Practical Methods of Comparing Vendor-Provided Cracking Distress Data to Agency-Produced Ground ~~Truth~~ Reference Data

Douglas Frith
Dennis Morian

QES Quality Engineering Solutions
Engineering • Inspection

Acknowledgements

TPF-5(299) Improving the Quality of Pavement Surface Distress and Transverse Profile Data Collection and Analysis

Dr. Shelley Stoffels – Penn State

Andy Mergenmeier – FHWA

Dennis Dvorak – FHWA

Project Objectives

- Establish a methodology for state agencies to use in selecting a pavement data collection vendor
- Establish a methodology for agencies to evaluate vendor equipment for purchase to conduct their own pavement data collection
- The TPF-5(299) technical advisory committee selected the HPMS definitions as the data objective to use for this study

Selection of Ground Reference Method

- Manual Surveys – Lane Closed to Traffic
- Manual Surveys – From Edge of Pavement
- Windshield Surveys
- Manual Ratings – From Digital Images
- Semi-Automated Ratings – From Digital Images

Location details and lane markings are critical!

Ground Reference Variability

- Manual Surveys – LTPP Study, Rada, et al, 1999
- Manual Surveys – QES staff, PCI Surveys, 2013-2019
- Manual Surveys – State A, 2018
- Manual Windshield Surveys – State B, PCC, June 2019

Accuracy and Precision for LTPP Manual Ratings

AC Pavements

Distress Type	Unit	Distress Severity	Pooled Reference	Group Statistics			
				Mean	Std. Dev.	COV (%)	Bias
Fatigue Cracking	Sq. meters	All Levels (Total)	14.2	16.5	6.2	38	2.3
Longitudinal Cracking WP	meters	All Levels (Total)	18.4	18.3	6.0	33	-0.2
Longitudinal Cracking NWP	meters	All Levels (Total)	75.0	70.7	14.7	21	-4.3
Transverse Cracking	number	All Levels (Total)	26.4	24.7	3.2	13	-1.7
Transverse Cracking	meters	All Levels (Total)	44.3	44.6	4.2	9	0.3

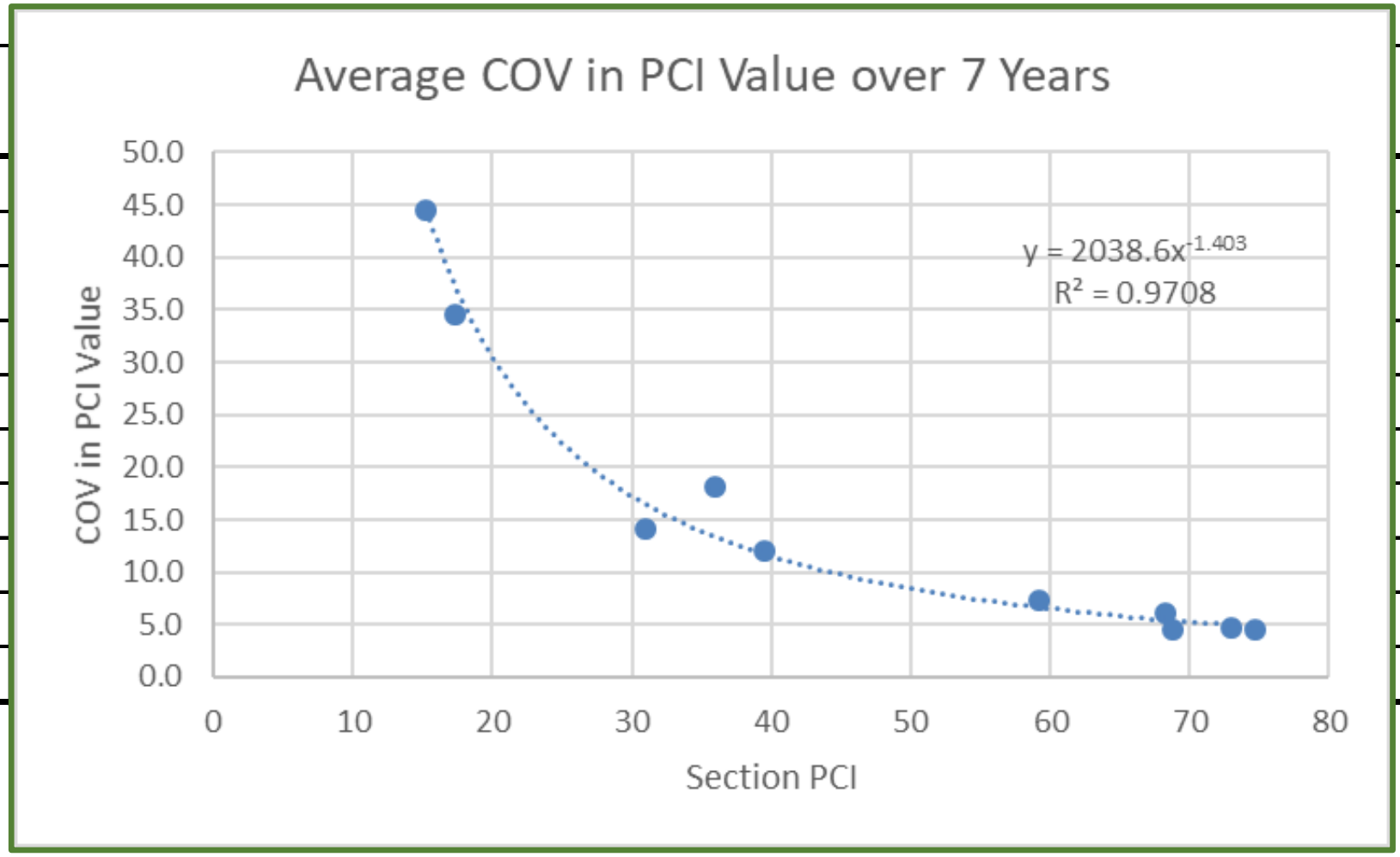
(Rada, et al., 1999)

PCC Pavements

Distress Type	Unit	Distress Severity	Pooled Reference	Group Statistics			
				Mean	Std. Dev.	COV (%)	Bias
Corner Breaks	number	All Levels (Total)	3.9	3.7	0.5	14	-0.2
Longitudinal Cracking	meters	All Levels (Total)	7.5	7.0	1.6	22	-0.5
Transverse Cracking	number	All Levels (Total)	9.4	9.6	1.4	15	0.2
Transverse Cracking	meters	All Levels (Total)	24.8	25.0	2.1	8	0.2
Spalling of Long. Joints	meters	All Levels (Total)	6.6	7.2	4.9	68	0.5
Spalling of Trans. Joints	number	All Levels (Total)	3.7	3.4	0.9	25	-0.3
Spalling of Trans. Joints	meters	All Levels (Total)	1.7	2.0	1.4	71	0.3

QES Staff PCI Surveys on AC pavements

Site #	Average All Years
1	4.5
2	34.5
3	7.3
4	4.5
5	14.1
6	4.7
7	18.1
8	12.0
9	6.1
10	44.5



State A Cracking Surveys

Absolute Difference between Reference Rater and Evaluator Rater

Distress Type	Trans 1	Trans 2	Trans 3	Trans Total
Average % Cracking Difference	0.2	0.8	0.1	0.3
Maximum % Cracking Difference	0.9	5.7	0.7	1.5
Distress Type	WP 1	WP 2	WP 3	WP Total
Average % Cracking Difference	0.4	5.9	0.4	0.3
Maximum % Cracking Difference	1.3	50.4	2.7	1.3
Distress Type	NWP 1	NWP 2	NWP 3	NWP Total
Average % Cracking Difference	0.4	8.4	0.0	0.3
Maximum % Cracking Difference	0.8	79.5	0.0	0.8

10 control sites, each 0.3 miles long
Two experienced raters: Reference Rater, Evaluator Rater

State B PCC Surveys

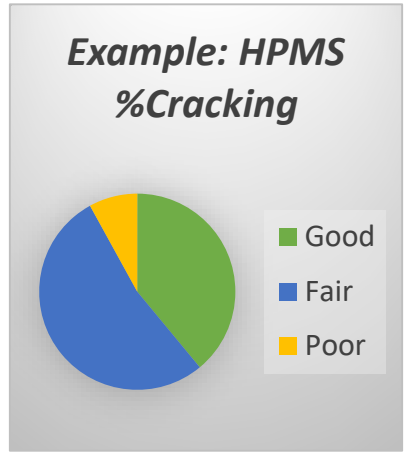
Overall Agreement Among Raters in Windshield Survey of PCC Pavements

Distress Type	Agreement Among Raters in Total Distress Amount (All Severities), 100 - COV (%)	Percentage of Ratings more than one STD away from AVG (outliers)
Transverse Cracking	82%	25%
Longitudinal Cracking	82%	28%
Spalling	64%	30%
Corner Cracking	56%	22%
Shattered Slabs	71%	26%
Patching	74%	33%
Surface Deterioration	52%	45%

Framework of Guidelines – Step 1

- Agency selects control sites
  - Recommended minimum length of 0.3 miles
- Sites should be chosen to represent the agency's
  - Pavement types
  - Surface textures
  - Frequently-occurring distress conditions

    Example: Sections with HPMS %Cracking Good, Fair and Poor
  - Distresses of high concern and impact in the decision-making process
- Minimum of six sites is recommended, but each is statistically evaluated independently



Framework of Guidelines – Step 1

- Agency performs manual cracking distress ratings under closed traffic lanes
  - Sections and rating intervals must be clearly marked: Start, End and Intermediate Pts
    - Rating intervals of minimum length 0.03 miles
    - Control sections should have minimum of 10 rating intervals
  - Replicate ratings are recommended for determination of the ground reference
    - Equivalence of ratings should be assessed
    - Single rating or consensus used as reference
    - Replicates should not be averaged
  - Mark or use template to identify wheel paths, following HPMS directives
  - Develop summarized cracking distresses at 0.03-mile intervals
    - HPMS directives
    - Agency definitions

Subsections for control site surveys must be clearly marked

This state agency uses a template for manual ground reference surveys

Framework of Guidelines – Step 2

- Vendors collect digital images over all control sites
- Vendors submit images, viewing, and manual rating software
- Vendors submit cracking distress summaries at 0.03-mile increments
  - Important that the increments match the pavement markings used for the ground reference ratings
- Agency may choose to view and/or perform independent ratings from the images

Framework of Guidelines – Step 3

- Agency completes a statistical evaluation of the submitted data
  - Determines a pass/fail for each control site
- Agency should identify pass/fail criteria before data is evaluated
- Agency may weight some control sites differently than others
  - Key distresses
  - Fine cracking
  - Pavement type

Framework of Guidelines – Step 4

Statistical Tests Evaluated:

- d2s: 95% limits on the difference between two test results (ASTM 1998)
- Student's t-Test
- Pearson Correlation Coefficient
- Paired t-Test
- Equivalence Test with Paired Data

Equivalence Testing with Paired Data

- Hypothesis tests are typically used to look for a **difference** in the results large enough to affect the outcome

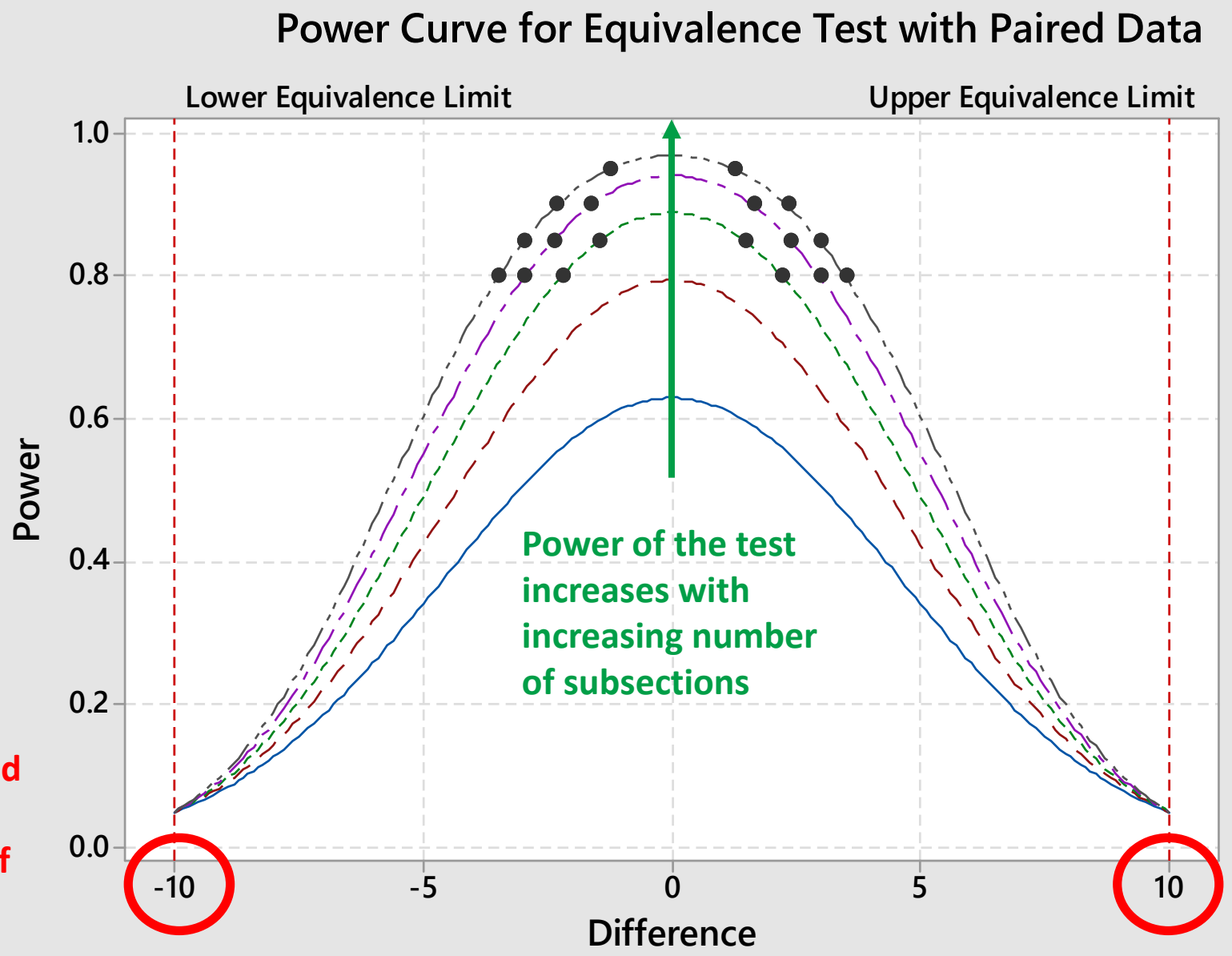
  Example: *Does the treatment significantly improve the condition?*

- Our goal is 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

Equivalence Testing with Paired Data

- In equivalence testing, **α (alpha) is the risk level of accepting a method as equivalent when it is not**
- β (beta) is the risk level of rejecting equivalence when the data is in fact equivalent
- **Power (1-β) is the likelihood of correctly concluding that the difference is within the equivalence limits, when this is true**
- N is the number of subsections; more subsections reduce risks and increase the power of the test
- The upper and lower limits should be set at the differences that would have a meaningful impact on the outcomes

Power is the likelihood of correctly concluding that the difference is within the equivalence limits



α is the agency's risk of accepting a vendor as equivalent when it is not

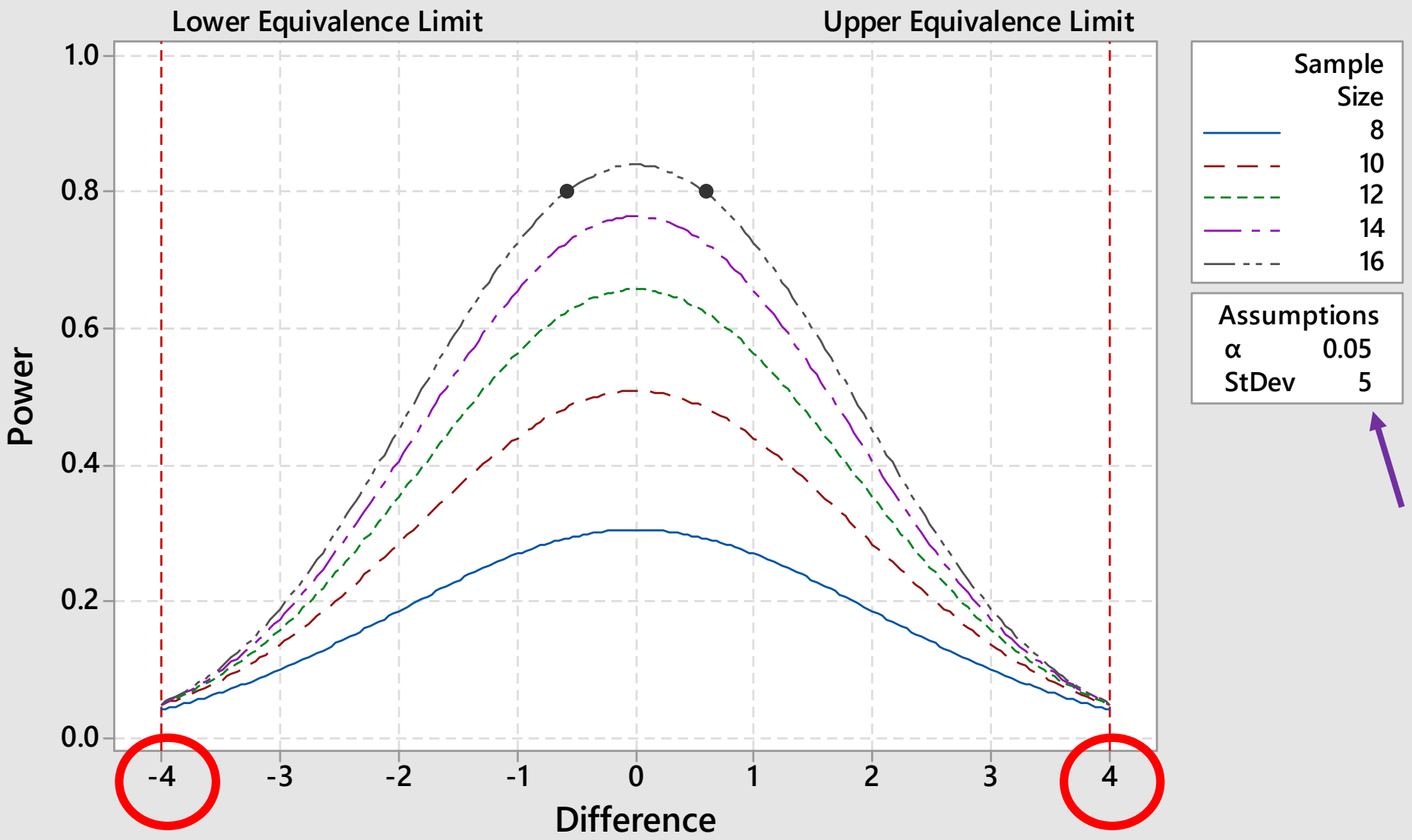
StDev based on historical data and can be updated

Limits are based upon the intended use of the data

PE 2019

QES

RPUG Road Profile Users' Group


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
-4
-3
-2
-1
0
1
2
3
4
Difference
Sample Size
8
10
12
14
16
Assumptions
α 0.05
StDev 5

Interpreting the results

- A fictional agency rates 6 control sites of AC and 3 control sites of PCC
- Vendor ratings are compared to the ground reference rating for each site
  - Some sites are found to be equivalent
  - Some sites are not found to be equivalent
- Drawing Conclusions
  - Based on a priority weighting of sections and or pavement types

Control Site ID	Pavement Type	HPMS % Cracking	Site Weight	Vendor A	Vendor B
A	AC	Good	2	EQ	Not EQ
B	AC	Fair	2	Not EQ	Not EQ
C	AC	Fair	1	EQ	EQ
D	AC	Fair	1	Not EQ	EQ
E	AC	Poor	1	EQ	EQ
F	AC	Poor	1	Not EQ	EQ
G	PCC	Good	1	EQ	EQ
H	PCC	Fair	2	EQ	EQ
I	PCC	Poor	1	Not EQ	EQ
			SCORE	7	8

Thank You!

QESpavements.com