**JOINT TRANSPORTATION RESEARCH PROGRAM**

INDIANA DEPARTMENT OF TRANSPORTATION
AND PURDUE UNIVERSITY

# Probability of Detection Study for Visual Inspection of Steel Bridges: Volume I—Executive Summary


**Leslie E. Campbell, Luke R. Snyder,
Julie M. Whitehead, Robert J. Connor, Jason B. Lloyd**

SPR-3820 • Report Number: FHWA/IN/JTRP-2019/21 • DOI: 10.5703/1288284317103

## RECOMMENDED CITATION

Campbell, L. E., Snyder, L. R., Whitehead, J. M., Connor, R. J., & Lloyd, J. B. (2019). *Probability of detection study for visual inspection of steel bridge: Volume I—Executive summary* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2019/21). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284317103

## AUTHORS

**Leslie E. Campbell**
**Julie M. Whitehead**
Graduate Research Assistants
Lyles School of Civil Engineering
Purdue University

**Luke R. Snyder**
Graduate Research Assistant
School of Electrical and Computer Engineering
Purdue University

**Robert J. Connor, PhD**
Jack and Kay Hockema Professor
Director of CAI and S-BRITE
Lyles School of Civil Engineering
Purdue University
(765) 496-8272
rconnor@purdue.edu
*Corresponding Author*

**Jason B. Lloyd, PhD, PE**
Bridge Steel Specialist, Western Market
National Steel Bridge Alliance

## JOINT TRANSPORTATION RESEARCH PROGRAM

The Joint Transportation Research Program serves as a vehicle for INDOT collaboration with higher education in-stitutions and industry in Indiana to facilitate innovation that results in continuous improvement in the planning, design, construction, operation, management and economic efficiency of the Indiana transportation infrastructure. https://engineering.purdue.edu/JTRP/index_html

Published reports of the Joint Transportation Research Program are available at http://docs.lib.purdue.edu/jtrp/.

## NOTICE

# TECHNICAL REPORT DOCUMENTATION PAGE

| | | |
|---|---|---|
| **1. Report No.**<br>FHWA/IN/JTRP-2019/21 | **2. Government Accession No.** | **3. Recipient's Catalog No.** |
| **4. Title and Subtitle**<br>Probability of Detection Study for Visual Inspection of Steel Bridges: Volume 1—Executive Summary | | **5. Report Date**<br>August 2019<br>**6. Performing Organization Code** |
| **7. Author(s)**<br>Leslie E. Campbell, Luke R. Snyder, Julie M. Whitehead, Robert J. Connor, and Jason B. Lloyd | | **8. Performing Organization Report No.**<br>FHWA/IN/JTRP-2019/21 |
| **9. Performing Organization Name and Address**<br>Joint Transportation Research Program<br>Hall for Discovery and Learning Research (DLR), Suite 204<br>207 S. Martin Jischke Drive<br>West Lafayette, IN 47907 | | **10. Work Unit No.**<br>**11. Contract or Grant No.**<br>SPR-3820 |
| **12. Sponsoring Agency Name and Address**<br>Indiana Department of Transportation (SPR)<br>State Office Building<br>100 North Senate Avenue<br>Indianapolis, IN 46204 | | **13. Type of Report and Period Covered**<br>Final Report<br>**14. Sponsoring Agency Code** |

**15. Supplementary Notes**

Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

**16. Abstract**

An inspector's ability to correctly identify surface and internal defects in steel bridge components is critical to protecting public safety. Ensuring that inspectors are properly trained and adequately equipped to detect these defects in locations that are difficult to access and/or in unfavorable environmental conditions must be a high priority. While the Federal Highway Administration and individual state departments of transportation have guidelines for inspector qualifications, trainings, and certifications, there is very little emphasis placed on evaluating or "testing" a given inspector's capability to characterize detect defects in the field. As a result, there is also very little if any data on how well a given inspector actually performs or the variability which can be expected between various inspectors. This comprehensive Probability of Detection (POD) study was conducted to establish the ability of an inspector with the current required training to locate cracks in steel bridge components using typical visual inspection techniques.

| | |
|---|---|
| **17. Key Words**<br>bridge inspection, fatigue cracking, probability of detection | **18. Distribution Statement**<br>No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161. |

| **19. Security Classif. (of this report)** | **20. Security Classif. (of this page)** | **21. No. of Pages** | **22. Price** |
|---|---|---|---|
| Unclassified | Unclassified | 24 | |

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

## CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

## INTRODUCTION

An inspector's ability to correctly identify surface and internal defects in steel bridge components is critical to protecting public safety. Ensuring that inspectors are properly trained and adequately equipped to detect these defects in locations that are difficult to access and/or in unfavorable environmental conditions must be a high priority. While the Federal Highway Administration and individual state departments of transportation have guidelines for inspector qualifications, trainings, and certifications, there is very little emphasis placed on evaluating or "testing" a given inspector's capability to characterize and detect defects in the field. As a result, there is also very little if any data on how well a given inspector actually performs or the variability which can be expected between various inspectors.

This comprehensive Probability of Detection (POD) study was conducted to establish the ability of an inspector with the current required training to locate cracks in steel bridge components using typical visual inspection techniques.

Specifically, during this research, a POD study focused on visual detection of fatigue cracks in steel bridge girders was performed. Trained and experienced bridge inspectors were asked to conduct routine and hands-on inspections of a simulated steel bridge components with known defects. The hands-on portion of the inspection course was conducted from a man-lift outdoors to better simulate field bridge inspections. The focus of the test inspection course was to locate and identify cracks in steel components. The details represented on the test bridge mimic many of those currently in the bridge inventory. Environmental and personal factors were tracked and compared to detection rates. This study is believed to be the first statistically significant study of its kind in the United States related to visual inspection of cracks in steel bridge components.

The study has shown that routine visual inspection from the ground and hands-on visual inspection of weathering and painted steel bridge girders are unreliable for finding cracks. In fact, the routine inspection scenario and the hands-on inspection of the weathering steel specimens were discontinued partway through the study due to low detection rates and to allow adequate time for the hands-on inspection of the painted specimens. Based on the results of the POD study, recommendations to improve the reliability of hands-on visual inspection of painted steel bridge girders have been made.

It is important to note that the intention of this study was to evaluate the reliability of visual inspection on steel bridges, not to find fault with individual inspectors. The research set out to answer the questions "What cracks are being missed?" and "Are our expectations for inspectors reasonable?" If inspectors are being expected to find cracks that cannot reasonably or reliably be found with the current procedures, the system should be reevaluated.

## METHODS

The goal of the POD study was to determine the effectiveness of visual inspection (VI) with respect finding cracks in steel bridge components. At its most basic level, a POD study predicts a probability of detection for a given inspection technique as a function of crack length. Visual inspection is the most fundamental and widely used technique for bridge inspection. It is very often used independently from other NDT methods, but also supplements techniques like magnetic particle (MPT) or dye penetrant testing (DPT).

To analyze the data using the industry standard probability of detection analysis tools, which were developed and outlined in Military Handbook 1823a (DoD, 2009), the test course must meet certain criteria to be statistically significant.

Probability of detection studies are used to determine the effectiveness of an inspection technique and produce curves that relate the likelihood of detection to a crack length. POD studies are widely used by other industries, like oil and gas, as well as the aircraft industry, and are typically used with other, less variable forms of NDE, like eddy current, MPT, DPT, or ultrasonic testing (UT).

There is a tendency in POD studies to reduce the number of specimens because of the associated cost of the specimens themselves and the process of installing realistic defects. POD researchers have suggested crack size ranges, numbers of cracks, and a specified noise ratio to ensure meaningful results. (Specimens or details without cracks are referred to as "noise.") It is important that the inspection course contain more defect free specimens than specimens with cracks. The most important part of a POD study is that the defects must be detectable by the method being evaluated.

A test frame was constructed at Purdue's S-BRITE Center to simulate a 2-span, 3-girder steel highway bridge. The frame supports 108 W-shape specimens and additional cover plate specimens. Each specimen is approximately 2 feet in length and generally contains more than one detail. Since the details of interest are compressed so to speak into each short specimen, the bridge actually represents three girder lines approximately 400 feet in length each. Each span is 40 feet long and girders are spaced at 8 feet 4 inches. The specimens are suspended approximately 25 feet above grade. To reflect INDOT's bridge inventory and that of other states in the pooled fund project TPF-5(281), one span was dedicated to weathered steel specimens and one to painted specimens. The test frame can be seen in Figure 1.

Several features were added to the test frame to improve the realism of the environment. By being 25 feet above the ground, a man-lift is needed for the hands-on inspection portion of the course. It was important to mimic the challenges of real inspection environments—to have the inspectors "bounce" in the man-lift and be constrained by their safety equipment. A wooden deck was installed to provide shading similar to that provided by a concrete deck. Wooden cross braces were attached

**Figure 1** POD test frame.

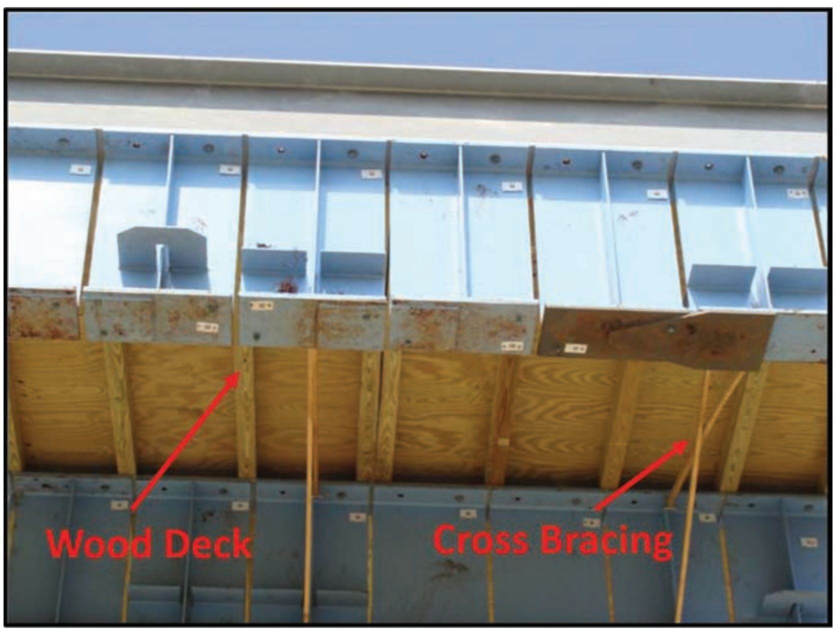


**Figure 2** Wood deck and cross bracing.

periodically along the girders to obstruct the inspectors' access, similar to other elements on real bridges that obstruct access. These features can be seen in Figure 2 and Figure 3.

The test course included a variety of specimen types. The W-shapes, which were either rolled W36 pieces or 36-inch-deep plate girders, included an assortment of attachments. These attachments included transverse stiffeners, longitudinal stiffeners, gusset plates, and bearing stiffeners. The pieces were each given an individual identification tag that identifies the attachment type and coating. Samples of the specimens can be seen below in Figure 4 and Figure 5.

There were two varieties of welded cover plate terminations included in the specimen matrix, tapered and square ended, which can be seen in Figure 6. There were also riveted plates, shown in Figure 7. These are 9-foot-long plates each with 54 rivets. Half of the riveted plate specimens are attached horizontally to the bottom flanges of the W-shape pieces to represent riveted cover plates, and half are mounted vertically on the columns to represent vertical truss members.

Three crack types were included in the study. The final inspection course had 70 possible hits on the painted portion of the course. The course also included 18 possible hits on the weathered specimens. However, due to the drastic differences in detection rates between the two coating types, the data were analyzed separately for the painted and weathered specimens. Out-of-plane cracks were created through cyclic fatigue loading in Bowen Laboratory. The weld toe cracks on cover plates were also introduced through cyclic fatigue loading at Bowen Laboratory. The crack lengths were specified and the specimens were cycled and monitored until the crack reached the desired lengths that satisfied the statistical requirements of the study. The cracks at the rivet holes were drawn and detailed at Purdue and then cut with an EDM (electrode discharge machine) wire. Pictures of each crack type can be seen in Figure 8, Figure 9, and Figure 10.

The cracks ranged in size from 1/2 inch to 5-3/8 inches. These cracks were split into four crack size ranges. Table 1 lists the number of cracks in each size range used for the painted portion of the hands-on inspection course.

It was essential that the inspection procedures be held constant among each of the inspectors so that the method of inspection and human factors could be evaluated. Each inspector received the same information via email before arriving at the test, were read the same instructions the morning of their test, given an identical binder of forms for their notes, asked to sign a confidentiality agreement stating that they would not discuss the course with any other inspectors, and required to complete two vision examinations. The test procedures provided to each inspector included a list of assumptions to make regarding the specimens and course. These assumptions were as follows:

1. The specimens are intended to represent 1960s to 1970s welded fabrication and weld quality. While weld quality may not meet modern standards, it is not the focus of the study. You are not required to comment on weld quality.
2. Assume the pieces suspended from the frame are fracture critical members.
3. Treat all gusset plates and vertical stiffeners as if bracing, floor beams, diaphragms, or cross-bracing are attached. For example, many specimens include gusset plates and vertical stiffeners welded to webs. Therefore, you are to assume there are lateral members attached to these components whether they are on the interior or exterior faces of the specimens or near the top or bottom flange. The bracing members have not been included to facilitate your access.
4. Assume all cover plate terminations are subject to tensile stress ranges.
5. Assume both flanges could be tension flanges.
6. The location of the specimen on the frame should not be used to "infer" the loading or stress state in the specimen. In other words, specimens installed near the ends of the support frame should not be viewed as being near a bearing. All specimens should be viewed as being subjected to the same stress state.
7. Interior and exterior specimens should be treated the same.
8. Both faces of each specimen should be treated the same.
9. Any specimen could have any type of crack or even multiple cracks.

The inspection order was predetermined and navigated by Purdue research staff so that each inspector saw the same specimens in the same order. In total, 30 inspectors (27 males and 3 females) participated in the study. This study included 13 inspectors from INDOT, three inspectors from the Illinois Department of Transportation, 12 inspectors from private engineering and/or inspection firms, and 2 inspectors from federal agencies.

Among the 30 inspectors tested, there was very little variability in vision results. All inspectors were able to read the smallest font on the Jaeger test. The Pelli Robson contrast test had some slight variation, but it was not found to impact the detection results.

After completing the inspection, the inspectors were given an Exit Survey. This survey asked for information regarding their age, years of experience, which tools that brought and used and which training courses they had completed. Additional data, including the time spent inspecting each portion as well as weather data was collected for each test.


**Figure 3** Access by man-lift.

**Figure 4** Weathered specimens.

**Figure 5** Painted specimen with gusset plate.

**Figure 6** Welded cover plate specimen.

(a)

(b)

**Figure 7** Riveted cover plate specimen mounted horizontally (a) and vertically (b).

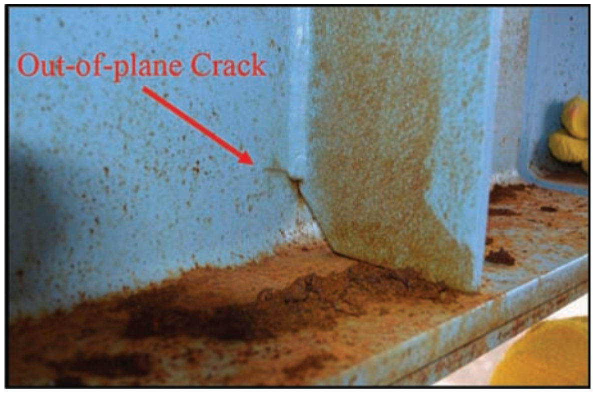


**Figure 8** Out-of-plane crack.


**Figure 9** Weld toe crack.

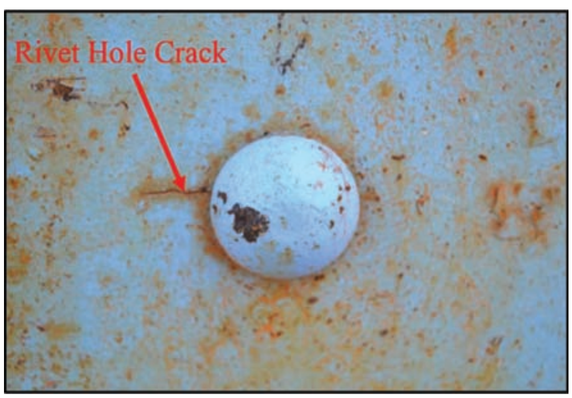


**Figure 10** Rivet hole crack.

TABLE 1
**Target Size and Distribution Summary**

| | Number of Hits per Size Range | | | |
|---|---|---|---|---|
| **Crack type** | **0.5″<1.5″** | **1.5″<2.5″** | **2.5″<3.5″** | **3.5″<5.5″** |
| **Out-of-plane** | 6 | 7 | 6 | 9 |
| **Cover plate weld toe** | 1 | 3 | 8 | 11 |
| **Rivet hole** | 17 | 2 | 0 | 0 |
| ***Total*** | *24* | *12* | *14* | *20* |

## RESULTS

### Routine Inspection (11 Participants)

The first inspection performed was designed to simulate a routine inspection of a highway bridge. The inspectors were permitted to walk beneath the structure. They were given an $11 \times 17$ sheet of paper with each specimen drawn on it for their comments. There was no imposed time limit. Seven participants used binoculars and one used a flood light.

The comments and notes were extremely varied. Three inspectors left general comments like "light to moderate corrosion on webs." Several inspectors used binoculars and carefully inspected each specimen. They noted cracks, possible cracks and provided dimensions. One inspector correctly identified five cracks. He/she also reported 22 false positives and spent over an hour on this inspection. Two inspectors were taking so long that this portion of the course was terminated to ensure that they would have time to complete the hands-on portion.

The average detection rate for fatigue cracks was 4%. However, 100% detection was not possible because some cracks physically could not be seen from the ground because they were obstructed by the bottom flange. After the first 11 inspectors, this inspection scenario was abandoned since inspectors could really only discern the general condition of the specimens, not identify fatigue cracks.

### Hands-On Inspection of Weathering Steel (11 Participants)

A small selection of 16 weathered W-shape specimens was included in the final course. There were 18 possible hits on these specimens. The average detection rate of the 11 inspectors was 11%. The average time for inspecting these 16 specimens was 41 minutes.

The hit/call rate was used to compare the number of hits and false positives an inspector had. A higher hit per call rate is desirable. A low rate means the inspector reported a lot of false positives. Obviously this is undesirable since it can lead to a waste of resources by spending more time investigating them and using other, more expensive, forms of NDE. The average hit/call rate was 28%.

Six of the 11 inspectors did not correctly identify any cracks. The most successful inspector found 7 of the 18 hits (39%) with a hit per call rate of 23%. Another notable observation regarding the detection of cracks on weathered specimens is that 20 of the total 21 hits were found with the use of a flashlight. The number of hits each inspector recorded, along with their corresponding false positives, can be seen in Figure 11.

After the first 11 inspectors, the inspections of the weathering steel specimens were discontinued because rust gradually formed across the cracks making them undetectable. In an actual bridge, regular cycling of the crack would maintain a "fresh" surface, but that is not possible with these static specimens.

### Hands-On Inspection of Painted Steel (30 Participants)

The participants inspected 72 faces of W-shape specimens (36 specimens), 16 nine-foot-long riveted cover plates, and 59 welded cover plate specimens, for a total of 147 painted specimens. The painted portion of the course included 70 possible hits. The results are summarized as follows:

- The average time to complete this scenario was just over 4 hours.
- The number of hits each inspector found along with their false positives can be seen in Figure 12. The most successful inspector found 60 out of 70 possible (86%). The worst performing inspector found only 22 of the 70 (31%). On average, the inspectors found 46 of the 70 cracks (65%), meaning they missed 24 cracks.
- The number of false calls made during the inspections ranged from 14 to 268. The average number of false calls was 90 with a standard deviation of 67. A false call was defined as a crack reported in a region of the specimen without a known defect. During the inspection, inspectors were instructed to use their best judgement to record only cracks, however, many inspectors recorded "possible or probable cracks" and noted the need for follow-up inspection or testing on their inspection forms. Since these supplementary efforts have economic implications, these indications were considered "false calls" when they did not correspond with known defects.
- Although the visibility of some of the cracks changed over time, the overall difficulty of the inspection seems to have remained relatively constant. Large differences in performance were recorded in inspections that occurred on consecutive days, indicating that the variability in results is likely due more to inspector characteristics and not changes in the conditions of the of the specimens.
- The best (or highest) hit/call rate was 75% (24BR-25), while the lowest was 13% (27PC-37), meaning 6.1 false calls for every crack correctly identified. More false calls were not an indicator of higher hits.

The variable of highest interest in terms of affecting detectability was crack length. Figure 13 shows the detection rates by crack size, broken into 1/2-inch size ranges. The largest crack range, between 5 and 5-1/2 inches, had a detection rate of 91%. The smallest crack grouping, between 1/2 and 1 inch, had a 46% detection rate. These show the average detection rates for each range.

Three cracks on the fixture were not found by any of the 30 inspectors. The largest undetected crack was 3-1/4 inches which was on the weld toe of a tapered cover plate specimen. The other two cracks missed were located at the ends of longitudinal stiffeners and were 3/4 and 1-3/4 inches in length. The smallest crack, measuring 1/2 inch, which was on at a rivet hole, was detected by 23 of the 30 inspectors. To complete the statistical POD analysis and generate the POD vs. "a" (i.e., crack length) curves, Military Handbook 1823a (DoD, 2009) and the accompanying software was used. There were 2,100 data entries (70 possible hits for each of the 30 inspectors). The $a_{50}$ value gives the crack length that corresponds to a 50% probability of detection for this data set. Likewise, $a_{90}$ provides the crack length for 90% POD.

Figure 14 shows the curve generated by the POD software. The y-axis is the probability of detection and the x-axis shows the crack sizes. Hits are shown at 1.0 at the top, and misses are shown at the bottom at 0. Cracks of all sizes were both found and missed.

For all inspectors and all specimens, the $a_{50}$ crack length is 1 inch and the $a_{90}$ crack length is 5-1/2 inches. Due to the variability in inspector performance, the 95% confidence limits could not be applied to the $a_{90}$ crack length. Figure 15 shows the individual inspectors' POD curves. For example, there were actually three inspectors which did not show an increase in likelihood of detection for longer cracks. After removing the results from these inspectors whose detection did not conform to the assumptions of the POD model (i.e., POD increasing with increasing crack length), the probability of detection curve was regenerated. For the 27 remaining inspectors and all the specimens, the $a_{50}$ crack length remains 1 inch, but the $a_{90}$ crack size is reduced to 4-1/2 inches.

Another method of statistical analysis, using a random parameters binary logit model, was used to determine how other variables interacted with crack length to affect the probability of detection.

After testing many combinations of variables, a group of three was found to significantly impact detection rates. Those were crack length, crack type, the number of years of inspection experience the inspector had, the inspection duration, and the elapsed time since the first inspection.

Marginal effects are used to show how each parameter affects the dependent variable, POD in this case. The marginal effect gives the change in probability of detection for a unit change in the independent variable. The marginal effects for the relevant parameters can be seen in Table 2. A larger marginal effect indicates a greater influence on the likelihood of detecting the crack while a smaller marginal effect indicates a lesser influence. For the dummy variables, marginal effects are computed as the difference in the estimated probabilities when the variable is changed from zero to one, while all the other variables are set equal to their means. For continuous variables, the marginal effects are computed from the partial derivative of the probability equation. In both cases, the reported marginal effect represents the average over all the observations.

The marginal effects of the six independent variables included in the binary logit model are shown in Table 2. Every additional inch of crack length, the POD increases by 16%. For every additional year of inspection experience, the POD decreases by 0.2%. This model shows that both out-of-plane distortion induced cracks and cover plate weld toe cracks, regardless of length, are approximately 16% less likely to be detected than rivet hole cracks. A one-minute increase in inspection time increases the likelihood of detecting a crack, regardless of length, by 0.06%. Finally, the likelihood of detecting each crack decreased approximately 8% from the first inspection to the last inspection.

In addition to recording the location of detected cracks, inspectors were asked to record the length of the cracks on their inspection forms. However, some inspectors were more disciplined about providing this measurement than others. Additionally, individuals took different approaches to determining these measurements with some carefully measuring each crack, and others visually estimating the length without a measuring scale. Most inspectors used some combination of the two strategies.

Figure 16 shows the crack length data for the girder specimens. The actual length of the crack is shown on the horizontal axis and the vertical axis displays the measured value reported by the inspectors. The diagonal 1:1 reference line represents exact agreement between the actual length and the measured length. For the majority of the cracks in the girder specimens, the average of the measured lengths plots below the 1:1 line indicating that the inspectors tended to underestimate the length of the crack. The average absolute error increased with crack size and the percent absolute error remained constant with crack size. The average length error was -0.37 inches and the standard deviation was 1.27 inches.

Figure 17 and Figure 18 present the crack length data for the welded cover plate and riveted plate specimens, respectively. In contrast to the girder specimens, the average of the measured lengths of these cracks is generally above the 1:1 line indicating that the inspectors had a tendency to overestimate the length. The average absolute error increased with crack size and the percent absolute error decreased with crack size. For the welded cover plates, the average length error was 0.51 inches and the standard deviation was 1.2 inches. For the riveted members, the average length error was 0.14 inches and the standard deviation was 0.32 inches.

A univariate analysis was performed to identify relationships between detection rates and other single factors including tools used, time spent inspecting, wind speed, and years of experience. Strong trends were not found, indicating that there is no single factor that greatly impacts detection. Interestingly, a slight *negative* trend between years of inspection experience and detection rates was seen. Figure 19 shows detection rates and years of experience, Figure 20 shows detection rate and average air temperature, Figure 21 shows the detection rate and inspection duration, and Figure 22 shows the detection rate and the number of training courses (out of the eight courses listed on the exit survey). No single variable showed a correlation with the number of false positives. Neither temperature, experience, training, nor duration showed more than a very slight trend with the number of false calls made by the inspector.

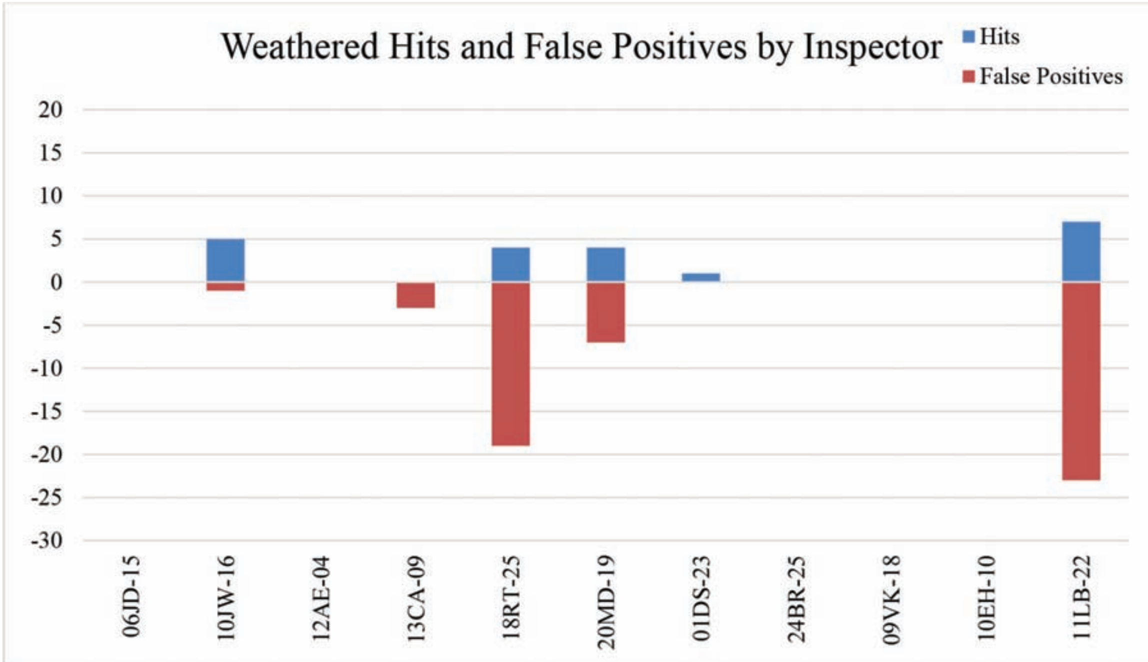


**Figure 11** Weathered hits and false positives.

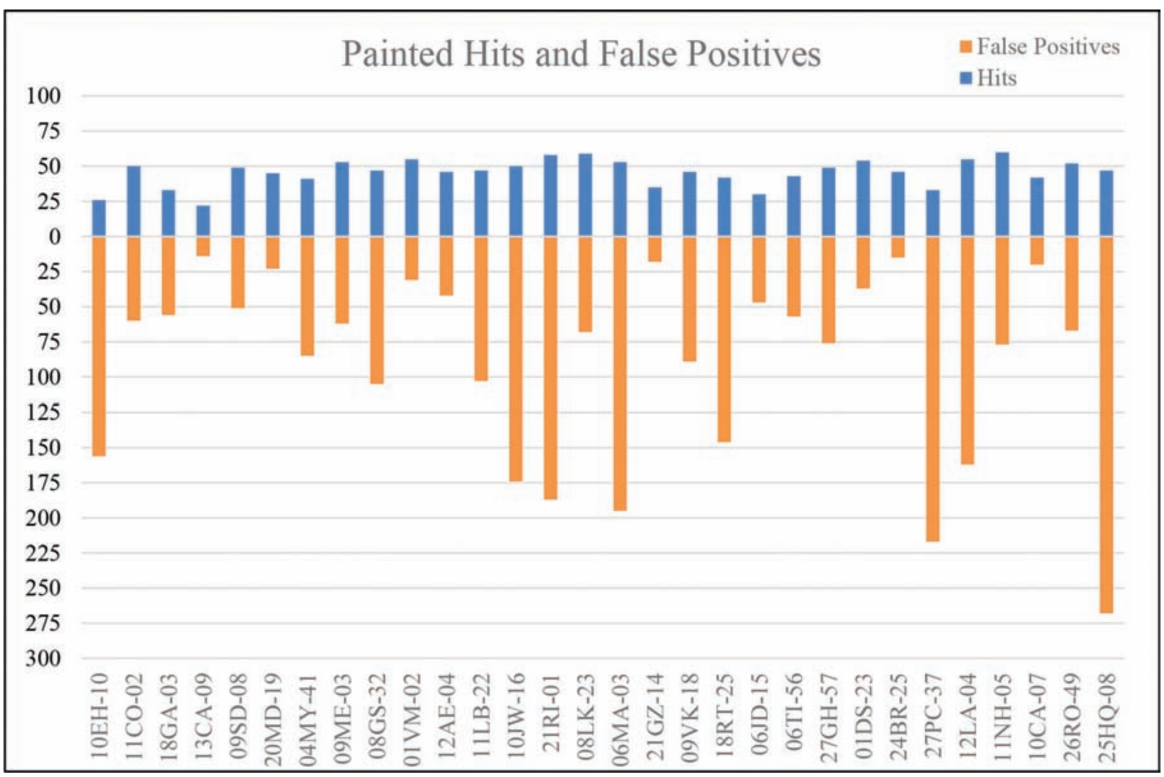


**Figure 12** Painted hits and false positives.

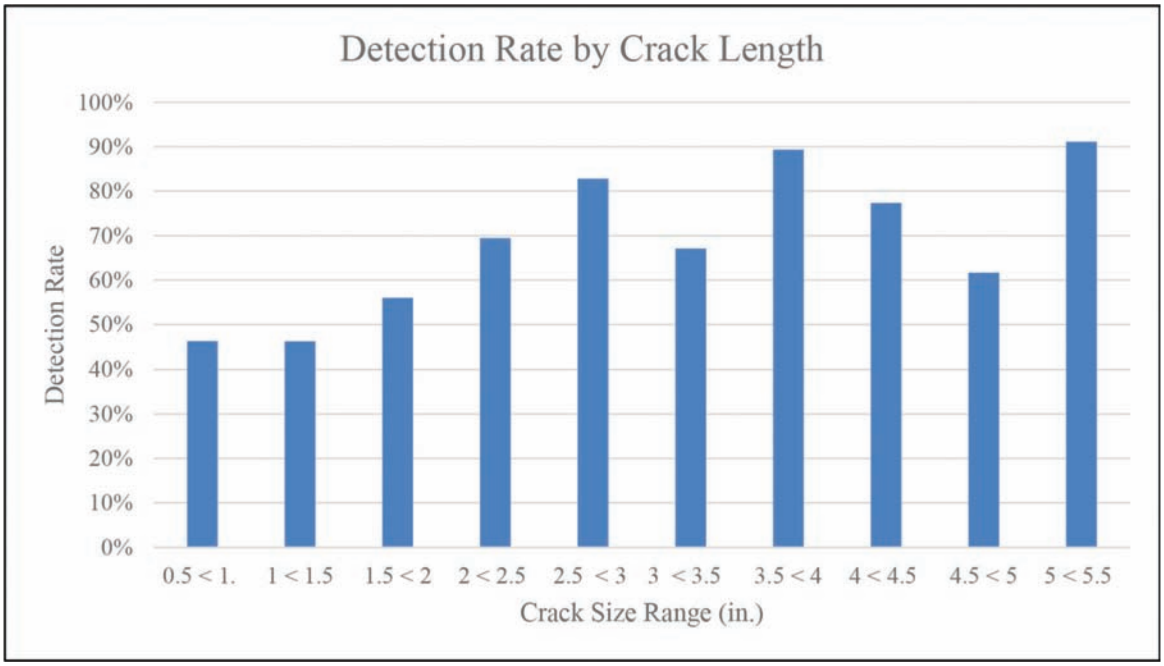


**Figure 13** Detection rate by crack length.

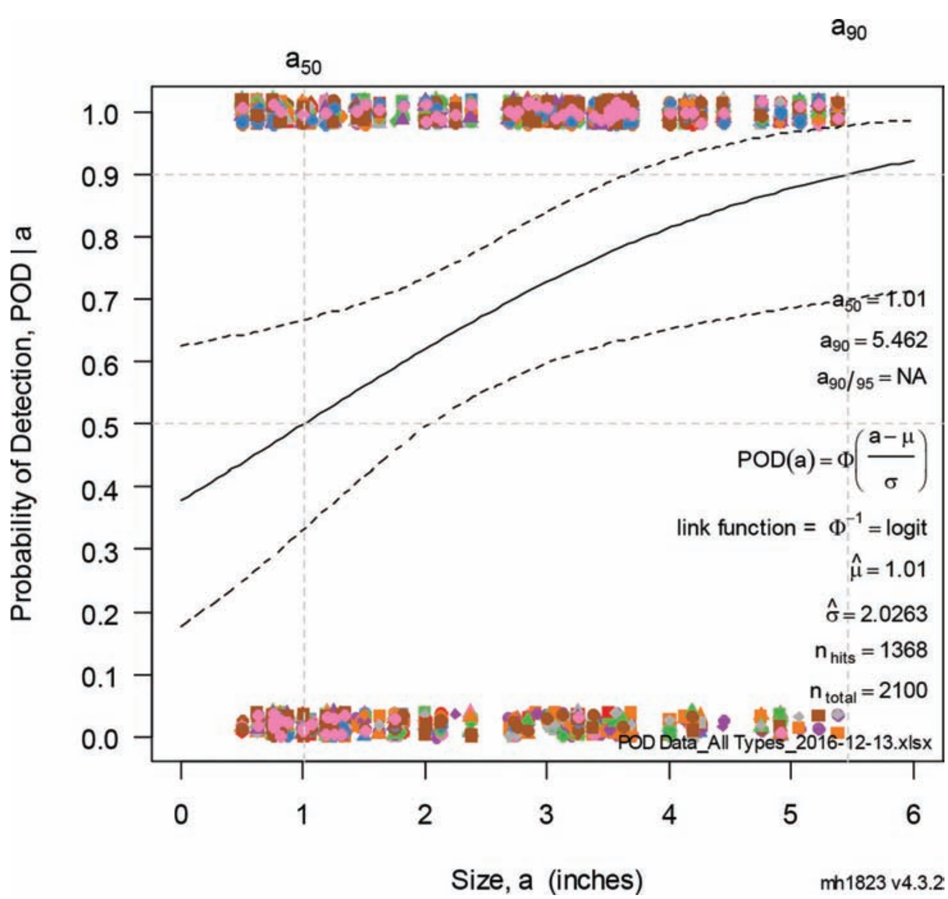


**Figure 14** Total data set POD vs. crack length.

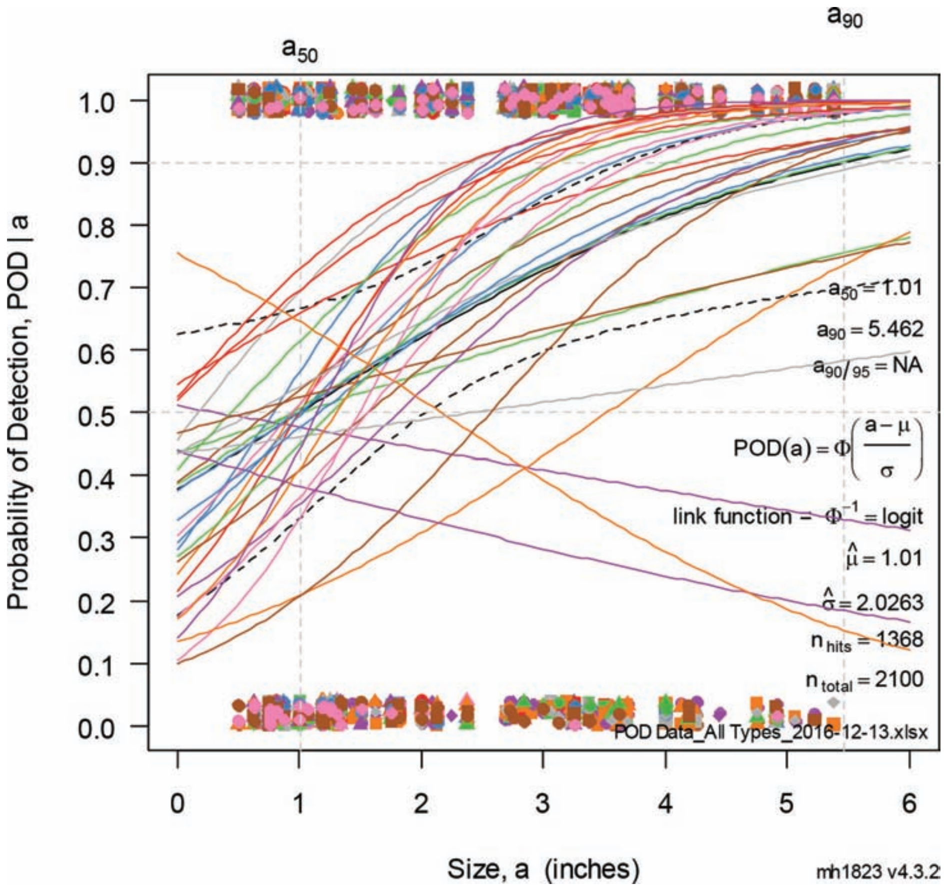


**Figure 15** Total data set individual POD vs. crack length curves.

TABLE 2
**Marginal Effects**

| Variable Description | Avg. Marginal Effect * (Std. Dev.) |
|---|---|
| Crack length (inches) | 0.156 (0.087) |
| Out-of-plane crack (1 if yes, 0 if not) | -0.159 (0.024) |
| Cover plate weld toe crack (1 if yes, 0 if not) | -0.153 (0.031) |
| Inspection duration | 6.41E-4 (2.73E-4) |
| Years of inspection experience | -0.002 (0.002) |
| Elapsed time since first inspection (days) | -1.11E-4 (4.72E-5) |

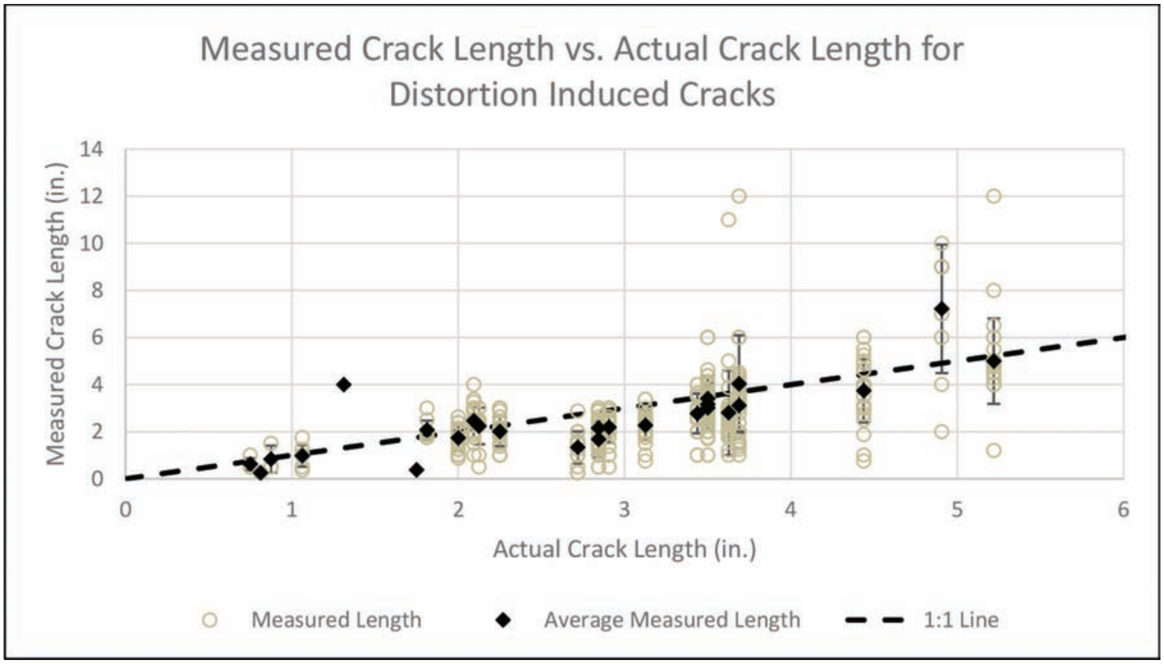


**Figure 16** Measured crack length versus actual crack length for out-of-plane distortion induced cracks in the W-shape specimens.

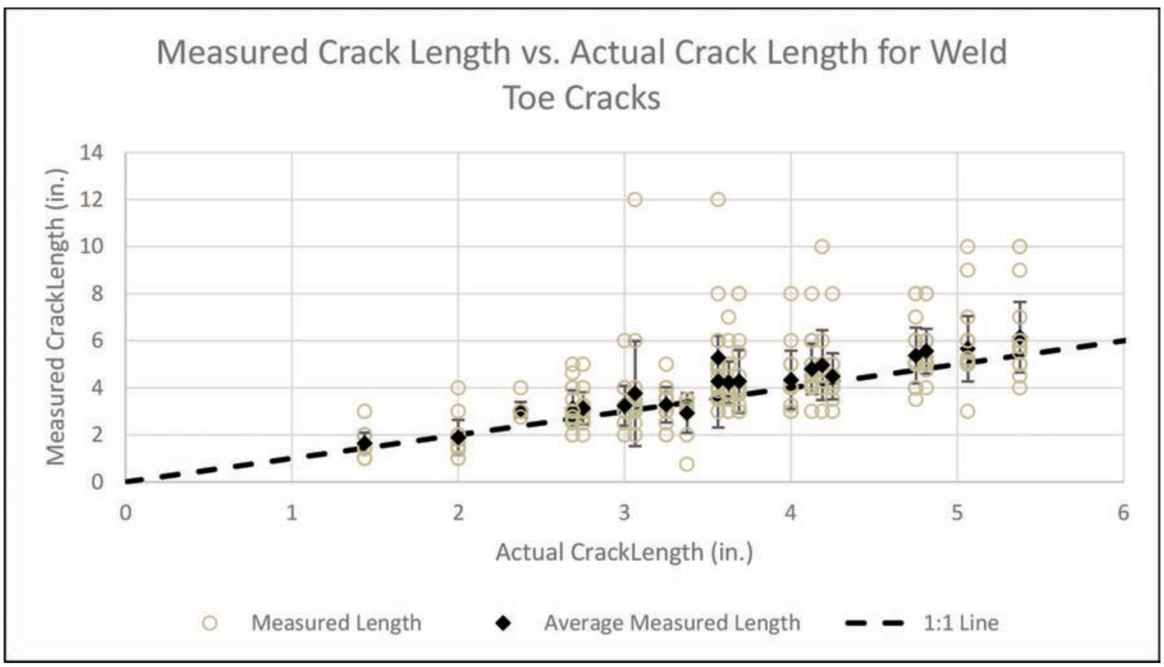


**Figure 17** Measured crack length versus actual crack length for weld toe cracks in welded cover plate specimens.

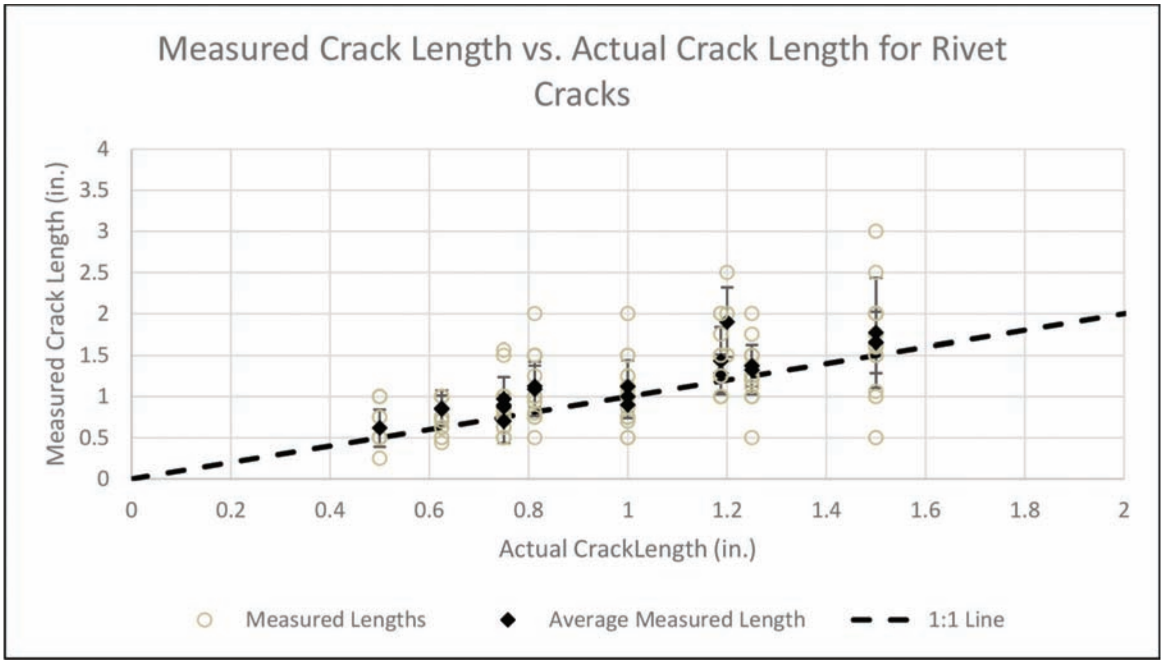


**Figure 18** Measured crack length versus actual crack length for cracks in the riveted plate specimens.


**Figure 19** Total detection rate vs. years of inspection experience.

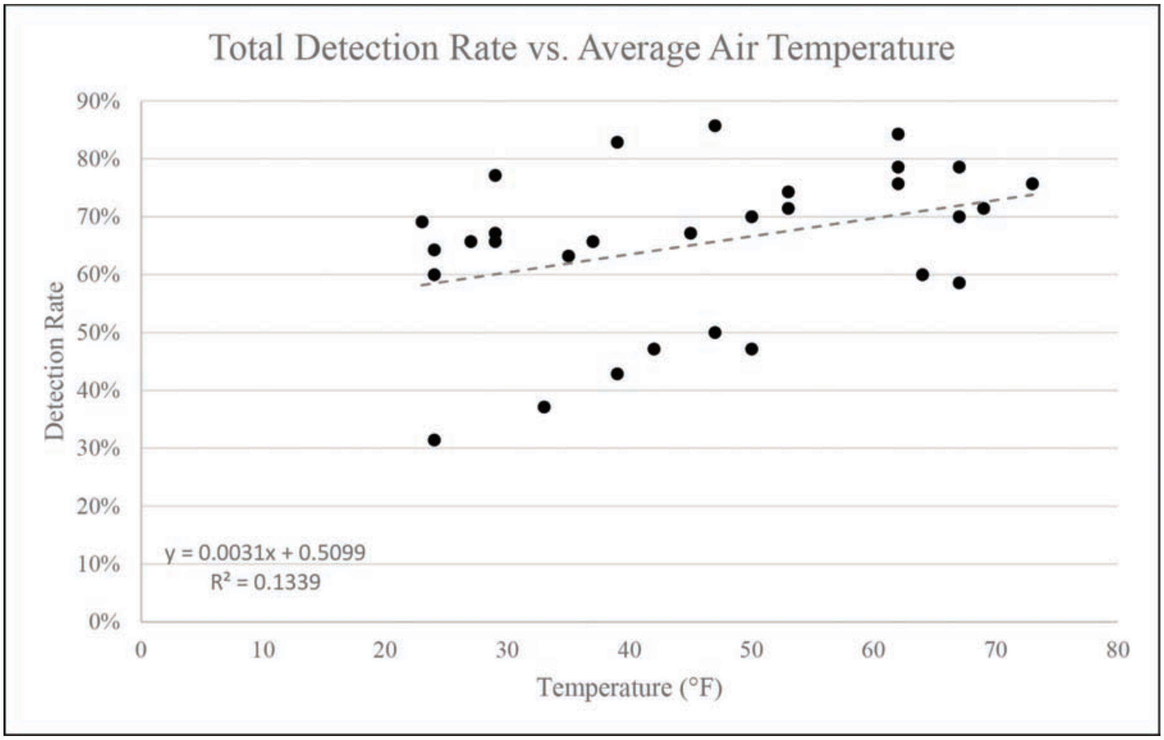


**Figure 20** Total detection rate vs. average air temperature.


**Figure 21** Total detection rate vs. inspection duration.

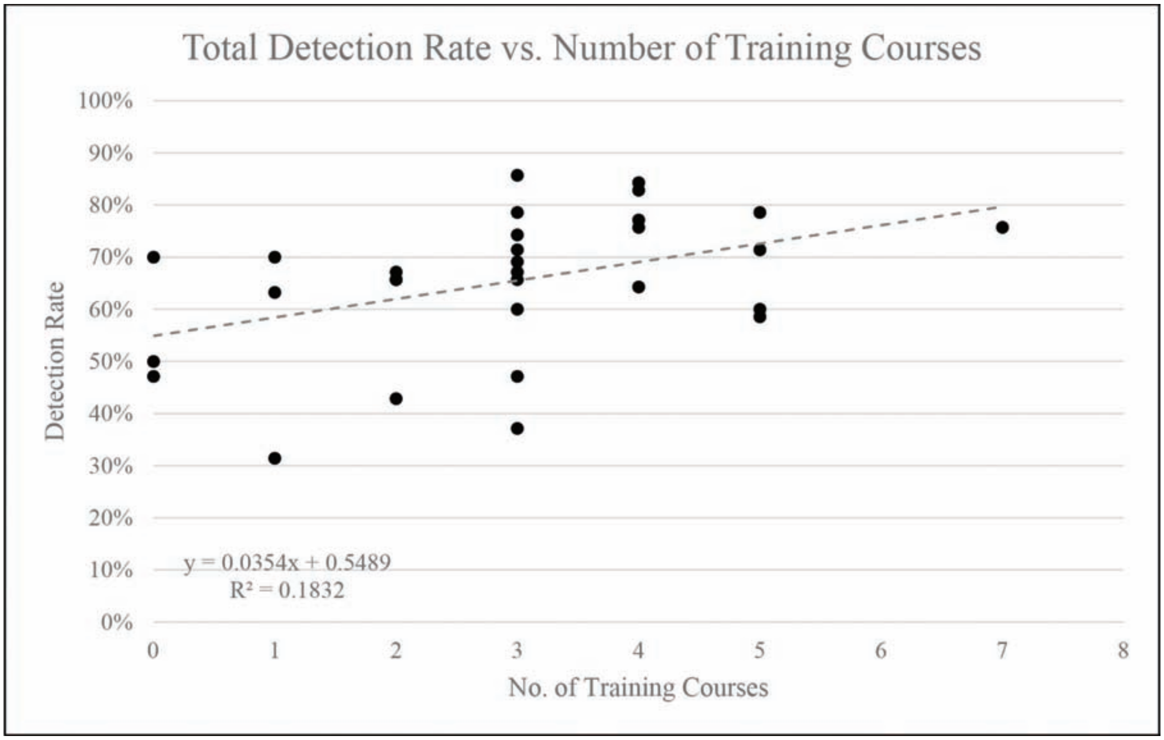


**Figure 22** Total detection rate vs. number of training courses.

## CONCLUSIONS

Variations in the reporting techniques for both routine and hands-on inspection of steel bridge superstructures were observed. While the documentation styles and level of detail may have been impacted by the study procedures provided or the fact that they were being observed in a test setting, it is clear that there is no consistent standard for identifying what information should and should not be recorded during a visual inspection.

Inspectors did not bring or use the same equipment to conduct hands-on visual inspections. Although no relationship was found between inspection performance and tool use, this is thought to be due to the lack of information collected during the study, and not because inspection tools have no effect on performance.

The routine inspection scenario produced highly varied inspection reports. The inspectors' documentation ranged from detailed cracks with dimensions, to unlabeled tick marks, to general comments on the overall condition of the bridge. Time spent completing the ground scenario ranged from 12 minutes to 82 minutes.

The hands-on inspection scenario for the weathered specimens resulted in an average detection rate of 11%. Of the 198 observations, only 21 hits were identified. Six of the 11 inspectors did not find any cracks.

The average detection rate for all 30 inspectors of all 70 cracks located on the painted specimens was 65%. Detection rates ranged from 31% to 86%. Univariate analysis between detection rates and other factors, like inspection time or the day's weather, revealed slight, but statistically significant, correlations. Detection rate increased with increasing inspection duration, temperature, and training, but decreased with increasing experience

For the majority of the inspectors, the likelihood of a crack being detected increased with crack length. However, this was not true for three of the participants.

The average number of false calls for all 30 inspectors on the 147 painted specimens was 90. The number of false calls ranged from 14 to 268. Univariate analysis between the number of false calls and other factors, like inspection time or the day's weather, did not reveal any significant correlations. However, the multivariate analysis revealed that the number of false calls was related to the inspector's employer, the maximum wind speed on the day of the inspection, the use of a tape measure, and completion of the Element Level Bridge Inspection training course. Inspectors that used a tape measure made fewer false calls while inspectors that were employed by a private inspection/engineering firm, experienced higher wind speeds, and attended the element level inspection course tended to make more false calls.

Probability of detection analysis using a log-odds model generated POD values for varying crack lengths. For the total population of cracks for this set of inspectors, a 1-inch-long crack had a 50% chance of being detected and a 5-1/2-inch-long crack had a 90% chance of being detected. The data was too scattered to assign a 95% confidence bound.

A random parameters binary logit model was used to identify variables beyond crack length that affected the probability that a crack would be detected. The analysis showed that crack length, the type of crack, the inspector's years of experience, the inspection duration, and the elapsed time since the first inspection significantly impacted the likelihood of a crack being detected.

## RECOMMENDATIONS

Recommendations for improving the reliability of visual inspection on steel bridges have been grouped into three categories: equipment, training, and procedures.

It is recommended that a standard set of equipment be provided to each inspector to ensure consistency among inspectors. Items for the standard tool set should include, but are not limited to a light-emitting diode (LED) or halogen flashlight with light output of at least 100 lumens and adjustable focus, wire brush, scraper, hammer, 5 × power and lighted 10 × power magnifying glasses, telescoping inspection mirror, easy to use measuring device, and 1/4-inch square rod for measuring web gaps. Requiring each inspector to bring and use the provided standard set of tools could better equip them for detecting defects and encourage them to follow proper inspection procedures. Inspectors should be provided with basic training outlining the proper use for each tool.

Providing detailed procedures for hands-on visual inspection of steel bridges may improve the method's reliability. The procedures should clearly state a process for inspecting different types of details, including using the prescribed tools to closely inspect areas prone to defects. The procedures should use the same terminology and descriptions that are used in the training courses. Requiring that the correct tools be used to closely inspect details prone to fatigue cracks may increase the likelihood of cracks being found. It is also recommended that an equipment checklist be provided for each inspection. The inspectors should indicate on the checklist that each tool was available and in working order for the inspection.

Additional procedural recommendations are offered based on visual inspection research, although their effectiveness for hands-on bridge inspections was not explicitly evaluated in this study.

- Encourage active observation.
- Hold regular calibration meetings and refresher training.
- Provide regular feedback.
- Rotate inspectors.
- Allow adequate time to complete each inspection and encourage inspectors to use the allotted time.

The large variability in inspection performance indicates that the current training program produces inspectors with differing inspection abilities. Although the content of the existing courses appears to be adequate, it seems that some inspectors are struggling to apply the lessons learned in the classroom in the field. Both a new half-day training course and a new training module were developed to address this deficiency. These trainings focus on the physical and mental factors of visual inspection and aim to teach inspectors "how to inspect" rather than "where to inspect" or "what to inspect." In order to evaluate the effectiveness of the new training, a small number of inspectors that participate in the trainings will be invited to complete the POD inspection using the same procedures as the benchmark inspectors.

The training module focuses on the physical tools and techniques of visual inspection and has been incorporated into the S-BRITE course *Inspecting Steel Bridges for Fatigue*. In speaking to the inspectors who participated in this study, it became clear that most have had very little exposure to real fatigue cracks during their inspection careers. After a brief introduction to the tools and techniques inspection, the module includes a practical component in which the inspectors are given the opportunity to see and touch specimens with true fatigue cracks.

The training course focuses on the cognitive processes used during inspection to improve observation and interpretation skills. The course subdivides the inspection activity into four distinct tasks (prepare, search, decide, and document) and then identifies the observation skills used in each task (perception and recognition, attention, memory, mental imaging and mental models, and judgement and decision-making). The course covers both the theory and application of these skills and offers techniques for improvement. This is intended to be a half-day standalone course offered through the S-BRITE Center.

Due to the large variability in results and the relatively weak predictive power of any of the variables expected to correlate with performance, it is recommended that performance testing be implemented. After completing the required classroom based and hands-on training courses, the inspectors should be required to pass a practical inspection test.

Four models of performance criteria have been outlined below for the practical inspection test. Due to the limited amount of available data, and the high variability of the data from this study, it is not yet possible to establish set requirements. The data collected from this inspector population was subjected to the proposed criteria and the outcomes are presented for discussion.

One option is to establish a flat detection rate for passing. For example, the inspectors would each be required to find 70% of the total cracks present on the course. This option is simple and easy to determine if an inspector meets the criteria. However, the flat rate does not take into account any differences between crack sizes. There is no weighting to place more importance on finding more critical cracks. Further, the question that really should be asked is how many cracks is it acceptable to miss.

A second option is to use a graduated scale for detection rates. For instance, to pass the practical exam, an inspector must find 50% of cracks shorter than 1 inch, 65% of cracks with lengths between 1 and 3 inches, and 80% of cracks greater than 3 inches long. This method stresses and presumes that inspectors should be able to find longer cracks. Again, criticality of the actual crack is not directly included in the test. For example, a 1-inch crack in a butt weld in a tension flange of a fracture critical girder is more critical than a 3-inch out-of-plane distortion crack in a web of girder in a multi-beam bridge.

A third option is to set passing criteria based on crack types. Setting the criteria based on the type of crack, and how detrimental the particular crack type is for the structure, promotes a more reasonable approach to bridge inspection. For example, load induced cracks emanating from butt welds grow perpendicular to the stress, and can quickly lead to member fracture. For this crack type, the criteria for finding cracks less than 1-inch-long may be set to be 70% of the possible cracks. For distortion induced cracks that typically grow more slowly and pose less risk, a lower acceptable rate, possibly 50%, may be acceptable.

A fourth option is to evaluate inspector performance based on both detection rate and the number of false calls. Large numbers of false positives result in higher costs (longer inspection times and additional equipment being used to further investigate the suspected, but "unreal" cracks) and can negatively affect a bridge inspection program. Inspectors could be rated based on both their detection rate and the number of false calls, with a minimum rating necessary for passing.

## REFERENCE

DoD. (2009, April 7). *Nondestructive evaluation system reliability assessment* (MIL-HDBK-1823A). U.S. Department of Defense. http://www.statisticalengineering.com/mh1823/MIL-HDBK-1823A(2009).pdf

## About the Joint Transportation Research Program (JTRP)

On March 11, 1937, the Indiana Legislature passed an act which authorized the Indiana State Highway Commission to cooperate with and assist Purdue University in developing the best methods of improving and maintaining the highways of the state and the respective counties thereof. That collaborative effort was called the Joint Highway Research Project (JHRP). In 1997 the collaborative venture was renamed as the Joint Transportation Research Program (JTRP) to reflect the state and national efforts to integrate the management and operation of various transportation modes.

The first studies of JHRP were concerned with Test Road No. 1—evaluation of the weathering characteristics of stabilized materials. After World War II, the JHRP program grew substantially and was regularly producing technical reports. Over 1,600 technical reports are now available, published as part of the JHRP and subsequently JTRP collaborative venture between Purdue University and what is now the Indiana Department of Transportation.

Free online access to all reports is provided through a unique collaboration between JTRP and Purdue Libraries. These are available at http://docs.lib.purdue.edu/jtrp.

Further information about JTRP and its current research program is available at http://www.purdue.edu/jtrp.

## About This Report

An open access version of this publication is available online. See the URL in the citation below.

**Recommended Citation**

Campbell, L. E., Snyder, L. R., Whitehead, J. M., Connor, R. J., & Lloyd, J. B. (2019). *Probability of detection study for visual inspection of steel bridge: Volume I—Executive summary* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2019/21). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284317103