10	TBD
ACP-3	TBD	ACP	0.3 mile	0.03 mile	10	TBD
ACP-…	TBD	ACP	0.3 mile	0.03 mile	10	TBD
JCP-1	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
JCP-2	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
JCP-3	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
JCP-…	TBD	JCP	0.5 mile (or 100 slabs)	0.05 mile (or 10 slabs)	10	TBD
CRCP-1	TBD	CRCP	0.3 mile	0.03 mile	10	TBD
CRCP-2	TBD	CRCP	0.3 mile	0.03 mile	10	TBD
CRCP-3	TBD	CRCP	0.3 mile	0.03 mile	10	TBD
CRCP-…	TBD	CRCP	0.3 mile	0.03 mile	10	TBD

The ground Reference Rating will be conducted by the agency, with the option of including the vendor personnel. All cracking equal to or greater than 1mm in width shall be reported. The sections and subsections will be clearly marked with Start, End, and Intermediate Points by the agency.

The vendor shall collect and report HPMS Cracking Percent data, to include the detailed measures as well as the HPMS Cracking Percentage on all designated reference sections. Data shall be reported at the specified subsection lengths. An agency may choose to require cracking data other than HPMS Cracking Percent. This alternative cracking data must be defined in the solicitation.

The HPMS Percent Cracking for asphalt concrete pavement (ACP), jointed concrete pavement (JCP) and continuously reinforced concrete pavement (CRCP) are defined as:

- For ACP, Cracking Percent is the percentage of the total area exhibiting visible fatigue type cracking (any cracking in the wheel path) for all severity levels in the wheel path in each section.
    - Cracking shall be measured and reported for both wheel paths. Measuring and reporting cracking outside of the wheel path areas is not required for HPMS reporting. Report the lineal length of cracking for each wheel path.
    - Any and all severity levels (sealed and unsealed) shall be reported.
    - Multiply the lineal length of wheel path with cracking by 39 in. to get the total area of wheel path cracking.
    - Cracking Percent is reported as the area of wheel path cracking divided by the total area of the section, multiplied by 100. To limit the potential variation in

reported lane width, for the purposes of this cracking comparison the lane width is set at 12 ft. wide.
    - Report Cracking Percent to the nearest one percent on each subsection.

- For JCP, Cracking Percent is the percentage of slabs within the section that exhibit transverse cracking. Partial slabs shall contribute to the section that contains the majority of the slab length.
    - Reported cracking for jointed concrete pavements excludes longitudinal cracks, corner breaks, D-cracking, and ASR cracking that may occur on a slab.
    - Report the actual total number of slabs in each subsection as well as the total number of slabs for the entire reference section.
    - Report the total number of slabs containing one or more transverse cracks extending for at least one-half the lane width. This is the number of cracked slabs.
    - The percentage of cracking reported is calculated as the number of cracked slabs, divided by the total number of slabs in the section, multiplied by 100.
    - Report Cracking Percent to the nearest one percent on each subsection.

- For CRCP, the Cracking Percent is the percentage of the area of the section exhibiting longitudinal cracking, punchouts, and/or patching. Transverse cracking shall not be considered in the Cracking Percent for CRCP.
    - Cracking and distresses may occur anywhere on the pavement. Transverse cracks that are at or near right angles to the direction of travel in the lane should not be included in the calculation.
    - Distresses to be included are longitudinal cracking (any severity), punchouts, and patched areas.
    - For longitudinal cracking, the cracked area is determined as the length of the crack multiplied by 1-ft. width.
    - For punchouts, the area is determined by the two transverse cracks and the edge of the pavement or longitudinal joint.
    - Report the area of all patches.
    - Percentage of Cracking for CRCP pavements is determined as the area of pavement where longitudinal cracking, punchouts, and patches are detected divided by the total area of the section, multiplied by 100. To limit the potential variation in reported lane width, for the purposes of this cracking comparison the lane width is set at 12 ft. wide.
    - Report Cracking Percent to the nearest one percent on each subsection.

HPMS Cracking Percent data shall be delivered to the agency in a data table following the formats illustrated in Tables 21, 22, and 23.

Table 21. Data Delivery Table for ACP.

Reference Site Number	Subsection Number	Subsection Station (mile)	Length of Wheel Path Cracking	HPMS Cracking Percent
ACP-1	1	0 to 0.03		
ACP-1	2	0.03 to 0.06		
ACP-1	…	…		
ACP-1	10	0.27 to 0.30		
ACP-2	1	0 to 0.03		
…	…	…		
ACP-3	10	0.27 to 0.30		
ACP-…				

Table 22. Data Delivery Table for JCP.

Reference Site Number	Subsection Number	Subsection Station (mile)	Total Number of Slabs	Number of Cracked Slabs	HPMS Cracking Percent
JCP-1	1	0 to 0.05			
JCP-1	2	0.05 to 0.10			
JCP-1	…	…			
JCP-1	10	0.45 to 0.50			
JCP-2	1	0 to 0.05			
…	…	…			
JCP-3	10	0.45 to 0.50			
JCP-…					

Table 23. Data Delivery Table for CRCP.

Reference Site Number	Subsection Number	Subsection Station (mile)	Length of Longitudinal Cracking	Area of Punchout	Area of Patching	Total Distress Area	HPMS Cracking Percent
CRCP-1	1	0 to 0.3					
CRCP-1	2	0.03 to 0.06					
CRCP-1	…	…					
CRCP-1	10	0.27 to 0.30					
CRCP-2	1	0 to 0.03					
…	…	…					
CRCP-3	10	0.27 to 0.30					
CRCP-…	…	…					

The agency will perform the statistical equivalence evaluation using the paired t-test for equivalence method. Each reference site will be evaluated on an independent basis. The risk will need to be balanced between the agency and vendor by establishing the statistical parameters (alpha, N, limits, power) used for the test in advance of the vendor's ratings. The agency and vendor can then determine, on a section by section basis, whether the vendor is producing cracking distress data that is equivalent to the reference ratings. For sections found not to be equivalent, the vendor will need to modify their equipment and/or crack evaluation processes until equivalent results are achieved. To accomplish the calibration, multiple attempts may be required by the vendor. If equivalency cannot be achieved with the predefined statistical parameters of alpha, power and limits, these can be modified until equivalency is achieved. The modified statistical parameters then indicate the increased risk to the agency associated with accepting vendor data on similarly distressed roadways.

REFERENCES

1. Mendelsohn, D.H., "Automated Pavement Crack Detection: An Assessment of Leading Technologies," Second North American Pavement Management Conference, 1987.

2. Paterson, W.D., "Proposal of Universal Cracking Indicator for Pavements," Transportation Research Record, 1994.

3. Multiple Articles and Authors, "Pavement Management Systems," Transportation Research Record, TRR No. 1455, 1994.

4. American Society for Testing and Materials, "ASTM E 1778 – 98a Standard Terminology Relating to Pavement Distress," 1998.

5. American Society for Testing and Materials, "ASTM D 6433 – 07 Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys," 2007.

6. Frith, D., S. Stoffels, D. Morian, and C. Larson, "Quality Analysis Methods for Pavement Distress Data," Transportation Research Board Poster Session, January 7-11, 2001, Washington, D.C.

7. Frith, D., D. Morian, and S. Stoffels, "Quality Management of Pavement Performance Data," Pavement Evaluation 2002 Proceedings, October 21-25, 2002, Roanoke, Virginia.

8. Long-Term Pavement Performance Program, "Distress Identification Manual for the Long-Term Pavement Performance Program," 4th Edition, 2003.

9. McGee, K.H., "Automated Pavement Distress Collection Techniques," NCHRP Synthesis 334, 2004.

10. Halcrow Group Limited, "SCANNER Surveys for Local Roads-Specification Volume 1-Advice to Local Authorities, Volume 3-Acceptance Testing and Accreditation," 2005.

11. Frith, D., R. Shekharan, T. Chowdhury, C. Larson, and D. Morian, "Effects of Comprehensive Quality Assurance-Quality Control Plan on Pavement Management," TRB Record No. 1228, Transportation Research Board, Washington, D.C., 2007.

12. Steven, B., J.T. Harvey, and B. Tsai, "An Assessment of Automated Pavement Distress Identification Technologies in California," California Department of Transportation (Caltrans) Technical Memorandum, 2008.

13. Vitillo, N., N. Gucunski, C. Rascoe, and S. Zaghloul, "Evaluation of the Automated Distress Survey Equipment," New Jersey Department of Transportation, Federal Highway Administration, 2009.

14. Ong, Pj. (Raymond), S. Noureldin, and K.C. Sinha, "Automated Pavement Condition Data Collection Quality Control, Quality Assurance, and Reliability," Indiana Department of Transportation, 2010.

15. Wang, K.C.P., "Automated Pavement Cracking Analysis from 2D to 3D Challenges and Opportunities," Pavement Evaluation Conference, 2010.

16. NCHRP Report 677 "Development of Service Levels for the Interstate Highway System," Dye Management Group with Paul. D. Thompson Consulting and Quality Engineering Solutions, Inc., 2010.

17. Wang, K., "Potential for Comprehensive Evaluation of Pavement Surface with 3D Laser Imaging," 23rd Annual Road Profile User's Group (RPUG), 2011.

18. Wang, K., "Automated Survey of Pavement Distress based on 2D and 3D Laser Images," University of Arkansas, 2011.

19. Wang, K.C.P. and O. Smadi, "Automated Imaging Technologies for Pavement Distress Surveys," TRB Circular E-C156, 2011.

20. Olsen, M.J., G.V. Roe, C. Glennie, F. Persi, M. Reedy, D. Hurwitz, K. Williams, H. Tuss, A. Squellati, and M. Knodler; "Guidelines for the Use of Mobile LIDAR in Transportation Applications," NCHRP Report 748, 2013.

21. Tsai, Y. (James) and Z. Wang, "A Remote Sensing and GIS-Enabled Asset Management System (RS-GAMS)," Final Report, Georgia Institute of Technology, 2013.

22. Tsai, Y. (James) and Z. Wang, "A Remote Sensing and GIS-Enabled Asset Management System (RS-GAMS) Phase 2," Final Report, Georgia Institute of Technology, 2013.

23. Serigos, P.A., M. Burton, A. Smit, J.A. Prozzi, and M.R. Murphy, "Field Evaluation of Automated Distress Measuring Equipment," Texas Department of Transportation, 2014.

24. Tsai, Y. (James) and C. Jiang, "An Implementable Framework for Standardizing National Pavement Crack Measures," 9th International Conference on Managing Pavement Assets, 2014.

25. Frith, D.J., R.A. Shekharan, D.A. Morian, and T. Chowdhury, “Ten Years of Pavement Distress Independent Verification & Validation (IV&V) in Virginia,” 9th International Conference on Managing Pavement Assets, Washington, D.C., May 18-21, 2015.

26. Frith, D.J., S.M. Stoffels, D.A. Morian, and N. Mastin, “A Case Study in Establishing Quality Assurance Limits for Automated Pavement Distress Data in North Carolina,” 9th International Conference on Managing Pavement Assets, Washington, D.C., May 18-21, 2015.

27. Mathavan, S., K. Kamal, and M. Rahman, “A Review of Three-Dimensional Imaging Technologies for Pavement Distress Detection and Measurements,” Institute of Electrical and Electronics Engineers, 2015.

28. Serigos, P.A., K. Chen, A. Smit, M.R. Murphy, and J.A. Prozzi, “Automated Distress Surveys: Analysis of Network Level Data (Phase III),” Texas Department of Transportation, 2015.

29. Wang, K.C.P., Q.J. Li, G. Yang, Y. Zhan, and Y. Qiu, “Network Level Pavement Evaluation with 1 mm 3D Survey System,” Journal of Traffic and Transportation Engineering, 2015.

30. Murly, B., M. Jordan, and D. O’Brien, “Automated Pavement Condition Assessment Using Laser Crack Measurement System (LCMS) on Airfield Pavements in Ireland,” 9th International Conference on Managing Pavement Assets, 2015.

31. Becerik-Gerber, B., S.F. Masri, and M.R. Jahanshahi, “An Inexpensive Vision-Based Approach for the Autonomous Detection, Localization, and Quantification of Pavement Defects,” Final Report, NCHRP IDEA Project 169, 2015.

32. Wang, K.C.P., Q. (Joshua) Li, and C. Chen, “Development of Standard Data Format for 2-Dimensional and 3-Dimensional (2D/3D) Pavement Image Data used to determine Pavement Surface Condition and Profiles,” Task 2 - Research Current Practices, Federal Highway Administration, 2016.

33. Wang, K.C.P., Q. (Joshua) Li, and C. Chen, “Development of Standard Data Format for 2-Dimensional and 3-Dimensional (2D/3D) Pavement Image Data used to determine Pavement Surface Condition and Profiles,” Task 3 - Evaluate Data Items and Formats, Federal Highway Administration, 2016.

34. Wang, K.C.P., Q. (Joshua) Li, and C. Chen, “Development of Standard Data Format for 2-Dimensional and 3-Dimensional (2D/3D) Pavement Image Data used to determine Pavement Surface Condition and Profiles,” Task 4 - Develop Metadata and Proposed Standards, Federal Highway Administration, 2016.

35. Committee of Transport Officials, “Automated Road Condition Assessments Part G: Imaging,” Technical Methods for Highways (TMH) 13, The South African National Roads Agency SOC Limited, 2016.

36. Wang, K.C.P., "Artificial Intelligence: Digital Image Library Ground Truthing and NCHRP 1-57A," TFP-5(299) meeting, RPUG meeting, 2017.

37. Michigan Department of Transportation, "Pavement Condition Data Collection and Processing RFP for Years 2018 – 2021," 2017.

38. Tsai, Y. (James) and A. Chatterjee, "Comprehensive, Quantitative Crack Detection Algorithm Performance Evaluation System," Journal of Computing in Civil Engineering, American Society of Civil Engineers (ASCE), 2017.

39. MSDOT 2018, Distress Classification Guide, a Supplemental Guide, Pathway Services Inc., revised June 2018.

40. Example Index and Distress Check Sheet, Oregon Department of Transportation.

41. https://www.oregon.gov/ODOT/Construction/Pages/Pavement-Management-System.aspx, July 1, 2018

42. Northwest Pavement Management Systems Users Group, R. Keith Kay, "Pavement Surface Condition Rating Manual, Northwest Pavement Management Systems Users Guide," Northwest Technology Transfer Center, Local Programs Division, Washington State Department of Transportation, March 1992. https://www.wsdot.wa.gov/NR/rdonlyres/1AB0E29D-72D7-466A-9547-C9F631B4CE6C/0/PavementSurfaceConditionRatingManual.pdf, July 1, 2018

43. A Guide to Evaluating Pavement Distress through the Use of Digital Images, Version 2.6, Virginia Department of Transportation, Maintenance Division, November 2012.

44. "Data Quality Management Plan v 0.98," Materials & Tests, Pavement Management Section, Alabama Department of Transportation, May 21, 2018.

45. Caltrans Final Data Quality Management Plan with FHWA Approval Letter, Pavement Program Maintenance Division, California Department of Transportation, March 2018.

46. "Pavement Data Quality Management Program," Final Draft, Colorado Department of Transportation, May 2018.

47. Weigel, S., Network-Level Pavement Condition Data Collection Quality Management Plan, North Dakota Department of Transportation, Planning/Asset Management Division, March 2018.

48. Pavement Data Quality Management Plan, Virginia Department of Transportation, May 2018.

49. Data Quality Management Plan for Pavement Condition National Highway Performance Program, Oregon Department of Transportation Pavement Services Unit, Version 1.1, October 2, 2018.

50. “Using the Preliminary Prioritization Report Generation Tool,” Materials & Tests, Pavement Management Section, Alabama Department of Transportation, March 23, 2016.

51. North Carolina Department of Transportation, Decision Trees: 10 Decision Tree Images.

52. North Dakota Department of Transportation, Decision Tree, March 2014.

53. General System Decision Tree, Ohio Department of Transportation, Office of Pavement Engineering, October 20, 2017.

54. Priority System Decision Tree, Ohio Department of Transportation, Office of Pavement Engineering, October 20, 2017.

55. Pavement Management System Decision Trees, Virginia Department of Transportation.

56. Request for Proposal (RFP) Number 56A0505, Automated Pavement Condition Survey (APCS), California Department of Transportation, May 2, 2017.

57. Pavement Data Collection 2017 Statement of Work, EC SMH, Colorado Department of Transportation.

58. Request for Proposals, Exhibit 2, Scope of Work 2017-2020, Mississippi Department of Transportation, March 2, 2017.

59. Request for Proposals (RFP) High Speed Data Collection, Final, North Carolina Department of Transportation, 2015.

60. Request for Proposals (RFP) Number 30-25421-14, Automated Pavement Data Management Collection, Oregon Department of Transportation, February 26, 2014.

61. 2018 Videologging RFP – Test Project & Work Statement for Automated Data Collection, Inventory and Analysis of Pavement Distress, Initial Submission, Pennsylvania Department of Transportation, August 21, 2017.

62. Specification No. TxDOT 968-63-65, Pavement Condition Data Collection Services, Texas Department of Transportation, Revised May 2017.

63. Request for Proposal, Requisition Number 2383, Collection & Processing of Detailed Pavement Surface Condition Metric & Roadway Imagery Data (2018-2021 Collection Seasons), Michigan Department of Transportation, December 14, 2017.

64. Request for Proposal: Videolog Digital Imaging Services, RFP# 48400-196-0000070467, Georgia Department of Transportation, April 2018.

65. Highway Performance Monitoring System (HPMS) Field Manual, Chapter 4.4, Item 52, Federal Highway Administration, December 2016.

66. Wang, K.C.P., et.al, "Cognition-Based Intelligent Solutions for Condition and Safety Surveys," presentation at 2018 RPUG, Rapid City, SD.

67. Tsai, Yichang (James), et.al, "Evaluation of Proposed Standard Data Format and Compression Algorithms for 2D/3D Pavement Surface Image", FHAW Report TBD, December 2019.

68. Quality Analysis Methods for Pavement Distress Data", Transportation Research Board Poster Session, January 7-11, 2001, Washington, D.C. with Stoffels, S.M., Morian, D.A., and Larson, C.

69. Hsu, Jason C., J. T. Gene Hwang, Hung-Kung Liu and Stephen J. Ruberg, "Confidence Intervals Associated with Tests for Bioequivalence," Biometrika, Oxford University Press, Vol. 81, No. 1 (Mar., 1994), pp. 103-114.

70. Lung, K. Rick, Mary A. Gorko, Jennifer Llewelyn and Norman Wiggins, "Statistical method for the determination of equivalence of automated test procedures," Journal of Automated Methods & Management in Chemistry, Vol. 25, No. 6 (November-December 2003) pp. 123-127.

71. Wellek, Stefan, Testing Statistical Hypotheses of Equivalence and Noninferiority, 2nd ed., CRC Press, Taylor & Francis Group, Boca Raton, FL, 2010.

72. Lakens, Daniel, "Equivalence Tests: A Practical Primer for Tests, Correlations, and Meta-Analyses," Social Psychological and Personality Science, Vol. 8(4), pp. 355-362, 2017.

73. Pardo, Scott, Equivalence and Noninferiority Tests for Quality, Manufacturing and Test Engineers, 1st ed., CRC Press, Taylor & Francis Group, Boca Raton, FL, June 2019.

74. Simpson, A. L., Rada, G. R., Visintine, B. A., & Groeger, J. (2016). Interstate Pavement Condition Sampling (No. FHWA-HIF-17-022). United States. Federal Highway Administration. Office of Asset Management.

75. Shahin, M.Y. and Kohn, S.D. (1981). Pavement Maintenance Management for Roads and Parking Lots. Technical Report No. M-294. U.S. Army Construction Engineering Laboratory.

76. https://www.law.cornell.edu/cfr/text/23/490.313 (November 25, 2019).

77. http://www.minitab.com/en-US/default.aspx (January 24, 2020).

78. Highway Performance Monitoring System (HPMS) Field Manual, Office of Highway Policy Information, Office of Management & Budget (OMB) Control No. 2125-0028, December 2016.

79. Tsai, Yichang (James), et.al, “Evaluation of Proposed Standard Data Format and Compression Algorithms for 2D/3D Pavement Surface Image,” Federal Highway Administration Report TBD, December 2019.