MOUNTAIN-PLAINS CONSORTIUM

MPC 15-288 | A. Hesse, R. Atadero, and M. Ozbek

Using Expert Opinion to Quantify Accuracy and Reliability of Nondestructive Evaluation on Bridges

MPC
MOUNTAIN-PLAINS CONSORTIUM

A University Transportation Center sponsored by the U.S. Department of Transportation serving the Mountain-Plains Region. Consortium members:

Colorado State University
North Dakota State University
South Dakota State University
University of Colorado Denver
University of Denver
University of Utah
Utah State University
University of Wyoming

Using Expert Opinion to Quantify Accuracy and Reliability of Nondestructive Evaluation of Bridges

Alex Hesse
Rebecca Atadero
Mehmet Ozbek

Colorado State University

August 2015

ACKNOWLEDGEMENTS

A special thanks to the Mountain-Plains Consortium, a USDOT University Transportation Center for the funding of this project. Also, thanks to the state DOT and private company respondents to the surveys conducted for this study. Your opinions and knowledge were invaluable.

ABSTRACT

Bridge inspection is an important phase in bridge management. In 2009, a joint American Society of Civil Engineers Structural Engineering Institute and AASHTO Ad-Hoc group was created to identify the issues in bridge safety and to study current bridge practices. This group recommended: "A more rational, risk-based approach to determining the appropriate inspection intervals for bridges is needed, as opposed to a set twenty-four month cycle for all bridges." The committee also recommended a wider use of NDE methods. The difficulty in increasing the use of these NDE methods is the increased costs and time spent. One way to deal with this is to implement risk-based planning. Although to do this, the accuracy, reliability, bias, and cost of each test must be quantified. This study attempts to quantify these parameters for common bridge NDE methods. This was done through two methods. First, a literature review was performed to determine common NDE methods and data were found for these methods. Second, a Delphi method survey was conducted to develop a broader range of data that matches real life practices. All of the data were then analyzed and conclusions were drawn to quantify the various parameters for these NDE methods.

TABLE OF CONTENTS

1. INTRODUCTION 1

1.1 Current Bridge Inspection Practice 1
1.2 Research Objectives 2
1.3 Research Methodology 2
1.4 Thesis Organization 3

2. BACKGROUND AND LITERATURE REVIEW 4

2.1 Introduction 4
2.2 Commonly Used NDE Methods and Current Research 4
2.2.1 Visual Inspection 4
2.2.2 Acoustic Emission 5
2.2.3 Cover Meter/ Pachometer 6
2.2.4 Electrical Potential 7
2.2.5 Impact Echo 7
2.2.6 Liquid Penetrant 8
2.2.7 Magnetic Particle 8
2.2.8 Mechanical Sounding 9
2.2.9 Radar 9
2.2.10 Radiography 9
2.2.11 Rebound Hammer 10
2.2.12 Thermal Imaging 10
2.2.13 Ultrasonic 10
2.3 Comparative Studies for Accuracy, Reliability, and Bias of NDE Methods 11
2.3.1 Comparing Radar, Chain Drag (Mechanical Sounding) and Electrical Potential – A Barnes and Trottier 2000 Study 11
2.3.2 Comparing Chain Drag (Mechanical Sounding) and Impact Echo to Concrete Core Results – A Scott et al. 2003 Study 12
2.3.3 Comparing Coin-Tap Test (Mechanical Sounding) to Infrared Thermography – A Clark et al. 2003 Study 12
2.3.4 Comparing Various NDE Methods on Deteriorating Concrete Bridges – A Rens et al. 2005 Report 12
2.3.5 Comparing Ultrasonic, Hammer Sounding (Mechanical Sounding), Surface Hardness (Rebound Hammer), and Radar – A Wood and Rens 2006 Case Study 14
2.3.6 Comparing Impact Echo, Ground Penetrating Radar, and Infrared Thermography – A Yehia et al. 2007 Study 15
2.3.7 Comprehensive NDE Concrete Bridge Deck 2013 Study by the Transportation Research Board 16
2.4 Comparative Studies for Various Costs of NDE Methods 18
2.4.1 Comparing Various NDE Methods on Deteriorating Concrete Bridges – A Rens et al. 2005 Report 18
2.4.2 Comprehensive NDE Concrete Bridge Deck 2013 Study by the Transportation Research Board 18
2.5 Previous Surveys 19
2.5.1 Rens et al. 1993 Survey 20
2.5.2 CALTRANS 1994 Survey 20
2.5.3 Rens and Transue 1996 Follow-Up Survey 21
2.5.4 FHWA 2001 Survey 21

2.6 The Delphi Method ... 22
2.6.1 Background of the Delphi Method ... 22
2.6.2 The Delphi Method Process ... 23
2.6.2.1 Round One of the Delphi Method ... 24
2.6.2.2 Round Two of the Delphi Method ... 24
2.6.2.3 Round Three and Subsequent Rounds of the Delphi Method ... 24
2.6.2.4 Determining Participants ... 24
2.6.2.5 Determining Size of a Delphi Method Study ... 24
2.6.2.6 Determining When Consensus and Stability is Reached ... 25
2.6.3 Previously Conducted Delphi Method Surveys in Civil Engineering ... 25

3. IMPLEMENTING THE DELPHI TECHNIQUE AS A SURVEY METHOD TO GATHER EXPERT OPINION ... 26
3.1 Selection of the Delphi Method ... 26
3.2 Determining Participants ... 26
3.3 Round One Questionnaire ... 26
3.3.1 Section One of the First Questionnaire ... 27
3.3.2 Section Two of the First Questionnaire ... 27
3.3.3 Section Three of the First Questionnaire ... 27
3.3.4 Questionnaire One Response Rate ... 28
3.4 Round Two Questionnaire ... 28
3.4.1 Subsection One of the Second Questionnaire: Bias ... 28
3.4.2 Subsection Two of the Second Questionnaire: Accuracy ... 29
3.4.3 Subsections Three and Four of the Second Questionnaire: Precision and Reliability ... 29
3.4.4 Subsection Five of the Second Questionnaire: Costs ... 30
3.4.5 Questionnaire Two Response Rate ... 30
3.5 Round Three Questionnaire ... 30
3.5.1 Questionnaire Three Response Rate ... 31
3.6 Round Four Questionnaire ... 31
3.6.1 Questionnaire Four Response Rate ... 31

4. RESULTS AND DISCUSSION ... 32
4.1 Introduction ... 32
4.2 Round One Questionnaire ... 32
4.2.1 Certification and Experience Level ... 32
4.2.2 Commonly Used NDE Methods ... 32
4.3 Determining Convergence and Stability ... 33
4.3.1 Using Spearman's Rank Correlation Coefficient Method to Calculate Stability ... 34
4.4 Results for the Bias of NDE Methods ... 34
4.5 Results for the Reliability of NDE Methods ... 36
4.6 Results for the Accuracy of NDE Methods ... 38
4.7 Results for the Costs Subsection of the Survey ... 40
4.8 Conclusion for the Results of the Surveys ... 45

5. CONCLUSIONS AND RECOMMENDATIONS ... 47
5.1 Summary ... 47
5.2 Conclusions ... 47
5.3 Suggestions for Future Work ... 48

BIBLIOGRAPHY ... 49

APPENDIX A: EXAMPLES OF THE SURVEYS IMPLEMENTED 53
APPENDIX B: STABILITY RESULTS.. 76

LIST OF TABLES

Table 2.1 Correlation Between the Stress Intensity Factor Range, ΔK, and Crack Intensity for Steel Bridges (Gong et al. 1992) 6

Table 2.2 Summary of Barnes and Trottier Study Results (2000) 11

Table 2.3 Summary of CALTRANS (Moore et al. 2001) NDT Survey Question 1: NDT Methods Currently in Use and Rens et al. (1997) NDE Survey Question: Domestic NDE Methods Currently in Use 21

Table 2.4 Summary of NDE Techniques used on Steel Structures from 2001 FHWA Survey (Moore et al 2001) 22

Table 2.5 Summary of NDE Techniques used on Concrete Structures from 2001 FHWA Survey (Moore et al. 2001) 22

Table 2.6 Selected Ranges of Coefficient of Variation Used to Determine Agreement (English and Kernan 1976) 25

Table 4.1 Number of People Indicating Experience with Each Method for Steel Bridge Members 33

Table 4.2 Number of People Indicating Experience with Each Method for Concrete Bridge Members 33

Table 4.3 Concrete and Steel NDE Methods Response Statistics for the Bias Subsection 35

Table 4.4 Concrete and Steel NDE Methods Response Statistics for the Reliability Subsection 36

Table 4.5 Concrete and Steel NDE Method Bias and COVs Indicated by Participant Responses 38

Table 4.6 Coefficient of Variation Results from the Accuracy Subsection from Round 3 and Round 4 39

Table 4.7 Stability Results from the Accuracy Subsection from Round 3 and Round 4 39

Table 4.8 Concrete and Steel NDE Method Normalized Accuracy and Standard Deviation Indicated by Participant Responses 40

Table 4.9 Concrete and Steel NDE Method Time to Run a Test Median, Standard Deviation, COV, and Correlated Range Indicated by Participant Responses 41

Table 4.10 Concrete and Steel NDE Method Time to Analyze Data Median, Standard Deviation, COV, and Correlated Range Indicated by Participant Responses 42

Table 4.11 Concrete and Steel NDE Method Time to Train an Inspector Median, Standard Deviation, COV, and Correlated Range Indicated by Participant Responses 43

Table 4.12 Concrete and Steel NDE Method Number of Inspectors Needed Median, Standard Deviation, COV, and Correlated Range Indicated by Participant Responses 44

Table 4.13 Concrete and Steel NDE Method Monetary Cost for Equipment Median, Standard Deviation, COV, and Correlated Range Indicated by Participant Responses 45

Table 4.14 Comprehensive Comparison by Ranking of Each NDE Method Indicated by Participant Responses 46

LIST OF FIGURES

Figure 2.1 Percentage of Inspectors to Identify Deficiencies During an In-Depth Inspection of two FHWA NDEVC Test Bridges (Phares et al. 2001) ... 5

Figure 2.2 Schematic of the Impact Echo Method Being Performed on a Damaged and Undamaged Concrete Deck ... 8

Figure 2.3 Summary of Rens et al. (2005) Concrete NDE Method Comparisons ... 13

Figure 2.4 Summary of Wood and Rens (2006) Case Study Results for NDE Tests Methods on the Lawrence Street Bridge ... 14

Figure 2.5 Percentage of Radar Tests to Identify Deficiencies in Concrete Specimens (Yehia et al. 2007) ... 15

Figure 2.6 Percentage of Infrared Tests to Identify Deficiencies in Concrete Specimens (Yehia et al. 2007) ... 16

Figure 2.7 Statistical Performance Grades of Each NDE Method Performed by Gucunski et al. 2013 . 17

Figure 2.8 Cost Performance Grades of Each NDE Method Performed by Gucunski et al. 2013 ... 19

Figure 3.1 Representation of the Scale Used for the Bias Subsection. ... 29

Figure 3.2 Representation of the Scale Used for the Precision and Reliability Subsections. ... 30

Figure 3.3 Example of a Personal Scale Made for Time Spent Running a Test ... 30

EXECUTIVE SUMMARY

The majority of the bridges in the United States were built during two periods. The first period of construction was in the 1930s during the Great Depression, and the second period was during the 1950s and 1960s with the implementation of the Federal-Aid Highway Act of 1956 (Rens et al. 2005). Ramey et al. (1997) shows that the American Association of State Highway and Transportation Officials (AASHTO) *Standard Specifications for Highway Bridges* implies a bridge design lifespan of 50 years, meaning the majority of the bridges in the United States have reached the end of their expected service life. Combining this with a limited maintenance and repair budget makes efficient bridge management essential for improving the safety and serviceability of the current bridge system in the United States. There are three main elements to the management of bridges and other types of infrastructure: inspection of the system, decision-making about maintenance and repair of the system, and the performance of maintenance and repair on the system. This research project considers the inspection phase of the management process for bridges, specifically nondestructive evaluation.

With recent bridge failures, specifically the I-35W bridge collapse in Minneapolis, bridge inspection practice in the United States has received much closer scrutiny. In response to this scrutiny, a joint American Society of Civil Engineers Structural Engineering Institute (ASCE/SEI) and AASHTO Ad-Hoc group was created to identify the issues to guarantee bridge safety and to study how current bridge inspection practices could be improved for the future (ASCE/SEI-AASHTO, 2009).

The ASCE/SEI-AASHTO ad-hoc group wrote a recent document entitled "White Paper on Bridge Inspection and Rating." In this paper, the group developed a listing of several deficiencies in existing inspection practice. One of this group's recommendations was, "A more rational, risk-based approach to determining the appropriate inspection intervals for bridges is needed, as opposed to a set twenty-four month cycle for all bridges" (ASCE/SEI-AASHTO, 2009). This paper went on to note that, "A more detailed inspection conducted less frequently may have a positive impact on the overall safety and maintenance of bridges in the U.S., allowing for broader application of Nondestructive Evaluation (NDE) technologies and a better understanding of the condition of individual bridges" (ASCE/SEI-AASHTO, 2009).

As per the ASCE/SEI-AASHTO ad-hoc group recommendation of a rational approach to bridge inspection, a wider use of NDE methods (specifically methods more advanced relative to visual inspection) is needed. The difficulty in increasing the use of these more complicated NDE methods is the increased costs and time spent using these methods. The ASCE/SEI-AASHTO ad-hoc group has a recommendation to attempt to solve this problem. Their suggestion of a more detailed inspection but conducted less frequently could result in a safer and more cost effective maintenance program. These NDE methods should contribute to a more detailed inspection procedure to make the inspection process more efficient and cost effective and bridges should become increasingly safer. This risk-based or reliability-based planning can be used to determine the appropriate inspection frequency, scope, intensity, and methodology. Doing this would help inspectors understand which bridges are high risk and what failure modes they may have. Although to do this, accuracy and reliability of each test must be quantified. This will allow inspectors to find a balance between these costs incurred and the accuracy and reliability of the methods. This will help develop a more efficient system for the inspection management process (ASCE/SEI-AASHTO, 2009).

An extensive literature search was performed and many NDE articles, papers, and reports for bridge engineering applications were found and reviewed. Based on the articles found, it was determined that there is a significant amount of research being conducted to determine the best situations (i.e., crack detection, rebar location, etc.) for specific NDE methods for bridges. Based on these studies, previous

surveys, and the results from the first round of the Delphi survey, commonly used NDE techniques for both concrete and steel bridges were determined. Various studies were also found that have been conducted to determine the accuracy and reliability of many of these methods relative to one another. These comparisons were both qualitative and quantitative in nature but only offered a limited insight to the accuracy and reliability of the various NDE tests.

Since it was determined that there is limited information quantifying the level of accuracy and reliability in the tests or to compare various tests to one another, a comprehensive survey to gather expert opinion was identified as a means to obtain the desired information. It was established that the Delphi method was an efficient and effective survey technique to gather this information. This survey aims to provide quantitative descriptions of accuracy, reliability, and bias (in terms of statistical descriptions) and a comprehensive comparison of the various tests to provide information to researchers and practitioners working in the fields of bridge management and inspection.

A total of four Delphi method rounds were conducted in order to determine quantitatively the accuracy, reliability, bias, and various costs of common NDE methods. The first survey was employed to determine background information of the participants and common NDE methods for bridges. The second and subsequent surveys were used along with various statistical scales in order to develop quantitative information based on each method.

The results of these surveys were used to develop quantitative information for each method. Based on these results, various conclusions were drawn. It was shown that most commonly used bridge NDE methods tend to be under biased, meaning the majority of the measured results are slightly less than the true value. However, these biases were shown to be less than 10%. Along with this, most commonly used bridge NDE methods tend to be relatively repeatable. Furthermore, it was shown that inspectors seem to have a relative understanding of the variability in different tests, but they tend to not have an understanding of the absolute scale of the variability. It was shown that the accuracy of commonly used bridge NDE methods tends to be relatively variable. For concrete testing, most tests had a true response percentage of about 80%. Finally, the various costs associated with the NDE methods examined tended to be quite variable, making this measure difficult to evaluate. However, there was a small trend that indicated tests that were cheaper in terms of equipment also tended to be easier and faster to perform.

1. INTRODUCTION

1.1 Current Bridge Inspection Practice

The majority of bridges in the United States were built during two periods. The first period of construction was in the 1930s during the Great Depression, and the second period was during the 1950s and 1960s with the implementation of the Federal-Aid Highway Act of 1956 (Rens et al., 2005). Ramey et al. (1997) show that the American Association of State Highway and Transportation Officials (AASHTO) *Standard Specifications for Highway Bridges* implies a bridge design lifespan of 50 years, meaning the majority of the bridges in the United States have reached the end of their expected service life. Combining this with a limited maintenance and repair budget makes efficient bridge management essential for improving the safety and serviceability of the current bridge system in the United States. There are three main elements to the management of bridges and other types of infrastructure: inspection of the system, decision-making about maintenance and repair of the system, and the performance of maintenance and repair on the system. This research project considers the inspection phase of the management process for bridges, specifically nondestructive evaluation.

The collapse of the Silver Bridge in West Virginia in 1967 started the first formal process for the inspection of bridges in the United States (Washer, 1998). After more than 40 years, visual inspection remains the most common inspection method (Phares et. al., 2004). Most of these inspections are conducted on a two-year cycle as required by the National Bridge Inspection Standards (NBIS) (Minchin et al., 2006). Due to the cyclical pattern of inspection, resources are used to re-inspect many bridges that may not need inspection. Examples of these are recently constructed bridges and standard bridges with proven track records and well understood deterioration modes. Using these inspection resources on bridges nearing the end of their service life may have prevented many of the recent bridge failures.

With recent bridge failures, specifically the I-35W bridge collapse in Minneapolis, bridge inspection practice in the United States has received much closer scrutiny. In response to this scrutiny, a joint American Society of Civil Engineers Structural Engineering Institute (ASCE/SEI) and AASHTO ad-hoc group was created to identify the issues to guarantee bridge safety and to study how current bridge inspection practices could be improved for the future (ASCE/SEI-AASHTO, 2009).

The ASCE/SEI-AASHTO ad-hoc group wrote a recent document entitled "White Paper on Bridge Inspection and Rating." In this paper, the group developed a listing of several deficiencies in existing inspection practice. One of this group's recommendations was, "A more rational, risk-based approach to determining the appropriate inspection intervals for bridges is needed, as opposed to a set twenty-four month cycle for all bridges" (ASCE/SEI-AASHTO, 2009). This paper went on to note that while in certain circumstances visual inspection is adequate, there are other cases when material defects and concealed elements are obstructed from view. Similarly, visual inspection is unable to detect micro defects and defects within the material, such as rebar corrosion. The paper also noted, "A more detailed inspection conducted less frequently may have a positive impact on the overall safety and maintenance of bridges in the U.S., allowing for broader application of Nondestructive Evaluation (NDE) technologies and a better understanding of the condition of individual bridges" (ASCE/SEI-AASHTO, 2009).

Nondestructive evaluation is a way to evaluate a structure without damaging the material's future usefulness. NDE is used in many fields, including mechanical engineering, civil engineering (including bridges), aeronautical engineering, medicine, and art. It should be noted that while nondestructive testing (NDT) and NDE are similar and the terms are often used interchangeably, they are not the same. NDT implies that only testing is being performed and data are being collected. Evaluation is the process of making judgments about the data gathered. Often, the evaluation in NDE implies both data collection and analysis are being done (Shull 2002).

There are many different NDE methods that have been developed in recent years by various organizations, including the Federal Highway Administration (FHWA), to evaluate different material properties and bridge conditions (see Section 0 for a description of commonly used bridge NDE methods). These various methods have become increasingly popular due to the nondestructive nature of the assessment. Visual inspection is considered to be a form of nondestructive evaluation and is the most widely used method due to the relatively low costs. Visual inspection has been shown, however, to have many flaws that can stem from, among other things, inspector bias, lack of experience, inability to "see" internal conditions, and concealed elements (Washer 1998).

As per the ASCE/SEI-AASHTO ad-hoc group recommendation of a rational approach to bridge inspection, a wider use of NDE methods (specifically methods more advanced relative to visual inspection) is needed. The difficulty in increasing the use of these more complicated NDE methods is the increased costs and time spent using these methods. The ASCE/SEI-AASHTO ad-hoc group has a recommendation to attempt to solve this problem. Their suggestion of a more detailed inspection but conducted less frequently could result in a safer and more cost effective maintenance program. These NDE methods should contribute to a more detailed inspection procedure to make the inspection process more efficient and cost effective and bridges should become increasingly safer. This risk-based or reliability-based planning can be used to determine the appropriate inspection frequency, scope, intensity, and methodology. Doing this would help inspectors understand which bridges are high risk and what failure modes they may have. Although to do this, accuracy and reliability of each test must be quantified. This will allow inspectors to find a balance between these costs incurred and the accuracy and reliability of the methods. This will help develop a more efficient system for the inspection management process (ASCE/SEI-AASHTO, 2009).

1.2 Research Objectives

The goal of this thesis is to facilitate risk-based inspection planning by quantifying the accuracy and reliability of common NDE methods for bridges. This will give bridge inspectors a better understanding of NDE methods relative to each other and will give managers the data they need to incorporate the uncertainty in inspection results in bridge management. This will allow them to create a more efficient evaluation process rather than using the current two-year cycle.

The specific objectives of this thesis are to:

- Determine the most common and practical NDE methods for steel and concrete bridges
- Determine the type of traits (accuracy, reliability, etc.) that should be analyzed in order to describe the uncertainty in NDE in a quantitative way
- Implement data collection to obtain quantitative data about the common NDE methods for bridges to facilitate risk-based inspection planning

1.3 Research Methodology

This thesis focuses on the accuracy, reliability, bias, and costs of common NDE methods, including visual inspection, which can be used to test certain bridge elements. Data about the accuracy, reliability, bias, and various costs of each NDE method were collected. This was done by two methods. First, a literature review was performed to determine common NDE methods being used and studied for bridge inspection. Also during the literature review, data pertaining to the specific traits of the NDE methods were sought. During the literature review process it was determined that there are very little quantitative data being published that could be used to establish general statistical descriptions for the uncertainty in various NDE methods or even give relative comparisons between tests. To complement the literature, a Delphi Method survey was conducted with experts in the NDE for bridges field in order to develop a broader range of data that match real life practices. All of the data were then analyzed and conclusions were drawn to quantify the accuracy, reliability, bias, and various costs incurred for common bridge NDE methods.

1.4 Thesis Organization

This thesis contains four additional chapters covering NDE methods and the accuracy of these methods for bridges. Section 2 contains a literature review describing the current state of NDE research for bridges along with previous surveys that were conducted involving NDE methods for bridges. This chapter also contains a brief description of each NDE method that was analyzed for this thesis and an explanation of each. Section 3 explains the implementation of the Delphi survey with experts in the bridge NDE field. The results obtained from the Delphi survey are presented and discussed in Section 4. Section 5 contains a summary of the work performed and conclusions that were drawn from this work. Furthermore, suggestions for future work are included in this chapter.

2. BACKGROUND AND LITERATURE REVIEW

2.1 Introduction

An extensive literature search was performed, and many NDE articles, papers, and reports for bridge engineering applications were found and reviewed. Based on the articles found, it was determined that there is a significant amount of research being conducted to determine the best situations (i.e., crack detection, rebar location, etc.) for specific NDE methods for bridges. Based on these studies, previous surveys (see Section 2.4), and the results from the first round of the Delphi survey, commonly used NDE techniques for both concrete and steel bridges were determined. A description of these methods, including available data about accuracy and reliability, can be found in Section 2.2. Various studies have also been conducted to determine the accuracy and reliability of many of these methods relative to one another. Section 2.3 describes these studies and presents data based on the findings. Furthermore, Section 2.4 describes studies that have been conducted to compare various costs of these methods. Finally, Section 2.5 gives a description of the Delphi method as a tool to gather information from experts in the field and how this method is implemented.

2.2 Commonly Used NDE Methods and Current Research

2.2.1 Visual Inspection

Visual inspection is usually one of the first NDE methods used for locating defects on all structural members. Visual inspection can also be used after more advanced methods identify a defect to give the inspector more detail (Mix 2005). Visual inspection refers to inspecting a structural member with the five senses and very basic tools (i.e., flashlights, tape measures, etc.). A subset of visual inspection, visual testing can also include more advanced optical devices such as borescopes and microscopes. It should be noted that the concepts of visual inspection and visual testing are slightly different and are often confused with one another. Given the more complicated nature of visual testing, visual inspection may include visual testing, but certain aspects of visual inspection may not be included within visual testing (Moore et al. 2001).

In 2001, Phares et al. conducted a study trying to quantify the reliability of bridge visual inspection. Forty-nine bridge inspectors from 25 state DOTs were asked to conduct seven routine visual inspections and three in-depth visual inspections on two of the FHWA's Nondestructive Evaluation Validation Center (NDEVC) test bridges while being monitored by NDEVC staff. A routine inspection was defined as inspecting and issuing an overall rating to the superstructure, substructure, and deck elements while an in-depth inspection is a more comprehensive inspection of specific aspects of these elements (welds, paint, pins, etc.). The results of this study indicated that the majority of the inspector assigned ratings were statistically different than the reference ratings established by NDEVC personnel. It was shown that during the routine inspection, visual inspection of the superstructure, substructure, and deck had an overall bias of +3%, -5%, and +5%, respectively. For example, the average response for the superstructure was a 5.61 rating while the average reference rating was 5.42, implying the inspectors determined the superstructures were in better condition than they actually were. The coefficients of variation (COV) for the ratings of the superstructure, substructure, and deck were 0.14, 0.12, and 0.16, respectively. Furthermore, it was determined that in-depth inspections may not yield any more detail than the routine inspections. It was also shown during the in-depth inspection that a low percentage of inspectors were able to identify localized deficiencies, as shown in Figure 2.1. Note that 42 inspectors worked on the STAR Bridge B544 and 44 inspectors worked on the Route 1 Bridge (there were a few details on the Route 1 Bridge with only 42 inspectors).

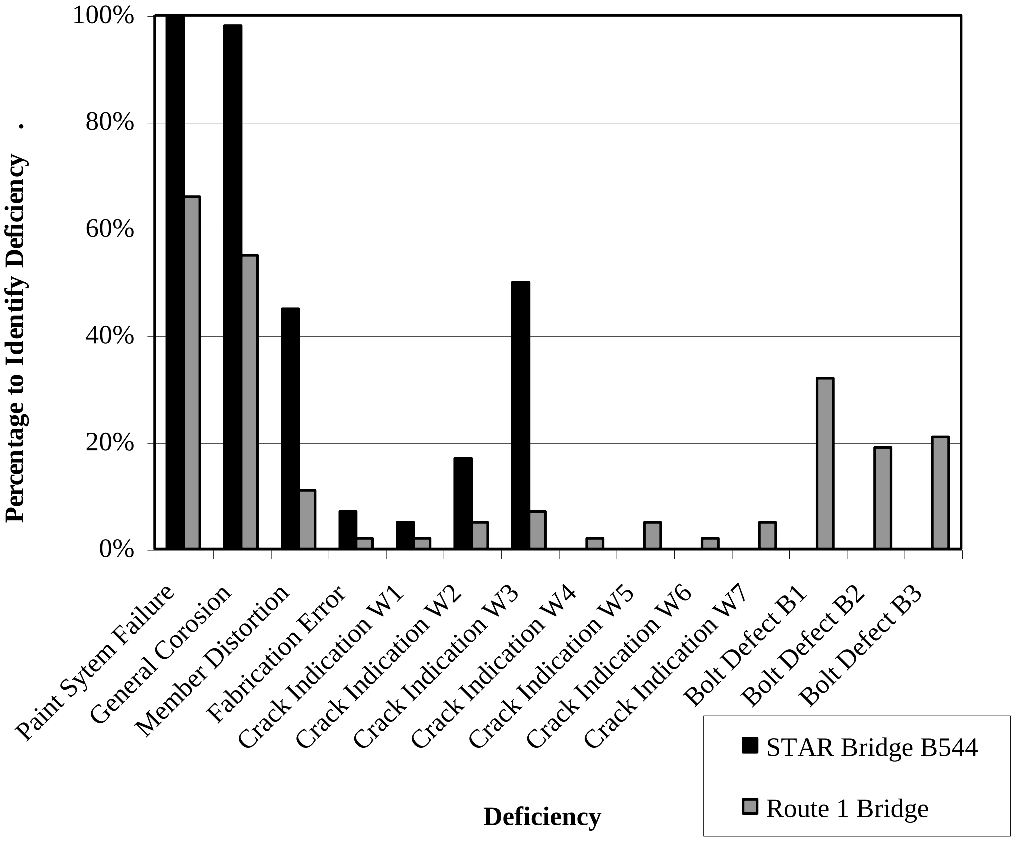


Figure 2.1 Percentage of Inspectors to Identify Deficiencies during an In-Depth Inspection of Two FHWA NDEVC Test Bridges (Phares et al. 2001)

This study also found various factors affected the inspector's reliability, including fear of traffic, near visual acuity, color vision attributes, formal bridge inspection training, and the inspector's perception of the bridge's maintenance, accessibility, and complexity. The study concluded that there are many aspects of visual inspection that need improvements. It should be noted that the visual inspections were not compared to any other methods (Phares et al. 2001).

2.2.2 Acoustic Emission

Acoustic emission is an NDE method in which the material being tested generates acoustic signals that warn of increased mechanical or thermal stress. The basis of acoustic emission is the fact that materials will emit a sonic or ultrasonic wave when stressed to the point where increased deformation or fracture occurs. This method measures the low amplitude signal that is produced when dislocations in the material's crystal lattice structure are created during plastic deformation. Due to the need of plastic deformation, acoustic emission for bridges is used mainly on steel members (i.e., girders and cable stays). Since this test measures flaws while they occur, the flaw cannot be retested and requires continuous monitoring (Mix 2005).

In 2001, Nair conducted a case study of a pre-stressed concrete bridge and a steel girder bridge using acoustic emission. Both bridges were loaded with static and dynamic loadings and were monitored for damage using acoustic emission. Through this study, Nair showed that when a material reaches a certain stress level it emits a signal that can be correlated into the severity of the damage. Similarly, Golaski et al. (2002) performed a case study of five concrete bridges of varying ages and degrees of damage. Similar conclusions were drawn from this study; however, neither study compared acoustic emission testing to other NDE methods or provided data to determine the accuracy or reliability of the method.

A study conducted by Gong et al. (1992) of 36 steel railroad bridges related the acoustic emission test output to a stress intensity factor, K. The stress intensity factor is a function of both the stress level and the crack length. By determining the range of the stress intensity factor, ΔK, the severity of the crack can be determined. A correlation between ΔK and crack intensity can be seen in Table 2.1.

Table 2.1 Correlation Between the Stress Intensity Factor Range, ΔK, and Crack Intensity for Steel Bridges (Gong et al. 1992)

Range of ΔK	Crack Description
$0 < \Delta K < 10$	Minor Defect
$10 \leq \Delta K < 20$	Slow Crack Growth
$20 \leq \Delta K < 30$	Requires Repair
$30 \leq \Delta K < 40$	Dangerous
$40 \leq \Delta K$	Imminent Failure

Bridge engineers use this system to plan, schedule, and prioritize maintenance. A ΔK reading of 10 or higher indicates a detailed inspection is required in the area. While there was no indication of the accuracy or reliability of this system, it was determined that noise could be caused by rubbing, hammering, rain, and electrical system noise, which could skew the results. There was no data that showed how much this could affect the results (Gong et al. 1992). Furthermore, Rens et al. (2005) compares acoustic emission to various other tests (see Section 2.3).

2.2.3 Cover Meter/ Pachometer

A cover meter is a method used to measure the concrete cover over the rebar. Along with this, the instrument can detect rebar size (if cover distance is known) and direction. This is done by creating an alternating magnetic field with a probe coil. The instrument is moved along the surface and metal objects can be detected with the range of the magnetic field. These measurements change the voltage of the output as a function of concrete cover and bar diameter. It should be noted that the instrument cannot measure both concrete cover and bar diameter simultaneously as the voltage is dependent on both variables. Oftentimes, the bar diameter is known and the concrete cover is verified (Song and Saraswathy, 2007).

In 2011, Algernon performed a study in which test blocks of known dimensions and rebar locations were created. Different blocks included bar sizes ranging from No.3 to No. 9 to determine the influence of bar diameter on measurements. Furthermore, another set of blocks was manufactured with varying bar spacings to measure how this affected accuracy. Various geometries were also created, including crossing rebar and layered rebar at various depths. The first measurement was at a section of No. 3 rebar with crossing bars. In a section with dense bar crossings, the cover meter measured 1.14 inches too low while in a section with the same rebar diameter and no cross bars the cover meter measured the cover to within a tolerance of 0.04 inches. A similar measurement was performed with decreased bar spacing. This measurement indicated slightly less cover due to the influence of neighboring bars.

The next measurement was taken at a section with No. 9 bars but with the cover meter equipment at the setting for No. 3 bars. As noted previously, cover and bar diameter cannot be measured simultaneously. One of the parameters must be known in order to measure the other one. Based on this setup, the cover was measured as slightly less than actual due to the wrong bar diameter input (Algernon 2011).

Based on these results, it was shown that cover meter measurements with known bar diameter, wide enough spacing (more than 3 inches) and no crossing layers can have good bias of about +/- 0.04 inches. However, the introduction of one or more of these parameters can reduce the cover depth reading providing a conservative reading (Algernon 2011). These readings are considered conservative because it is generally better to measure the cover to be less than the actual as more cover means more protection for rebar. Furthermore, Rens et al. (2005) provide a qualitative comparison of cover meter with other methods (see Section 2.3).

2.2.4 Electrical Potential

Electrical potential measures the ability of an electric current to flow within a material. This indicates the material's transfer properties. Electrical potential techniques for bridges have mainly been developed and used to detect steel reinforcement corrosion in concrete structures. Steel corrosion in concrete is mainly dependent on moisture and chloride content. These factors also influence the electrical properties of the concrete. By measuring the electrical potential, the content of moisture and chloride content can be determined. From this, corrosion in the steel is not directly measured, but rather the probability of corrosion is measured indirectly through these electrical properties in the concrete (Maierhofer et al. 2010).

Gucunski et al. (2010) points out that electrical potential measurements cannot produce quantitative data. Rather, this method measures the potential in the concrete. The more negative the potential, the higher the chance of corrosion. ASTM C876 provides general guidelines for evaluating the potential readings. In general, if a potential reading is higher than –0.2V there is a 90% chance there is no corrosion, while if the reading is lower than –0.35V there is a 90% chance there is corrosion (ASTM International 2009). Furthermore, Barnes and Trottier (2000) performed a case study in which deterioration was compared to other methods and Rens et al. (2005) provide a qualitative comparison of electrical potential with other methods (see Section 2.3).

2.2.5 Impact Echo

Impact echo is based on the material's vibrational response when it has been impacted. This method is typically applied to materials with two parallel surfaces. After impact, waves will propagate within the material and will be reflected from the boundaries. The reflected waves have a maximum peak in the frequency signal, which can be used to determine the material's thickness. This method is typically used on concrete slabs to determine the thickness of the material. Along with this, technicians can also determine if an area in the slab has delamination or spalls based on an observed reduced thickness relative to the rest of the slab, as seen in Figure 2.2. While it is much less common, impact echo can also be used to determine the thickness of steel members as well (Maierhofer et al. 2010).

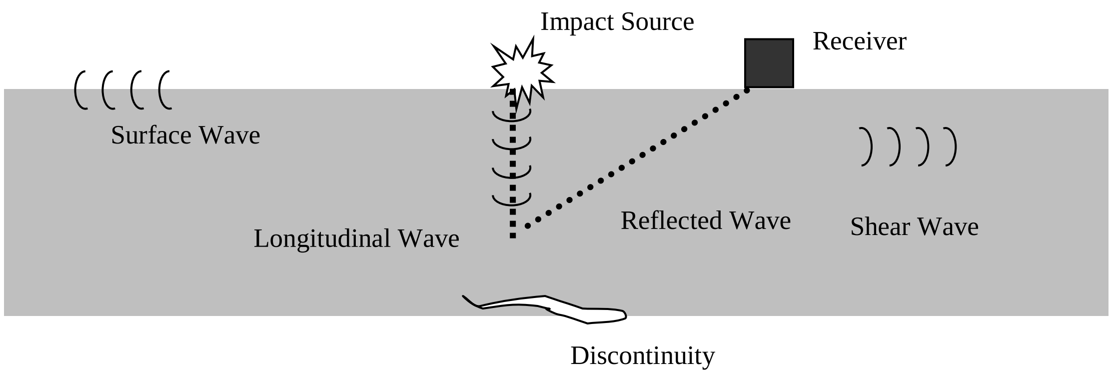


Figure 2.2 Schematic of the Impact Echo Method Being Performed on a Damaged and Undamaged Concrete Deck

Various studies have been conducted to research the accuracy and reliability of impact echo. Of these studies, Watanabe et al. (2004) performed a case study in which voids of various sizes and depth were placed in a concrete member and tested using impact echo. According to the authors, it was shown that for concrete with no rebar, impact echo could very accurately identify nearly the exact void sizes; however, no numbers were provided. It was discovered, however, that by introducing rebar, the accuracy was reduced. While this reduction in accuracy was not quantified, it was shown that the voids could still be detected, but the sizes were harder to determine as the rebar spacing decreased. Furthermore, two studies (Yahia et al. 2007 and Scott et al. 2003) attempted to compare impact echo to other methods with varying results (see Section 2.3).

2.2.6 Liquid Penetrant

The liquid penetrant method is performed by applying a liquid dye to the surface of a member (commonly steel) and allowing it to stand for a period of time. The penetrant is absorbed by capillary action into any surface discontinuities. Once absorbed into the discontinuities, excess liquid is removed and a light-colored developer is applied that draws some of the liquid penetrant out of the discontinuity making the flaw easier to see. Since the penetrant is carried into the defects by capillary action, the method can only be used to detect surface defects such as cracks and poor welds. Along with this, liquid penetrant cannot be used with porous material, making it difficult to be used on concrete members (Mix 2005).

McCrea et al. (2002) looked at various parameters that could affect the accuracy and reliability of the liquid penetrant test. Among these parameters are the defect size and wait time of the penetrant. The smaller the defect is the less liquid will be absorbed into the defect, making the defect harder to detect. No minimum volume was determined, however. Along with this, if the area is not properly cleaned prior to administering the penetrant or the penetrant is not given enough time on the sample, the liquid will have difficulties being absorbed.

2.2.7 Magnetic Particle

The magnetic particle test is a method similar to the liquid penetrant method in that it can only detect surface flaws. To do this, the material is magnetized by a magnetic coil and fine ferromagnetic particles are poured onto the surface. Defects in the material will affect the magnetic field from the magnetization causing the particles to attract to it. These particles outline the surface defects. Since the sample must be magnetized, this method is limited to magnetic materials such as ferromagnetic steel (Shull 2002).

McCrea et al. (2002) looked at various parameters that could affect the detection of surface defects. They found that the size of the coil could affect the sensitivity of the test; the larger the coil the stronger the magnetic field is, making the test more sensitive. Also, every component must be tested at least twice to ensure the magnetic field travels perpendicular to the defect. Defects that run parallel to the field may not be detected by this method. Furthermore, Shull (2002) shows that there is almost no limitation to the size or shape of the flaw being tested. It was also shown that magnetic particle testing could detect limited subsurface defects up to a maximum depth of about 6.35 mm.

2.2.8 Mechanical Sounding

Mechanical sounding is a broad term used for testing of concrete members. The method is done by either dragging an instrument (often chains) across the surface or lightly hitting the surface with a hammer like tool. This is done to identify delamination in the concrete. By using mechanical sounding, voids can be detected by the deep thud sound rather than a clear ringing sound that would occur for solid concrete (Scott et al. 2003).

There have been various studies that have used mechanical sounding to validate other test methods. This is because mechanical sounding is a relatively old and well-known method. It is not, however, necessarily more accurate or efficient than the newer more advanced methods (i.e., impact echo and radar). Among these studies were Scott et al. (2003), Barnes and Trottier (2000), Wood and Rens (2006), and Clark et al. (2003). All these studies either compared the ability to detect deterioration or to determine the amount (area) of deterioration (see Section 2.3).

2.2.9 Radar

The radar method (also known as ground penetrating radar) is an electromagnetic method. For this method, a transmitter emits an electromagnetic pulse. This pulse is then reflected to the receiver or transmitted through the material to a receiver on the other side. The travel time of the pulse is measured for the determination of various geometric and internal properties of the material. Radar is commonly used for concrete member applications to determine member thickness, layer thicknesses of rebar, and location of rebar, ducts, anchors, and cavities within the concrete (Maierhofer et al. 2010).

Various studies have been conducted to research the accuracy and reliability of radar (specifically ground penetrating radar). Among these studies were Barnes and Trottier (2000), Yehia et al. (2005 and 2007), and Wood and Rens (2006). All these studies either provided qualitative data on the ability of radar to detect deterioration when compared with other methods or quantitative data to determine the amount (area) of deterioration (see Section 2.3).

2.2.10 Radiography

Radiography is the use of electromagnetic waves (often X-rays) to look internally at a material. These waves are emitted and travel through the material and received by the detector. By doing this, the waves can “see through” objects that are opaque. The intensity of waves that pass through the material is based on material composition, density, and thickness allowing for these properties to be measured. This method is commonly used on steel bridge members to determine thickness of the member, detect fractures, and inspect welds (Shull 2002) and voids and cavities in concrete bridges (McCrea et al. 2002). While there is little being done to determine the accuracy and reliability of radiography, McCrea et al. (2002) discuss various parameters that can affect the data. Among these are the exposure time, focal size, and defect orientation relative to the electromagnetic waves. Along with this, Rens et al. (2005)

qualitatively determined the adequacy of radiography for concrete and compared it with other methods (see Section 2.3).

2.2.11 Rebound Hammer

The rebound hammer method (also known as Schmidt hammer) is a test on concrete to determine the concrete compressive strength. This is done by impacting the surface of the member with the hammer and measuring the rebound. This rebound is then translated to the rebound number, which is directly proportional to the concrete's compressive strength. A high rebound number corresponds to a high compressive strength and a low rebound number corresponds to a low compressive strength (Rens 2006).

In 2006, Wood and Rens conducted a case study on the Lawrence Street Bridge in Denver. In this study, rebound hammer testing was compared with strength results from core samples. Along with this, the method was also compared with other NDE methods. Furthermore, Rens et al. (2005) provides a qualitative comparison of the rebound hammer with other methods (see Section 2.3). It has also been shown by various studies, including Qasrawi (2000), that there is a correlation between concrete quality, rebound hammer results, and ultrasonic results. It was shown that both of these methods could be used to indirectly determine concrete quality by nondestructive means, but no numbers related to accuracy or reliability were provided.

2.2.12 Thermal Imaging

Thermal imaging uses special cameras to detect infrared radiation. This radiation can be used to determine the temperature of a material's surface. The camera is pointed at the material and a spectrum of colors representing different temperatures can be seen. This method is commonly used for concrete bridge members to determine regions of voids or delamination. This can be done because areas of voids tend to be cooler than the surrounding area (Clark et al. 2003). Some research is being done to identify the accuracy and reliability of the thermal method. Two of these studies (Clark et al. 2003 and Yahia et al. 2007) developed comparisons of thermal testing to various other NDE methods (see Section 2.3).

2.2.13 Ultrasonic

The ultrasonic method is a method that uses high frequency (ultrasonic) waves. These waves are emitted by a transducer and are either received by another transducer or reflected back to the original transducer. The waves are then transformed into an electrical pulse and observed on an oscilloscope. Based on the wave propagation through the material, various material properties can be measured (Shull 2002). Among others, flaws, fractures, corrosion, thickness, weld imperfections, and pin discontinuities can be measured in steel members. For concrete members, ultrasonic testing is commonly used to determine thickness, locate rebar, and detect voids (Maierhofer 2010).

Many studies have been conducted to identify accuracy and reliability of the ultrasonic method. As mentioned previously, a comparison was made by Qasrawi (2000) to correlate concrete quality, rebound hammer results, and ultrasonic results. Furthermore, Rens et al. (2005) qualitatively determined the adequacy of the ultrasonic method and compared it with other methods while Wood and Rens (2006) compared the ultrasonic method with other methods through a case study (see Section 2.3).

2.3 Comparative Studies for Accuracy, Reliability, and Bias of NDE Methods

There were a number of studies and papers discovered during the literature review that compared various NDE methods to one another. These comparisons were both qualitative and quantitative in nature and offer a limited insight to the accuracy and reliability of the various tests described previously and a comparison of these parameters for different test methods. The following is a discussion of these studies.

2.3.1 Comparing Radar, Chain Drag (Mechanical Sounding) and Electrical Potential – A Barnes and Trottier 2000 Study

In 2000, Barnes and Trottier conducted a study on nine concrete bridges using ground penetrating radar, chain drag, and electrical potential to determine the accuracy of these methods to identify delamination and voids. The results were expressed in terms of percentage of area of the total bridge deck that was found to have deterioration. These results were then compared with the actual percentage of deterioration found and repaired, which was assumed to be the true value. It should be noted that the repair percentage was based on the chain drag results and more area was repaired as seen fit. While the area repaired may not be the absolute true value, it was assumed to be a representation of the true value in terms of maintenance planning. These values could be used to allocate expenses based on the results of a test. An example of this is in the case of the chain drag method. According to these results, this method tends to relatively reliably under-predict the true value. With this being the case, a larger budget should be allocated to offset the result. The results of the study can be seen in Table 2.2.

Table 2.2 Summary of Barnes and Trottier Study Results (2000)

Structure	Radar Percentage	Chain Drag Percentage	Electrical Potential Percentage	Area Repaired Percentage	Radar Bias Factor	Chain Drag Bias Factor	Electrical Potential Bias Factor
Stewiacke River Bridge	44.5	53.9	50.5	59.7	1.342	1.108	1.182
Skye River Bridge	42.3	34.4	41.9	38.7	0.915	1.125	0.924
Baddeck River Bridge	37.4	34.9	46	40.1	1.072	1.149	0.872
Shubenacadie CNR Overpass	28.5	35.1	39.5	35.3	1.239	1.006	0.894
Grand Pre Overpass	15	9.1	8.3	11.2	0.747	1.231	1.349
Deep Hollow Overpass	70.1	54	31.4	54.6	0.779	1.011	1.739
Victoria Bridge	13.6	5.2	0.1	11.1	0.816	2.135	111.000
Rough Brook Bridge	21.2	27.2	N/A	29.4	1.387	1.081	-
Glendale Bridge	16.2	20.1	9.3	22.6	1.395	1.124	2.430
Average	32.089	30.433	28.375	33.633	1.077	1.219	15.049
Std Dev	18.538	17.221	19.575	17.103	0.271	0.350	38.774
COV	0.578	0.566	0.690	0.509	0.251	0.287	2.577

Based on these results, it can be seen that ground penetrating radar tended to be the least biased (percentage repaired divided by percentage measured) with an average bias ratio of 1.077. Chain drag and electrical potential were more biased with an average bias ratio of 1.219 and 15.049, respectively. Radar also tended to be the most reliable with the smallest COV of 0.251. This was followed by a COV of 0.287 and 2.577 for chain drag and electrical potential, respectively. Note that by removing the outlying data point of the Victoria Bridge from the electrical potential method the data seems more reasonable with a bias factor and COV of 1.341 and 0.427, respectively. This shows, however, that there is a larger level of uncertainty when it comes to electrical potential. It should be noted that Barnes and Trottier performed another similar study with similar results published in 2004. However, no data about the repaired area were provided. Without the provided repaired area, or assumed true value, no computation of bias could be made It should be noted, however, that the COVs from the 2004 study for radar, chain drag, and electrical potential were found to be 0.61, 1.12, and 1.11, respectively. As shown, these values are slightly higher than the 2000 study.

2.3.2 Comparing Chain Drag (Mechanical Sounding) and Impact Echo to Concrete Core Results – A Scott et al. 2003 Study

Scott et al. (2003) conducted a study to identify delamination on the Van Buren Road Bridge. The results of the chain drag (a mechanical sounding method) and impact echo were compared to actual concrete cores. To conduct this study, NDEVC staff performed a comprehensive chain drag survey on the entire concrete deck. Impact echo measurements were then taken at 10 predetermined grid locations along the deck and where the core samples were to be taken. The chain drag and impact echo results perfectly matched the core sample results at these grid locations. Furthermore, the impact echo test conducted at the predetermined grid locations along the deck matched the actual core results 70% of the time (20% of the tests produced readings that indicated distress but were not accurate enough to definitively determine delamination). It should be noted that a ground penetrating radar study was also performed on this bridge with two types of systems (one system commercially available and one that was under development for FHWA). While these results were not directly compared, they were shown to be reasonably accurate relative to the other methods.

2.3.3 Comparing Coin-Tap Test (Mechanical Sounding) to Infrared Thermography – A Clark et al. 2003 Study

Clark et al. (2003) conducted a case study to determine the accuracy and reliability of the infrared method. Five spans of a concrete bridge in Northamptonshire in the United Kingdom were first tested with the coin-tap test, a mechanical sounding technique. These spans were then tested with an infrared camera. Both techniques were used to determine locations of delamination. It was shown in this study that the infrared imaging and coin-tap tests matched in five of the eight tests. Furthermore, in two instances, the infrared test located a delamination that the coin-tap test did not find. In one case, at the south abutment, the infrared test gave mixed results due to a damp patch while the coin-tap test measured delamination. This patch gave inaccurate temperature readings. It should be noted that these results were not confirmed with actual core samples as the bridge was still in use.

2.3.4 Comparing Various NDE Methods on Deteriorating Concrete Bridges – A Rens et al. 2005 Report

In this report, Rens et al. (2005) explain the use of the bridge management system for the city and county of Denver. While the only testing was done with the ultrasonic method, Rens et al. compare various tests that can be used on concrete to determine efflorescence, cracking, and delamination and spall. These methods are also compared on a relative cost basis (see Section 2.4). These common methods were determined through the surveys conducted by Rens et al. (1997) and Rens and Transue (1998). The results of this comparison can be seen in Figure 2.3. These results for efflorescence, cracking, and delamination and spall are based on a three-point Likert scale and no quantitative data was reported for these tests. It should be noted that the ultrasonic test performed was able to locate the location and size of vertical cracks within the bridge pier cap by taking three measurements vertically along the member, but these readings were not compared to any other method.

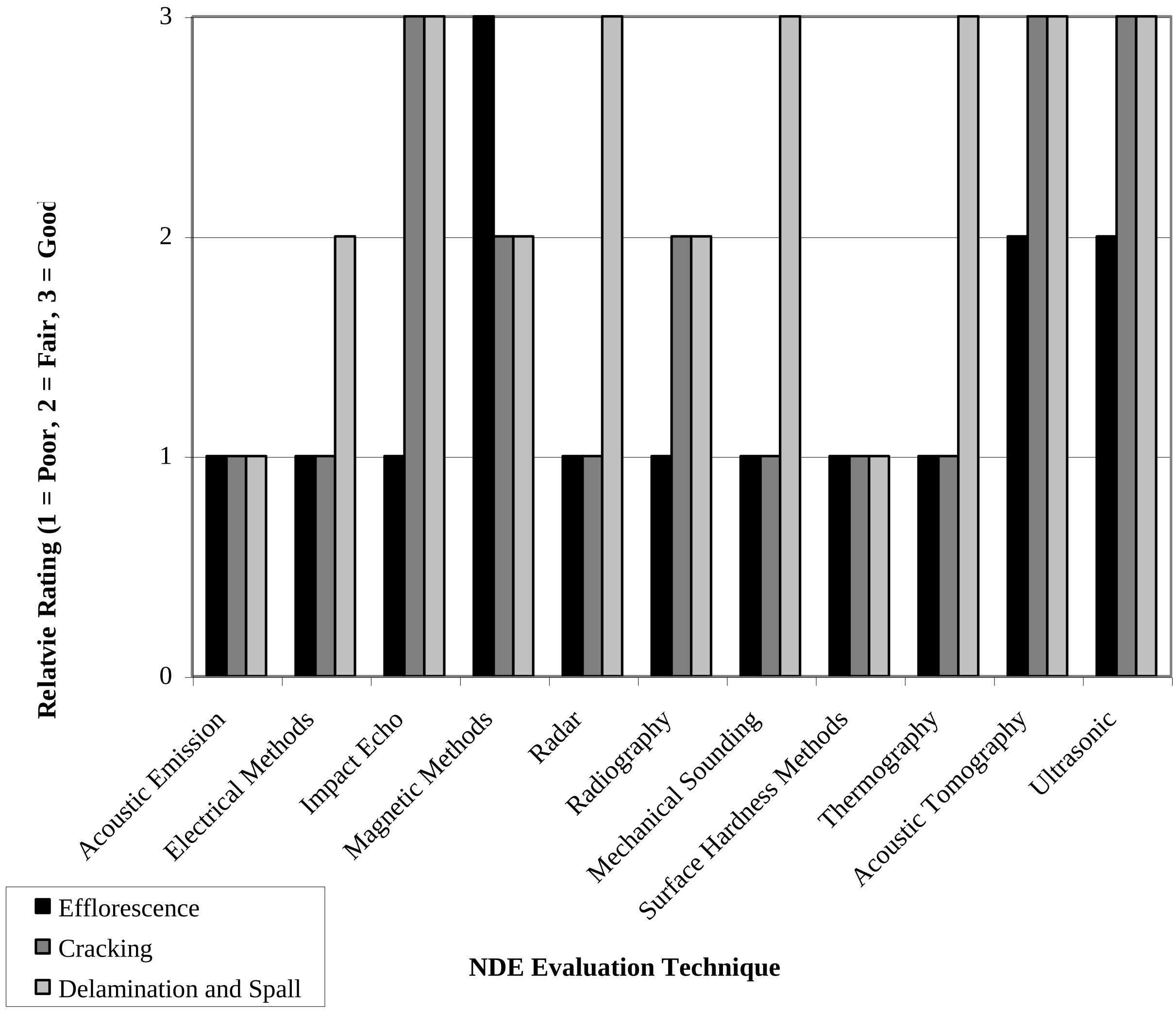


Figure 2.3 Summary of Rens et al. (2005) Concrete NDE Method Comparisons

2.3.5 Comparing Ultrasonic, Hammer Sounding (Mechanical Sounding), Surface Hardness (Rebound Hammer), and Radar – A Wood and Rens 2006 Case Study

In 2006, Wood and Rens conducted a case study of the Lawrence Street Bridge in Denver. Due to water penetration and freeze thaw cycles, the pier cap substructure of the bridge was deteriorating. The study was conducted to understand the amount of deterioration and to compare the results of various NDE methods. A total of five NDE methods were performed at five locations along the pier cap. All tests were performed to determine the amount of cracking in the concrete structure.

After the completion and compilation of the NDE data, computer programs were used to visually represent the data with colored regions. Based on these regions, each location was given a rating for each test. Core samples were then taken at each of the five locations and compared with the NDE results. The results of these tests can be seen in Figure 2.4. As shown, there was a wide variation among the NDE tests and the comparisons of these tests to the core results. Ultrasonic and hammer sounding tended to give similar results to one another while surface hardness and ground penetrating radar tended to match each other. Furthermore, surface hardness and ground penetrating radar tended to be more consistent with the relative core conditions. Note that no ground penetrating radar data were given for location A.

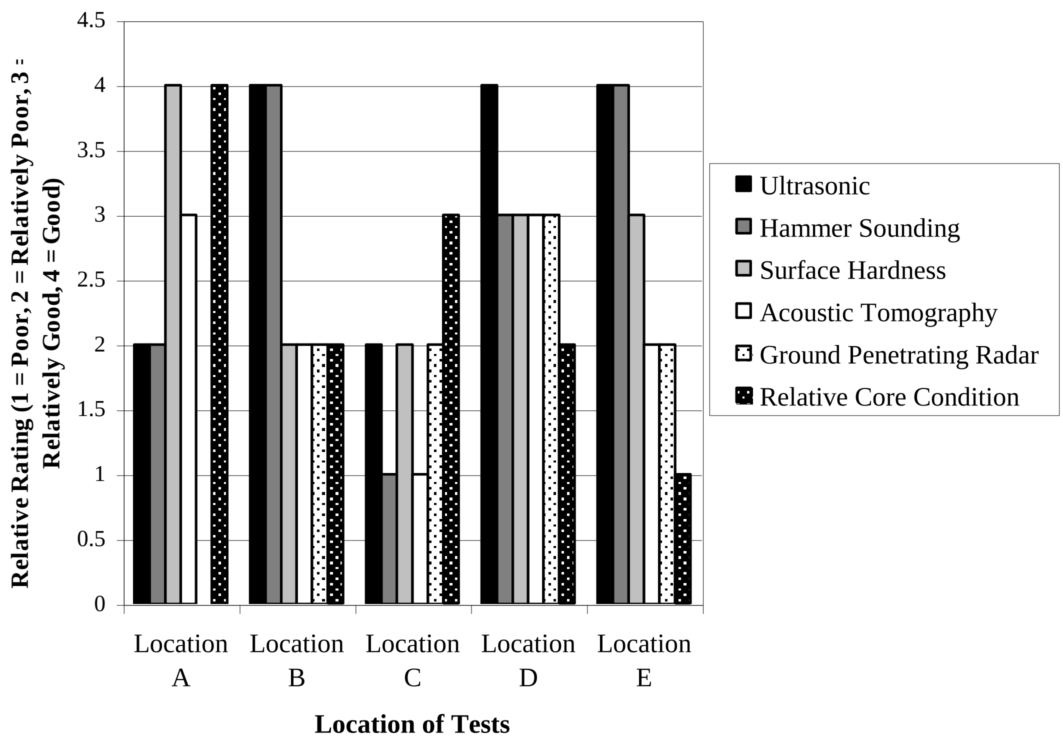


Figure 2.4 Summary of Wood and Rens (2006) Case Study Results for NDE Tests Methods on the Lawrence Street Bridge

2.3.6 Comparing Impact Echo, Ground Penetrating Radar, and Infrared Thermography – A Yehia et al. 2007 Study

Yehia et al. performed a study to determine the reliability of impact echo, ground penetrating radar, and infrared thermography on concrete bridge decks. Three types of flaws—cracks, delaminations, and voids—of known location and dimension were introduced to specimens and the ability of each method to identify these flaws was tested. The results of the ground penetrating radar tests can be seen in Figure 2.5. Note that a similar procedure was conducted with the impact echo method and the method detected each deficiency 100% of the time. As shown, radar and impact echo were both fairly accurate when detecting delaminations and voids. In the cases where radar did not detect these flaws, it was determined the flaw was either too close to the surface (less than 1.25 inches) or too small (less than 0.25 inches in diameter). Both tests were able to measure the depth of detectable voids with accuracy exceeding 95%. Note that while radar was unable to detect cracks, impact echo was able to detect cracks with 100% accuracy.

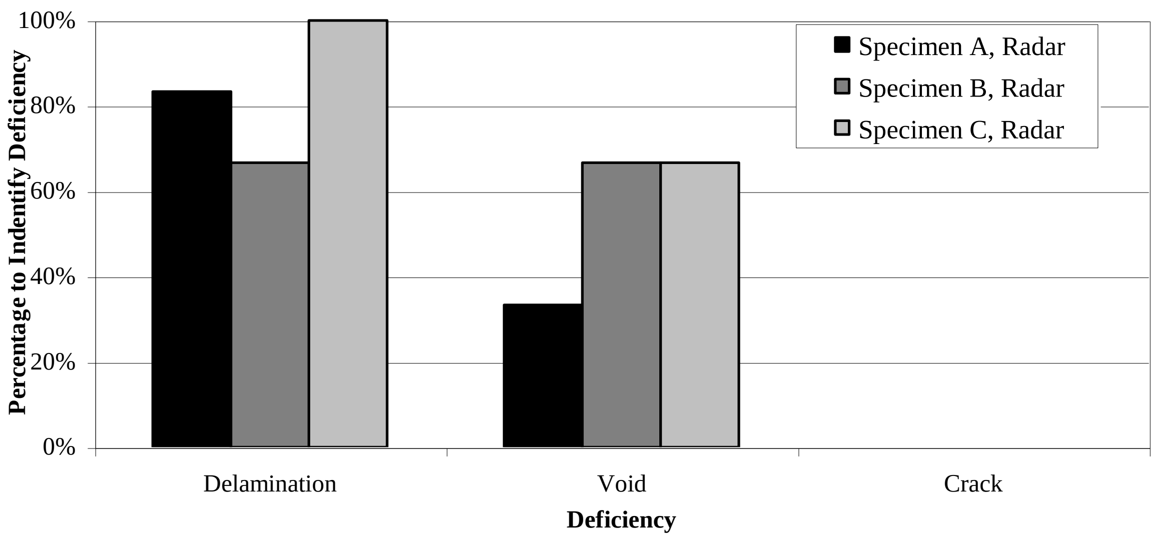


Figure 2.5 Percentage of Radar Tests to Identify Deficiencies in Concrete Specimens (Yehia et al. 2007)

The results of the infrared thermography test can be seen in Figure 2.6. For the infrared method, specimens were tested at different times of day in order to understand the effects of temperature variations during these times on the results. It was shown that the ambient temperature did not matter as much as the amount of sun exposed to the material. Because of this, there was no detection of flaws during the nighttime hours. Furthermore, it was determined the deeper (more than 2 inches) and smaller the flaws were, the harder it was to detect them. This could be seen in Specimen C (not represented in Figure 2.6) where all defects were deeper than 2.25 inches and no flaws were detected.

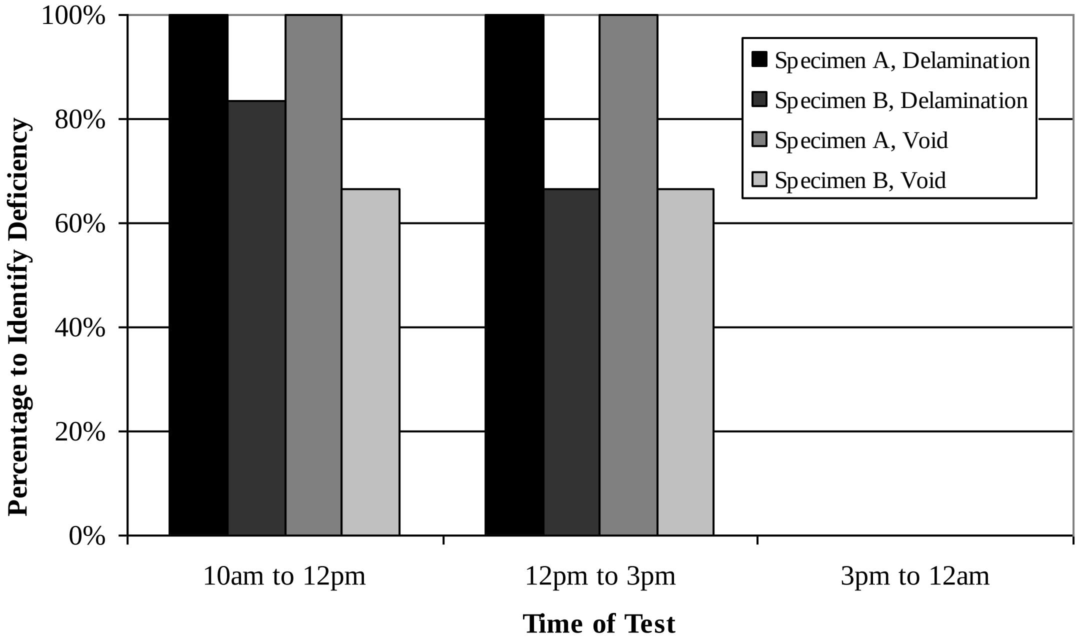


Figure 2.6 Percentage of Infrared Tests to Identify Deficiencies in Concrete Specimens (Yehia et al. 2007)

2.3.7 Comprehensive NDE Concrete Bridge Deck 2013 Study by the Transportation Research Board

In 2013, Gucunski et al. published a comprehensive study of common NDE methods that are used to identify concrete bridge deck deterioration through the Transportation Research Board (TRB). Prior to the testing, nine NDE methods that can be used to detect deterioration of concrete bridge decks were selected via a literature review. Of these nine methods, six (impact echo, ultrasonic, electrical potential, radar, mechanical sounding, and infrared) were included in the Delphi survey conducted for this research. Only these methods will be analyzed below (Gucunski et al. 2013).

During this study, ten organizations (industry vendors and research centers) used two different methods of validation testing. These methods were field and laboratory testing. For both methods, predetermined grids were used to identify locations on the bridge, and detailed testing instructions were provided to all participants prior to testing to ensure the same testing procedures. After testing was completed, cores were removed from the sample to provide ground truth data. For the field validation testing, a portion of the Route 15 bridge over I-66 in Haymarket, Virginia, was selected. For laboratory testing, two test decks were prepared. The first was a newly fabricated deck. This deck had nine delaminated areas, two corroded rebar mats, and four vertical cracks built into it. The second test deck was removed from a distressed highway bridge along Interstate 10 near El Paso, Texas, and taken to the lab. All participants were asked to submit the analyzed data from both the laboratory and field testing no later than two weeks after testing was completed (Gucunski et al. 2013).

After the testing was completed, the data were analyzed and two statistical performance measures were considered to help rank the methods. Cost performance measures were also analyzed and are discussed in Section 2.4. The statistical measures were accuracy and repeatability. Each measure was given a rank of 1 (not favorable), 3 (favorable), or 5 (very favorable) for each method. The ultimate goal of this study was to develop a computer bank of commonly used NDE methods that can be used by bridge deck inspectors.

This computer program will have will have an excess of information about these methods, including a description, the physical principle of the method, applications, performance, limitation, equipment needed, test procedures, and samples of data output (Gucunski et al. 2013).

Accuracy was judged on three criteria: detectability extent, detectability threshold, and severity of deterioration. Detectability was considered the most important parameter because if a certain effect cannot be detected, the other four measures are meaningless. Detectability is the ability for a method to detect a flaw and not report an intact location as defective (meaning false-positives and false-negatives should be minimized). Based on the results of the tests, these methods were given an average grade. A representation of the accuracy for the test methods included in the Delphi survey can be seen in **Figure** (Gucunski et al. 2013).

The repeatability of a test was also examined. One approach to measure repeatability was to use the COV of each method. This was not used, however, because not all of the results from each test can be used to determine COVs and some of the participants submitted raw data that could not be used to calculate the COV. Note that the COV was calculated for impact echo, electrical potential, and radar and all values were less than 0.25, indicating these tests were relatively reliable. Instead of using COV values, grading of the repeatability of each test was based on graphical presentation of the results. While this was somewhat subjective, it provided a more comprehensive analysis. A representation of the repeatability for the test methods included in the Delphi survey can be seen in Figure 2.7.

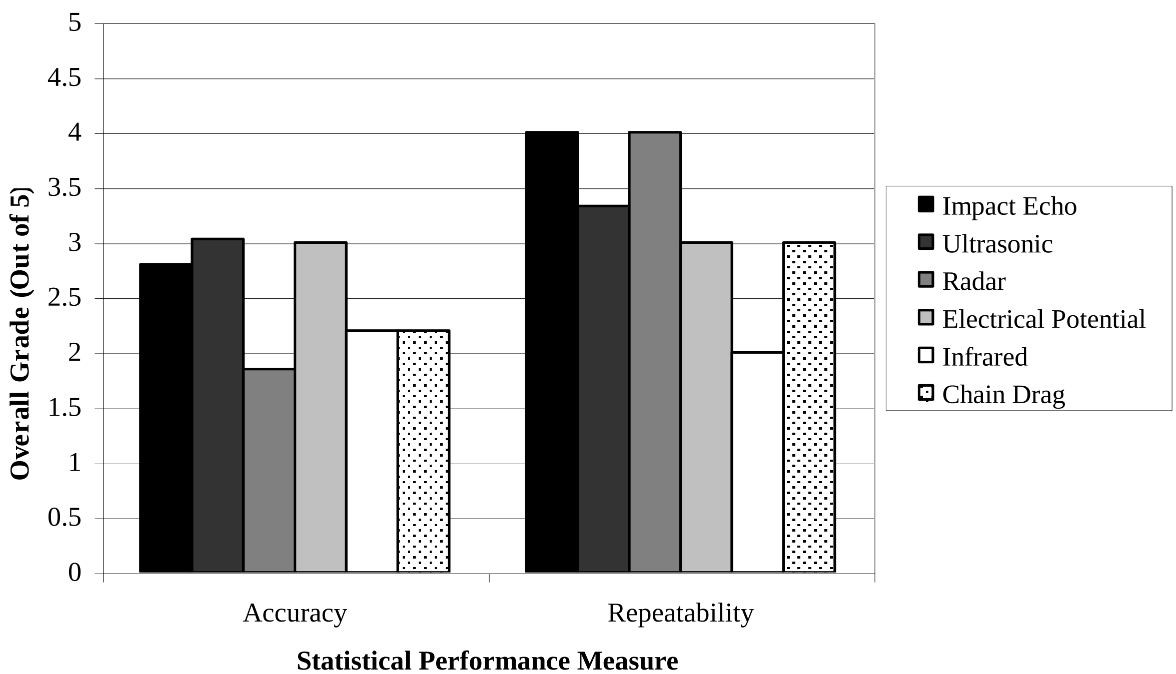


Figure 2.7 Statistical Performance Grades of Each NDE Method Performed by Gucunski et al. 2013

As shown in Figure 2.7, impact echo, electrical potential, and ultrasonic methods tended to have more accurate measurements with ground penetrating radar, infrared, and chain drag being slightly less accurate. Note that chain drag was not very successful at detecting defects that were relatively small or deep. There was also some concern with the infrared imaging test as testing was environmentally dependent and could only be done in a small window during the day. The rebar corrosion tests could be skewed because the methods were developed to measure corrosive activity (the environment in the concrete that promotes corrosion) and not corrosion itself. Furthermore, most tests were relatively repeatable. Infrared was the only method with less than satisfactory results. This was probably due to the method's susceptibility to the environment, including debris, shadows, markings, and time of day (Gucunski et al. 2013).

2.4 Comparative Studies for Various Costs of NDE Methods

While most studies found during the literature review were conducted as attempts to compare the results of various NDE methods, some studies also measured various costs of these methods. These comparisons were both qualitative and quantitative in nature and offer a limited insight to the various costs of common NDE methods. The following is a discussion of these studies.

2.4.1 Comparing Various NDE Methods on Deteriorating Concrete Bridges – A Rens et al. 2005 Report

As mentioned previously, Rens et al. (2005) published a report explaining the use of the bridge management system for the city and county of Denver. In this report, Rens et al. compared various tests that can be used on concrete to determine various defects. These methods were also compared on a relative cost basis. The relative cost was either determined to be high or low. It was shown that acoustic emission, radar, radiography, and thermography had relatively high costs, while electrical methods, impact echo, magnetic methods, mechanical sounding, surface hardness methods, acoustic tomography, and ultrasonic had relatively low costs.

2.4.2 Comprehensive NDE Concrete Bridge Deck 2013 Study by the Transportation Research Board

The comprehensive study published by Gucunski et al. (2013) also compared various factors that could affect the cost of a method. As mention previously, six NDE methods that were analyzed were included in the Delphi survey conducted for this research. Only these methods will be discussed here. Again, ten organizations used two different methods of validation testing (field and laboratory testing). After the testing was completed, the data were analyzed and three cost performance measures were considered to help rank the methods. These measures were 1) ease of data collection, analysis, and interpretation, 2) speed of data collection and analysis, and 3) cost of data collection and analysis (Gucunski et al. 2013).

The first cost performance measure that was analyzed was ease of use. For this measure, seven components were considered: expertise in data collection, number of operators, ease of maneuvering, physical effort for the setup, expertise in data analysis, and potential for automation. The grades for these components were based on both information provided by participants and observations by the research team. These grades were combined to create one ease of use grade for each method (shown in Figure 2.8) (Gucunski et al. 2013).

Speed was another performance measure that was analyzed. For this, there were two main components that classified speed. The first was the speed of data collection. Some methods collect data continuously while others collect data points. Therefore, data collection speed was determined by the area covered per hour of collection. The other component was speed of data analysis. This was defined as the time it took to process raw data into usable data. These values were combined to create one speed grade for each method, as shown in Figure 2.8 (Gucunski et al. 2013).

The final performance measure that was considered was the monetary cost of each method. For the cost measure, participants were asked to provide a cost estimate for bridge decks with an area of 5,000 ft^2 and 10,000 ft^2. The components considered for this measure were the cost of data collection and the cost of data analysis and interpretation. Grades were assigned to the cost based on a unit cost. The grades for data collection and analysis were then combined to create one cost grade for each method (shown in Figure 2.7) (Gucunski et al. 2013).

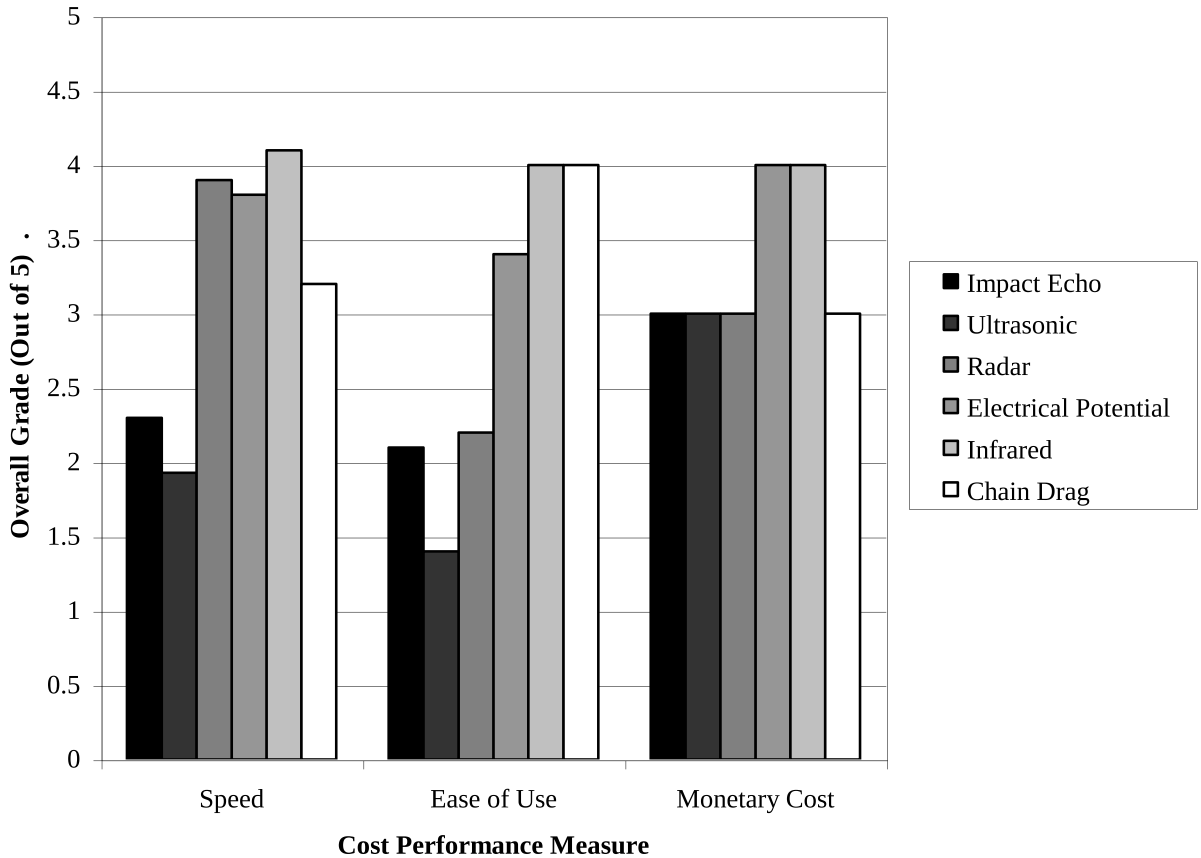


Figure 2.8 Cost Performance Grades of Each NDE Method Performed by Gucunski et al. 2013

The speed and ease of use measures were based in a similar scale as the statistical measures mentioned previously (1 indicated not favorable, 3 indicated favorable, and 5 indicated very favorable responses). As shown, infrared, chain drag, and electrical potential tended to be relatively easy to use, while the rest of the methods had various reasons that made them harder to use. Furthermore, radar, infrared, chain drag, and electrical potential tended to be relatively quick, while impact echo and ultrasonic tended to be slower. A different scale determined the monetary cost measure. Methods with costs less than $0.5/ft^2 were given a grade of five. The grade was decreased for every additional increase of $0.25/ft^2. As shown, all methods tended to be fairly cost effective with infrared and electrical potential being the most cost effective (Gucunski et al. 2013).

2.5 Previous Surveys

Four previous surveys on the use of NDE methods on highway structures were discovered during the literature review. Relevant findings from these surveys were used to form the framework of the Delphi method survey conducted as part of the present project. The previous surveys included a 1993 study by Rens, et al. for the American Association of Railroads, a 1994 unpublished study by the California

Department of Transportation (CALTRANS), a follow-up study by Rens and Transue in 1996, and a 2001 FHWA survey.

2.5.1 Rens et al. 1993 Survey

In 1993, Rens et al. conducted an international survey on general NDE use. A total of 96 surveys were sent. Of these surveys, 50 were sent to state DOTs and eight were sent to domestic industry organizations. The return rate was approximately 90% for these domestic surveys. Furthermore, a total of 38 surveys were sent to international public works organizations. The response rate for the international surveys was approximately 10%. The survey questions focused on what methods were being used in bridge inspection programs, the applications of these methods, and the qualifications of the personnel (Rens et al. 1997).

The key results from the survey used for this study were the types of methods being used in bridge inspection programs. The results of the domestic responses of the NDE techniques most commonly used are summarized in Table 2.2 at the end of Section 2.5.2. Note that the use of visual inspection or visual testing was not questioned. The three most common methods were ultrasonic testing, magnetic particle testing, and liquid penetrant testing. These results compare very well to the results from the CALTRANS 1994 survey (Table 2.3), as expected (Rens et al. 1997).

2.5.2 CALTRANS 1994 Survey

In 1994, CALTRANS conducted an unpublished survey of state departments of transportation (DOTs). The CALTRANS survey as described by Moore et al. had 37 state DOTs respond to nine questions about NDT (2001). Note that the questions were specifically about NDT, not NDE. The questions on this survey focused on what methods were being used in the bridge inspection programs, the procedures for these programs, the personnel performing the test, and the qualifications of the personnel (Moore et al. 2001).

Moore et al. present all the CALTRANS survey results, but only the key results from the survey for the purposes of this research are outlined. These results were the types of methods being used in bridge inspection programs. The survey asked what NDT methods were currently being used in state DOT bridge inspection programs. If only visual inspection was used participants were asked to note that. A summary of the responses can be seen in Table 2.3 at the end of this section. Table 2.3 shows that the most common methods were ultrasonic testing, liquid penetrant testing, visual testing, and magnetic particle testing (Moore et al. 2001).

Note that while this question asked about general NDT use, it was implied that all participants used visual inspection. However, the questions were compiled in the form of visual testing, not visual inspection. Confusion about what was included with visual inspection is probably why visual testing was listed less frequently than other methods, such as ultrasonic testing or liquid penetrant testing. The remaining questions gave results about the various procedures for the bridge inspection programs, the personnel performing the test, and the qualifications of the personnel but are not highlighted in this paper, as they did not directly pertain to the scope of the research (Moore et al. 2001).

Table 2.3 Summary of CALTRANS (Moore et al. 2001) NDT Survey Question 1: NDT Methods Currently in Use and Rens et al. (1997) NDE Survey Question: Domestic NDE Methods Currently in Use

NDE Method	Number of Responses from Moore et al. 2001 (37 total state DOTs)	Number of Responses from Rens et al. 1997 (52 total)
Ultrasonic Testing	26	37
Liquid Penetrant Testing	25	21
Visual Testing	19	-
Magnetic Particle Testing	17	13
Radiographic Testing	5	6
Acoustic Emission	2	-
Eddy Current Testing	1	6
Radar Locator	-	6
Schmidt Hammer	-	6
Contract Out NDE Work	-	6
Do Not Use NDE Techniques	-	5
Voltmeter	-	4
Other	-	7

2.5.3 Rens and Transue 1996 Follow-Up Survey

Rens and Transue conducted a follow-up survey to the 1993 Rens et al. survey in 1996. The same respondents from the 1993 survey were sent the new questionnaire. There was an 86% response rate. The purpose of this survey was to identify what information users seek during a bridge evaluation and what aspects of the bridges were deemed difficult to test. The results indicated that bridge decks were the most difficult bridge element to test. There was also an indication that determining the location of flaws was also difficult. The results also showed that for concrete structures, approximately 74% of respondents used NDE methods to determine reinforcement details, while approximately 84% of respondents used NDE methods to determine crack location and extent of cracking in steel structures (Rens and Transue 1998).

2.5.4 FHWA 2001 Survey

In 2001, the FHWA published a survey through the NDE Validation Center at the Tuner-Fairbank Highway Research Center. The survey focused on evaluating the current policies that might influence the reliability of visual inspection. The study focused on three main objectives: 1) developing a state-of-practice report for bridge inspection, particularly visual inspection, 2) gathering information on bridge inspection management, and 3) gathering data about the current use of NDE technologies (Moore et al. 2001).

The survey was sent to 52 FHWA state division bridge engineers, 99 Iowa county DOTs, and 15 bridge inspection contractors. Of these surveys, there were 42 state DOT responses (81%), 72 county DOT responses (73%), and six contractor responses (40%). This resulted in a combined response rate of 72% (Moore et al. 2001).

The key results from the survey for the purposes of this research were the current NDE techniques being used. This question was asked for steel, concrete, and timber bridges, but only steel and concrete NDE techniques will be outlined here. For steel bridges, the most common NDE techniques were visual inspection, liquid penetrant testing, ultrasonic testing, and magnetic particle testing. The results for steel NDE techniques can be seen in Table 2.4. Note that methods having minimal results (less than 10% of state DOT responses) were omitted from this table.

Table 2.4 Summary of NDE Techniques Used on Steel Structures from 2001 FHWA Survey (Moore et al 2001)

Steel NDE Technique	State DOT	County DOT	Contractors
Visual Inspection	95%	64%	100%
Liquid Penetrant Testing	81%	3%	67%
Ultrasonic Testing	81%	0%	67%
Magnetic Particle Testing	64%	0%	67%
Radiographic Testing	17%	0%	17%
Acoustic Emission	12%	1%	33%

For concrete bridges, the most common NDE techniques were visual inspection, mechanical sounding, cover meter, and rebound hammer. The results for concrete NDE techniques can be seen in Table 2.5. Note that methods having minimal results (less than 10% of state DOT responses) were omitted from this table.

Table 2.5 Summary of NDE Techniques Used on Concrete Structures from 2001 FHWA Survey (Moore et al. 2001)

Concrete NDE Technique	State DOT	County DOT	Contractors
Visual Inspection	90%	64%	100%
Mechanical Sounding	76%	43%	67%
Cover Meter	50%	0%	33%
Rebound Hammer	45%	13%	33%
Electrical Potential Measurements	26%	0%	33%
Radar	21%	0%	17%
Impact Echo	19%	0%	17%
Thermal/Infrared	12%	1%	17%

There were also questions asked about experience level and number of bridges tested. While these results were not directly used, they served as a framework in developing similar questions for the Delphi method survey that was conducted (Moore et al. 2001).

2.6 The Delphi Method

2.6.1 Background of the Delphi Method

The Delphi method was originally developed in the 1950s by Olaf Helmer and associates at the RAND Corporation (Yousuf 2007). The method is defined as “a group process involving an interaction between the researcher and a group of identified experts on a specific topic, usually through a series of questionnaires” (Yousuf 2007). The process is useful to gather opinions on complex topics when exact information is unavailable, making it a good tool to gather quantitative information of NDE methods based on expert opinions.

Originally, the Delphi method was used as a forecasting technique to predict the probability of future events (Yousuf 2007). Since then the method has been used for various reasons, including investigating the implication of historical events, determining possible budget distributions, planning curriculum, and determining potential policy options. These are not the only applications of the Delphi method;

determining the appropriateness of the method is not always clear (Linstone and Turoff 2002). Due to this uncertainty, Linstone and Turoff suggest the technique can be used when one or more of the following situations occur (2002):

- The problem is difficult to accurately study analytically but lends itself to be analyzed from subjective judgments
- The individuals needed to contribute to the study do not have a history of satisfactory communication and may have different backgrounds
- More individuals are needed than are feasible to interact face to face, making costs such as time and money for regular group meetings impractical
- Efficiency can be increased through supplemental communications rather than face-to-face meetings
- The communication process needs to be monitored and/or kept anonymous due to the strong disagreements of individuals
- The avoidance of a dominant quality of an individual or group is needed

The overall goal of the method is to reach a consensus within a group of experts (Okoli and Pawlowski 2004). This can be done by using a sequence of questionnaires to collect data and opinions from the group of experts. The process utilizes several iterations to provide feedback to the participants. This feedback allows the participants to reconsider their original opinion. Consequently, the results from previous iterations can change or be adapted by individual participants in later iterations based on the feedback of the group. With this feedback loop, the Delphi method attempts to reach a consensus within the group (Hsu and Sanford 2007).

The primary characteristic of the Delphi method is participant confidentiality, which is achieved through the use of mail or email to exchange information. This aspect of the method is designed to reduce the effects of dominant participants, which is often a concern in group-based methods such as brainstorming conferences. With this, certain adverse aspects of face-to-face participation, such as manipulation of participants or conformity to the group, can be reduced. Along with this, by conducting the method by mail or email, the cost of travel time and expenses are eliminated (Hsu and Sanford 2007).

The controlled feedback process for this method is designed to remove noise. Noise can skew the results that occur when the participants focus on group and/or individual interests rather than focusing on problem solving. The feedback process consists of a representation of the prior iteration making it so each participant can see the opinions of the entire group. This allows each participant to make additional conclusions and clarify the information from previous iterations based on these results. Through this process, the participants tend to become better problem solvers and focused on making their opinions more insightful. This minimizes the noise in the responses (Hsu and Sanford 2007).

2.6.2 The Delphi Method Process

The Delphi method is an iterative process until a consensus of experts' opinions has been achieved. To do this, multiple rounds of mailed or emailed surveys are sent to the participants (Hsu and Sanford 2007). While the process could be continuously implemented, Hsu and Sanford show the process takes about three iterations in order to achieve consensus. The following discussion outlines the series of iteration rounds (2007).

2.6.2.1 Round One of the Delphi Method

In the first round of the Delphi method an open-ended questionnaire is used. The open-ended questionnaire serves as the basis of the Delphi method. The questions are often developed from literature reviews or past surveys. The purpose of this questionnaire is to gather the type and level of expertise of the respondents and specific information about the topics in question. An example of this would be a question that asks the respondents if they use a specific NDE test (i.e., visual inspection) and, if so, how often. This question would give the investigators two types of information: how often the method is used and how experienced the individual is with the method. After receiving the respondent's answers, the information is compiled and organized. These results are then used to develop the second round of the survey (Hsu and Sanford 2007).

2.6.2.2 Round Two of the Delphi Method

The second round of the survey is developed by using responses collected from the open-ended questionnaire in the first round. These responses are used to develop closed-ended questions that require the participants to rank and order specific responses developed by the surveyors. In some cases participants are asked to provide rationale for their responses. The responses are then compiled into a review sheet that summarizes the responses of all the respondents (Hsu and Sanford 2007).

2.6.2.3 Round Three and Subsequent Rounds of the Delphi Method

In the third and subsequent rounds, a similar (or often the same) survey from round two is sent to participants along with the summary sheet of all the responses. The participants are then asked to review the summary sheet and answer the questions again based on their prior opinions and these results. These rounds give the respondents an opportunity to revise and change their responses based the overall responses of the group. They are also given the opportunity to specify reasons if they chose to remain outside the consensus. This process is then repeated with all the respondents' responses until it is determined a consensus is reached. On average, this process takes about four total rounds (Hsu and Sanford 2007).

2.6.2.4 Determining Participants

Determining appropriate participants is the most important step in the entire process of the Delphi method. The expertise of the participants directly relates to the quality of the results. However, there are no current standards of selecting participants for a Delphi method survey. It is a general criterion for surveyors to consider individuals who have backgrounds and experiences concerning the survey topic. Participants who are capable of critical thinking, providing helpful inputs, and willing to revise their initial judgments in order to reach a consensus are also sought (Hsu and Sanford 2007).

2.6.2.5 Determining Size of a Delphi Method Study

Witkin and Altschuld (1995) note that the approximate size of a Delphi study should be under 50, while Ludwig (1997) shows that most studies use between 15 and 20 participants. Concerning the response rate of the target group, Moore et al. (2001) showed that a response rate of 81% could be achieved from surveying 52 state DOTs, and a response rate of 73% could be attained from surveying 99 Iowa county DOTs. It was also shown, however, that a response rate of 40% was achieved from surveying 15 NDE contractors. The combined response rate of these groups was 72%. As shown, this rate is slightly skewed due to the amount of DOT participants relative to the amount of contractors surveyed.

2.6.2.6 Determining When Consensus and Stability is Reached

Data analysis of the results from each round after the first round can be employed to determine the agreement and stability of the results for each question. Agreement can be shown using various methods. Hsu and Sanford (2007) show that agreement can be determined if a certain percentage of responses falls within a set range. There are various opinions as to what percentage is needed. Hsu and Sanford (2007) also show that consensus is met when 80% of the responses fall within two points on a seven-point scale. Likewise, English and Kernan (1976) show that the COV can be used to determine agreement by evaluating the COV of each question for each round in conjunction with a decision rule of predetermined selected ranges. While the selected ranges are arbitrary, English and Kernan (1976) developed reasonable cutoff ranges as seen in Table 2.6.

Table 2.6 Selected Ranges of Coefficient of Variation Used to Determine Agreement (English and Kernan 1976)

Coefficient of Variation (COV)		Decision Rule
	0 < COV ≤ 0.5	Good degree of consensus; no need for an additional round
	0.5 < COV ≤ 0.8	Less than a satisfactory degree of consensus; possible need for an additional round
	COV > 0.8	Poor degree of consensus; definite need for an additional round

While agreement is an important measure, it is also important to measure the stability of each response from round to round. Stability is a representation of how much the responses change from one round to the next. Kalaian and Kasim (2012) present various parametric (absolute COV difference, F-ratio, Pearson Correlation Coefficient, and Paired t-test) and nonparametric (McNemar Change Test, Spearman's Rank Correlation Coefficient, and Wilcoxon Paired Signed-Ranks t-test) methods to determine stability. It is shown that the various methods can be used for specific circumstances. These conditions depend on, among other things, the type of data collected (Likert scale, dichotomous, etc.), number of people in the study, and the distribution of the data.

2.6.3 Previously Conducted Delphi Method Surveys in Civil Engineering

There have been a variety of studies in civil engineering that have used the Delphi method to gather information. The majority of these studies have been on the topic of management or planning. These studies asked experts their opinions in specific topics and how to handle specific situations. Two examples of these studies are Yasamis-Speroni et al. (2012) and Gad and Shane (2012). Both these studies looked at various factors that affected certain decision processes in management. All these studies produced qualitative information with no way of developing this information into quantitative data. Another study (Saito and Sinha 1991) used the Delphi method to study bridge condition ratings by inspectors. Again, this study produced completely qualitative results. These results indicated that more unified criteria and guidelines needed to be established to produce consistent bridge ratings.

3. IMPLEMENTING THE DELPHI TECHNIQUE AS A SURVEY METHOD TO GATHER EXPERT OPINION

3.1 Selection of the Delphi Method

Since it was determined that there is limited work being done to quantify the level of accuracy in the tests or to compare various tests to one another, a comprehensive survey to gather expert opinion was identified as a means to gather the desired information. It was established that the Delphi method was an efficient and effective survey technique to gather this information. This survey aims to provide quantitative descriptions of accuracy, reliability, and bias (in terms of statistical descriptions) and a comprehensive comparison of the various tests to provide information to researchers and practitioners working in the fields of bridge management and inspection.

3.2 Determining Participants

Prospective participants of the survey were determined through an Internet search of current DOT employees who were involved in bridge design and evaluation. The directory of the AASHTO Subcommittee on Bridges and Structures was also used to determine DOT employees who have experience with NDE techniques. An Internet search was also conducted to determine private companies in the United States that work in the NDE field. These prospective participants were then contacted by mail and email and asked to participate. A total of 36 DOTs were contacted. Not all 50 states were included due to difficulty of obtaining the required contact information for certain state DOTs and employees. A total of 27 private companies from around the country were also contacted. These companies all primarily work in the NDE field and have experience conducting tests on bridges. It should be noted that all prospective participants were contacted by mail. Of these, 25 (all being DOTs) were also emailed. The lower number of emailed participants was due to lack of provided email addresses. All surveys were sent to individuals who were deemed experienced with NDE methods when possible. Determining individual people with extensive NDE experience at a specific DOT was often difficult. To attempt to get a better response rate, the survey was sent to heads of structural engineering departments when experienced individuals could not be identified. They were asked to complete the survey or pass it along to someone who they felt was knowledgeable in NDE methods.

The number of participants contacted was determined in part by the response rate of the FHWA 2001 survey. This survey showed that about an 81% response rate could be achieved from surveying DOTs while about a 40% response rate could be achieved from surveying private contractors. Based on these results and the recommended survey size by Witkin and Altschuld (1995) and Ludwig (1997), it was determined about 60 possible participants should be contacted. This number was determined in conjunction with the number of DOT contact information that could be determined. Since only 36 DOTs had contact information that was easily accessible, it was determined the remaining number would be made up by private contractors. Prior to implementing the survey, the procedure and a description of the possible participants were submitted for review to the Institutional Review Board (IRB) and approved by the IRB for implementation.

3.3 Round One Questionnaire

A total of 63 people were contacted and asked to participate in the survey. They were mailed a packet, which included the six-page questionnaire, a cover letter that explained the survey and acted as the release form, and a self-addressed stamped return envelope to return the survey. The people who were also emailed were sent the questionnaire and cover letter. The cover letter included a description of the Delphi method, the goal of the survey, a description of the first round of the questionnaire, an explanation of the confidentiality of the survey, and the implied release of any known risks. A copy of the cover letter can be seen in Appendix A. A brief reminder letter was sent a week before the deadline to help improve the response rate. The following is a discussion of the first questionnaire. The questionnaire can be seen in Appendix A.

3.3.1 Section One of the First Questionnaire

Prior to completing the first section of the questionnaire, participants were given a detailed description of the questionnaire and the survey process as a whole. They were also asked to include contact information for the surveyors to identify the respondent of this and subsequent questionnaires. The first section of the questionnaire then asked participants about their background and general experience with NDE. They were asked their current education level, current NDE certification level, how long they have been working with NDE, the types of tasks they perform when working with NDE, and the number of bridges their organization and each participant individually evaluates in a given year. Note that the private contractors were also asked in what geographic region they perform NDE. It was implied that the DOT personnel only perform NDE in their respective state.

3.3.2 Section Two of the First Questionnaire

Section two dealt with various NDE methods for steel bridges. Participants were given a list of common NDE methods that are used on steel members. This list was developed from the literature review, the CALTRANS 1994 survey, and the FHWA 2001 survey. The methods in the list were acoustic emission, liquid penetrant testing, thermal/infrared, visual inspection, eddy current, magnetic particle testing, ultrasonic testing, radiography, and vibration analysis. Space was provided to note any test the participants commonly used but were not listed. Respondents were asked to list the types of conditions their organization sought to identify with each technique. If their organization did not use a specific method they were asked to leave the space blank. They were also asked to identify each method from the list their organization used at least once every month. If they did not use a specific technique at least once a month, they were asked to indicate which two methods they used the most. There were two purposes for the questions in this section. The first was to develop a list of the most widely used NDE methods based on the responses of the participants. The second purpose was to compile a list of common conditions that were tested for each NDE method.

3.3.3 Section Three of the First Questionnaire

The third section of the survey had a similar purpose and questions as the second section; however, these questions dealt with concrete bridges. Again, participants were given a list of common NDE methods but this time the methods were for concrete members. This list was also developed from the literature review, the CALTRANS 1994 survey, and the FHWA 2001 survey. The methods in the list were acoustic emission, mechanical sounding, rebound hammer, impact echo, cover meters/pachometers, radar, thermal/infrared, vibration analysis, electrical potential measurements, radiography, ultrasonic testing, and visual inspection. Respondents were again asked to list the common conditions that each method was used to identify and to list their organization's commonly used methods. Following this section, participants were also asked if they had ever stopped using any NDE methods on bridges in the past and, if so, to explain why. Note that respondents were also asked if they would like to be contacted with the subsequent questionnaires by mail or email. This question was used to attempt to increase the response rate by ensuring all subsequent correspondence would be by the participant's preferred means.

3.3.4 Questionnaire One Response Rate

The participants were given about a month to complete and return the first questionnaire. In order to help the response rate, each survey was written to take an estimated 20 minutes to complete. Also, a reminder letter was sent to all possible participants who had not responded about a week before the deadline. Of the 36 DOTs contacted, 11 responded to the survey. Moreover, of the 27 contractors contacted there were three responses. This resulted in a 22% response rate (31% from DOTs and 8% from contractors). While the response rate was lower than expected, the number of participants was deemed acceptable based on Ludwig's recommendation (1997).

There are various possible reasons for the low response rate. One reason could be busy schedules of the contacted participants. While the survey was written to take 20 minutes, the burden of the possibility of completing four to five questionnaires may have been too much time for some people to spend, which is a common limitation to the Delphi method. Another reason could have been miscommunication between department personnel. As mentioned previously, if specific people with extensive NDE experience at a DOT could not be determined, department heads were contacted. These department heads could have neglected the survey due to busy schedules, the survey could not have been forwarded to the correct personnel in a timely fashion, or the survey could have gotten lost in the process. Another reason could be an individual simply not wanting to complete a survey.

3.4 Round Two Questionnaire

All 14 people who responded to the first questionnaire were contacted and asked to participate in the second. Of these 14 people, 11 were from state DOTs and three were from private companies. These people were contacted either by mail (two people) or email (12 people) as indicated by their response to the last question in the first questionnaire. They were sent the 16-page questionnaire, which included directions on how to complete and return it. This questionnaire was composed of two sections. The sections were similar as they both dealt with questions about specific NDE methods. The first section was for NDE methods on concrete while the second section was for NDE methods on steel. Each section contained five subsections. These subsections included bias, accuracy (the tendency of a test to measure true results), precision (the reproducibility of a test in a controlled environment), reliability (the reproducibility of a test in an uncontrolled environment), and various costs of each method. It should be noted that all these definitions were developed only for the purposes of this survey. The methods asked about were determined from the responses of the first survey and results from the FHWA 2001 survey. The mailed questionnaires also included a self-addressed stamped return envelope. Again, a brief reminder letter was sent a couple days before the deadline. Another reminder letter was sent a few days after the deadline as not all participants had responded. The following is a discussion of the second questionnaire. The questionnaire can be seen in Appendix A.

3.4.1 Subsection One of the Second Questionnaire: Bias

The first subsection was used to determine how biased each test could be relative to the true value. Bias was defined as the tendency of a test to consistently measure either higher or lower than the actual or perceived value. Participants were given a bias scale from 1 to 11 to use when answering the questions. This scale can be seen in Figure 3.1.

-50% or More	- 40%	-30%	-20%	-10%		+10%	+20%	+30%	+40%	+50% or More
1	2	3	4	5	6	7	8	9	10	11

LOWER THAN TRUE VALUE ↑ TRUE VALUE HIGHER THAN TRUE VALUE

Figure 3.1 Representation of the Scale Used for the Bias Subsection

For this scale, a response of six represented the true value. Any incremental responses lower than six indicated an extra 10% bias from the true value (i.e., a response of five would have a bias of 10% lower than the true value, a four would have a bias of 20% below to the true value and so on). This same relationship was represented on the responses larger than six. Not all methods were included in this subsection. If a method was determined to not give a quantitative result, that method was not included in this section. The methods questioned were: cover meters/pachometers, impact echo, radar, ultrasonic testing, and visual inspection for concrete members and acoustic emission, radiography, ultrasonic testing, and visual inspection for steel members. These methods were not included in the accuracy subsection because they can produce quantitative results. It should be noted that ultrasonic testing for steel members was broken into three sections: crack detection, pin inspection, and weld inspection. This was done because, unlike other methods, this method is consistently used to identify all three of these defects, rather than just a single defect. In doing this, the test for each defect could have a different accuracy or bias.

3.4.2 Subsection Two of the Second Questionnaire: Accuracy

The second subsection was used to determine how accurate each test could be relative to the correct identification of the condition. This subsection only included methods that were deemed to have qualitative results. For the purposes of this survey, accuracy was defined as the tendency of a test to measure true results. Participants were given three options: false positive, false negative, and true response. False negative was defined as a test that measures no damage, but there is damage; false positive was defined as a test that measures damage, but there is no damage. A true response was defined as a test that measures damage and there is damage or a test that measures no damage and there is no damage. Participants were asked to estimate the percentage of time each test would have each result. They were told their percentages should add up to 100%.

3.4.3 Subsections Three and Four of the Second Questionnaire: Precision and Reliability

The third and fourth subsections were used to determine the precision and reliability of each method and included all the methods in question. For the purposes of this survey, precision was defined as the reproducibility of a test in a controlled environment (i.e., a lab setting) while reliability was defined as the reproducibility of a test in an uncontrolled environment (i.e., in the field). Participants were given a scale based on hypothetical means and standard deviations. They were also given COVs corresponding to these numbers and a graphical representation of the corresponding distribution, which was assumed to be a normal distribution. The same scale was used for both precision and reliability and can be seen in Figure 3.2. Participants were asked to indicate either the reliability or precision for each method based on the scale provided.

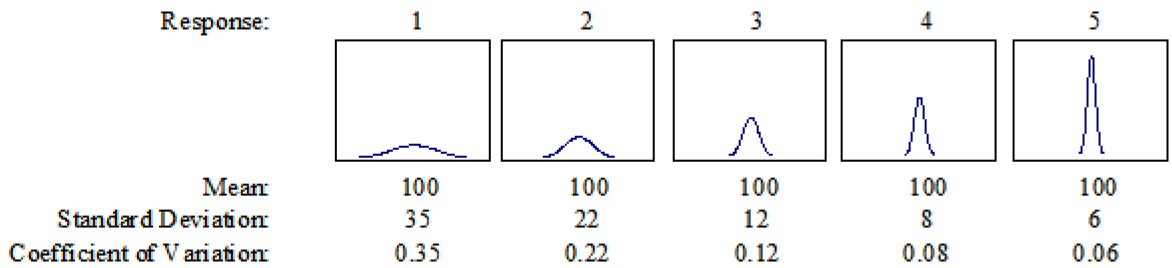


Figure 3.2 Representation of the Scale Used for the Precision and Reliability Subsections

3.4.4 Subsection Five of the Second Questionnaire: Costs

The fifth and final subsection of the second questionnaire pertained to various costs for each method. The costs for each method were: time spent running a test, time spent analyzing data, time to train an inspector, monetary cost for the equipment, and number of inspectors needed. For each cost, the participants were asked to develop a scale of five ranges. An example of this scale can be seen in Figure 3.3.

	Very Low	Low	Moderate	High	Very High
Scale	0 - 4 Hours	3 - 8 Hours	8 - 10 Hours	10 - 15 Hours	15+ Hours

Figure 3.3 Example of a Personal Scale Made for Time Spent Running a Test

The participants were then asked to categorize each method based on the user-developed scale. The user-developed scale was implemented because it was possible two participants may have substantially differing opinions on what constitutes a "very low" or "very high" cost. This procedure was repeated for all five costs. It should be noted that due to the implementation of the user-developed scales, some responses were vastly different than others. To account for this, the data for each scale were compiled and compared. A standard scale was then developed for each cost based on the participant-developed scales. These scales were developed to include as many responses as possible while staying relatively close to the average response for each range. The scales were then implemented for questionnaires three and four.

3.4.5 Questionnaire Two Response Rate

The participants were once again given about a month to complete and return the second questionnaire. Again, in order to help the response rate, the questionnaire was written to take an estimated 20 minutes to complete. Also, a reminder letter was sent to all possible participants who had not responded about a week before the deadline and again a few days after the deadline. Of the 11 DOTs contacted, eight responded to the second survey. Moreover, of the three contractors contacted there were two responses. While the reasons for most participants to discontinue their participation are unknown, one respondent had to drop out due to an increased workload and lack of time. While the response rate was lower than expected, the number of participants was still deemed acceptable.

3.5 Round Three Questionnaire

All 10 people (eight DOT and two private companies) who responded to the second questionnaire were contacted and asked to participate in the third questionnaire. They were sent the 14-page questionnaire, which included directions on how to complete and return the questionnaire. Also, included with this questionnaire was a results packet that contained the individual's response (a unique response packet was used for each participant) along with the average group response for the accuracy, bias, and reliability subsections. This questionnaire was nearly identical to the second questionnaire and participants were asked to complete the questionnaire in conjunction with the results packet. The goal was for the participants to iterate their response based on their prior response and the average group response.

While the survey was nearly identical to the previous survey, there were a few minor changes. The first change was the removal of the precision subsection. Based on the results of the second questionnaire, it was shown that precision and reliability were nearly identical. To shorten the survey to help keep the response rate high, the precision subsection was removed. The second change was the inclusion of predetermined scales based on prior responses for the costs subsection. These scales were developed to include as many of the responses as possible while staying relatively close to the average responses. Participants were asked to rate each method based on these scales. By doing this, the responses were much more uniform relative to survey two.

3.5.1 Questionnaire Three Response Rate

The participants were once again given about a month to complete and return the third questionnaire. Also, a reminder letter was sent to all possible participants who had not responded about a week before the deadline and again a few days after the deadline. Of the eight DOTs contacted, seven responded to the third survey. Moreover, both contractors responded to the survey. While the reasons for most participants to discontinue their participation are unknown, there was again some indication that increased workload and amount of time needed to complete the survey were a concern. Note that the participant who discontinued participation was not a significant outlier relative to the average group response.

3.6 Round Four Questionnaire

All nine people (seven DOT and two private companies) who responded to the third questionnaire were contacted and asked to participate in the fourth questionnaire. They were sent the 10-page questionnaire, which included directions on how to complete and return the questionnaire. Also, included with this questionnaire was a results packet that contained the individual's response (a unique response packet was used for each participant) along with the average group response for the accuracy and costs subsections from the third questionnaire. This questionnaire was nearly identical to the second and third questionnaires, and participants were asked to complete the questionnaire in conjunction with the results packet.

While the survey was nearly identical to the previous survey, there were a few minor changes. The bias and reliability subsections were removed from this round. Based on the results of the third questionnaire, it was shown that the responses from these subsections had converged and became stable so no further questioning was needed. The removal of these subsections also helped to shorten the survey. This meant only the accuracy and costs subsections were included in this questionnaire.

3.6.1 Questionnaire Four Response Rate

The participants were once again given about a month to complete and return the fourth questionnaire. Also, a reminder letter was sent to all possible participants who had not responded three days before the deadline and again a few days after the deadline. All nine participants responded to the survey.

4. RESULTS AND DISCUSSION

4.1 Introduction

A total of four Delphi method rounds were conducted in order to determine quantitatively the accuracy, reliability, bias, and various costs of common NDE methods. The first survey was employed to determine background information of the participants and common NDE methods for bridges. The second and subsequent surveys were used along with various statistical scales in order to develop quantitative information based on each method. The following is a discussion of the results of each survey.

4.2 Round One Questionnaire

The first questionnaire was used as a foundation for the second and subsequent questionnaires. The survey gave valuable information about the experience and certification level of all participants. Information about the types of methods being used and the flaws these methods were being used to detect was also gathered.

4.2.1 Certification and Experience Level

The certification and experience level information was sought after to ensure the participants could be considered knowledgeable in bridge NDE methods. According to the 14 original respondents, the average experience level of the participants with bridge NDE was 17.8 years with a maximum of 40 years and a minimum of five years. Most of these people were managers, but also assisted in data analysis, bridge inspection, and report writing. The most common education level was a four-year degree (10 participants). Moreover, two respondents had a master's degree and two respondents had a high school diploma. One person did not respond. It was also determined that an average organization tested about 2,000 bridges annually with an average respondent personally testing about 75 bridges per year. The certification level of the participants varied much more than the experience. Of the 15 original respondents, 73% of them possessed at least a professional engineering license. Along with this, three participants had an American Society for Nondestructive Testing (ASNT) NDT level II certification for at least one NDE method. Based on these results it was determined that all participants could be considered knowledgeable about bridge NDE methods.

4.2.2 Commonly Used NDE Methods

This section of the survey was used to determine the common methods that are currently in practice and the types of flaws being tested for. This was done both to compile a list of commonly used methods and to ensure the participants were knowledgeable with these methods. Participants were given a list of NDE methods for bridges (mostly compiled from the FHWA 2001 survey). They were asked to indicate what condition each technique they have experience with was used to identify or assess. Respondents were also asked to indicate which methods they used at least once a month on average or, if that was not applicable, to list the two most commonly used techniques they use. These questions were asked for both concrete and steel bridge members. The results for steel members can be seen in Table 4.1 and for concrete members in Table 4.2. Also included in these tables are the types of flaws that were commonly measured.

Table 4.1 Number of People Indicating Experience with Each Method for Steel Bridge Members

NDE Method	Frequency	Type of Flaw
Liquid Penetrant	12	Weld Imperfection, Crack Detection
Visual	12	General Flaws
Ultrasonic	12	Weld Imperfection, Crack Detection, Corrosion Detection, Thickness Measurement, Pin Inspection
Magnetic Particle	10	Weld Imperfection, Crack Detection
Radiography	7	Weld Imperfection, Crack Detection
Thermal	2	Deck Inspection
Acoustic Emission	1	Monitor Stay Cables
Eddy Current	1	No Response
Vibration Analysis	1	Force Measurement
Strain Gauges*	1	No Response

* Write in Response

Table 4.2 Number of People Indicating Experience with Each Method for Concrete Bridge Members

NDE Method	Frequency	Type of Flaw
Visual	12	General Flaws
Mechanical Sounding	10	Delamination
Cover Meters/Pachometer	8	Located Rebar, Determine Cover
Rebound Hammer	6	Test Compressive Strength
Thermal	5	Delamination
Impact Echo	4	Determine Thickness, Delamination
Radar	4	Located Rebar, Determine Thickness
Ultrasonic	4	Delamination
Acoustic Emission	3	Monitor Stay Cables
Electrical Potential	3	Detect Corrosion
Vibration	2	Force Measurement
Chloride Samples*	1	No Response
Radiography	0	-

* Write in Response

Based on the results from the Delphi survey and the FHWA 2001 survey, the methods for steel members that were removed from subsequent surveys were eddy current, thermal, vibration analysis, and strain gauges. The methods for concrete members that were removed from subsequent surveys were acoustic emission, vibration analysis, chloride samples, and radiography. Note there were various tests that were indicated to have stopped being used, but no reasons were given. Those methods were liquid penetrant (one person), ultrasonic (one person), and radiography (one person).

4.3 Determining Convergence and Stability

Determining when the responses have converged to a single value and the responses are stable is key when implementing a Delphi study. Convergence and stability are used to determine when the study should be terminated. Convergence is used to determine whether the responses are converging on a single value during a given round, while stability is used to determine the amount of change of the responses from one round to the next. English and Kernan's (1976) decision rule with the use of COV ranges was used to determine if a consensus was reached. All results with a COV lower than 0.5 were considered to be converging to the mean value. If a response was in a range from 0.5 to 0.8, the response was considered to be nearing convergence and was analyzed in more to detail to understand the trend. A response with a COV greater than 0.8 was not considered to be converging to a mean value. All questions that were considered to be converging were then analyzed to determine the stability of the response.

As shown by Kalaian and Kasim (2012), there are various parametric and non-parametric statistical methods to determine stability. Based on Kalaian and Kasim's recommendation, parametric methods should only be used if the subject group is larger than 30 and/or the responses have a normal distribution. Based on the response rates of each round and results from the surveys, it was determined that a non-parametric method should be employed. Of the non-parametric options, the McNemar Change Test could not be used because the results must be dichotomous (yes/no response). Similarly, the Wilcoxon Paired Signed-Ranks t-test could not be used because if there is no change in all responses from round to round the equation is unstable. Due to these limitations, Spearman's Rank Correlation Coefficient method was used to calculate stability.

4.3.1 Using Spearman's Rank Correlation Coefficient Method to Calculate Stability

To determine stability, Spearman's rho, r_S, must first be calculated using:

$$r_S = 1 - \frac{6\sum d_i^{\,2}}{n(n^2-1)} \qquad \text{Equation 4.1}$$

where, d_i is the difference between ranks of the respondents for the i^{th} question and n is the number of respondents. Note that due to people dropping out from round to round, n was used from the third survey when comparing round two to three. The rank correlation is then compared to a critical value determined from a table of critical values for Spearman's rho (Sheshkin 2004). If the calculated value is greater than the critical value, the response is determined to be stable. From this, the closer the value is to one the more stable it is and, conversely, a value close to zero indicates no stability. For this study a one-tailed level of significance of $\alpha = 0.05$ was used based on Kalaian and Kasim's recommendation (2012). Each question that was found to have converged to a value was then tested for stability. If the question was found to also be stable, the question was removed from subsequent rounds.

4.4 Results for the Bias of NDE Methods

It was determined, based on the convergence and stability analysis, that after round three all responses in the bias subsection had both converged and became stable and were removed from subsequent rounds. From this, the results of the third survey were considered to be the final values based on the responses from the nine respondents of that round. The response average, standard deviation, and COV can be seen in Table 4.3. Values from stability calculations can be seen in Appendix B.

Table 4.3 Concrete and Steel NDE Methods Response Statistics for the Bias Subsection

Concrete Methods			
	Average	Standard Deviation	Coefficient of Variation
Cover Meters/Pachometer	6.00	0.926	0.154
Impact Echo	5.40	0.894	0.166
Radar	5.57	1.272	0.228
Ultrasonic Testing	5.71	0.488	0.085
Visual Inspection	5.89	0.782	0.133
Steel Methods			
	Average	Standard Deviation	Coefficient of Variation
Acoustic Emission	5.50	1.000	0.182
Radiography	5.71	0.488	0.085
Ultrasonic Testing - Crack Detection	5.88	0.835	0.142
Ultrasonic Testing - Pin Inspection	5.89	0.782	0.133
Ultrasonic Testing - Weld Inspection	5.86	0.900	0.154
Visual Inspection	5.67	0.707	0.125

Based on the results, respondents felt that all the methods for concrete and steel bridges were slightly under-biased with an average response in the third round of all methods being between a response of 5 and 6 (5 being about 10% under-biased and 6 being the true value). All methods were also shown to have a good convergence (calculated COV less than 0.5) to the mean value.

Based on the limited data available for comparison, the survey results seem to show reasonable agreement to experimental findings. As shown by the Scott et al. (2003) study, when impact echo was performed at the location of a known flaw, the flaw was always detected. If the test was performed at predetermined grid locations (as is usually the case), the test tended to either suspect distress (indicating a possibility of a flaw) or missing it altogether. This seems to agree with the 5.4 average response of the participants. Furthermore, it was shown by Phares et al. (2001) that during a routine inspection, visual inspection of the superstructure, substructure, and deck had an overall bias of +3%, -5%, and +5%, respectively. Note a positive bias means the inspectors determined the bridge element was in better condition than it actually was. These numbers are close to (but slightly higher than) the numbers determined by the respondents. Conversely, Barnes and Trottier (2000) showed that radar tends to be slightly under-biased, which is reflected in the participant's responses with the responses indicating a little more bias than found in the study. Algernon's (2011) cover meter measurements with the ideal spacing and no layering of rebar produced very accurate results, which matches the participants' responses. However, if the parameters are changed, the test becomes more biased by indicating the cover is less than the actual measurement.

Based on this reasonable agreement, the data could be used to produce a bias factor. Table 4.4 shows how the response of the participants correlates to the bias of each test, or the bias factor (i.e., a bias factor of 1.0 means no bias, a bias factor of 0.9 means a response of 7 with a bias of +10% etc.). The bias factor was determined by fitting a trend line to the response and the bias representation (i.e., a response of 5 meant the test was under-biased by 10%). This factor could be used with an individual method's nominal value to give the inspector a more accurate representation of the true value. Therefore, multiplying the bias factor by the measured value would yield a more valid result. Note that not all methods were included in the bias subsection because not all methods provide quantifiable data.

4.5 Results for the Reliability of NDE Methods

As noted previously, responses for precision and reliability in round two were nearly identical so questions involving this parameter were combined into a single subsection just involving reliability for round three. With this, the following discussion is in terms of reliability.

It was determined that after round three all responses in the reliability subsection had both converged and became stable and were removed from subsequent rounds. Based on these calculations, the results of the third survey were considered to be the final values. The response average, standard deviation, and COV can be seen in Table 4.3. Furthermore, Table 4.4 shows how the response of the participants correlates to the reliability of each test in terms of a COV. Values from stability calculations can be seen in Appendix B. Based on the results, most methods had an average response between 3 and 4 (a response of 3 indicated a method with a COV of 0.12 and a response of 4 indicated a method with a COV of 0.08).

Table 4.4 Concrete and Steel NDE Methods Response Statistics for the Reliability Subsection

Concrete Methods			
	Average	Standard Deviation	Coefficient of Variation
Cover Meters/Pachometer	4.13	0.83	0.20
Electrical Potential	3.50	0.84	0.24
Impact Echo	3.40	0.55	0.16
Mechanical Sounding	4.11	0.78	0.19
Radar	3.71	1.11	0.30
Rebound Hammer	2.67	0.82	0.31
Thermal	2.60	0.89	0.34
Ultrasonic Testing	4.14	0.90	0.22
Visual Inspection	4.00	0.87	0.22
Steel Methods			
	Average	Standard Deviation	Coefficient of Variation
Acoustic Emission	4.00	-	-
Liquid Penetrant Testing	3.67	1.21	0.33
Magnetic Particle Testing - Crack Detection	4.00	0.71	0.18
Magnetic Particle Testing - Weld Inspection	3.75	0.50	0.13
Radiography	4.25	0.50	0.12
Ultrasonic Testing - Crack Detection	3.80	0.45	0.12
Ultrasonic Testing - Pin Inspection	3.83	0.75	0.20
Ultrasonic Testing - Weld Inspection	4.00	0.82	0.20
Visual Inspection	3.67	1.03	0.28

The two methods that fell below this range (rebound hammer and thermal) indicate that respondents felt these methods were less reliable than most other methods. The thermal method can be very dependent on both sun exposure and depth of flaw. Yehia (2007) showed that both of these factors could produce weak readings, causing a decrease in surface area detected or no detection. Similarly, Rens et al. (2005) showed the rebound hammer method did a "poor" job at detecting deterioration while Wood and Rens (2006) showed the method can be highly variable. Note that while it was shown that ultrasonic, mechanical sounding, and radar can also be relatively variable, they were not as unreliable as the rebound hammer method, as supported by the respondents (Wood and Rens 2006). No numbers were provided by the previous studies so no quantitative comparisons could be made to the participant's responses.

Phares et al. (2001) showed that even during routine visual inspections, the inspectors provided values that were statistically different. The inspector's average standard deviation for the superstructure, substructure, and deck were 0.76, 0.75, and 0.83, respectively. Based on these standard deviations and average reference rating, the COV for the superstructure, substructure, and deck responses were 0.14, 0.12, and 0.16, respectively. These responses correlate to about a response of 3 on the reliability scale provided. This shows that while the participants' responses were close they may have been a little too confident in the reliability of visual inspection. Similarly, Barnes and Trottier (2000) showed that radar, chain drag (mechanical sounding), and electrical potential have a COV of 0.258, 0.183, and 0.536, respectively. While these values do not quite match (participant responses indicate a lower COV for each method), the relative reliability of the different methods based on the study and the respondents do agree.

The Gucunski et al. (2013) study on various NDE methods used on concrete bridge decks showed that all NDE methods tested and conducive to data analysis had an average COV of less than 0.25. These results tend to agree when compared to the participant's responses. Furthermore, based on the repeatability grade for each method it was shown that impact echo, ultrasonic, radar, electrical potential, and mechanical sounding all had similar reliability and were relatively more repeatable when compared with infrared. Again, this agrees with the results from the survey. While it was impossible to compare the COV from the study to the COV as determined by the participants, the relative values of each test from the study tend to agree with the relative values based on the survey results.

Using only the limited data available for comparison, it can be shown the survey results indicate reasonable agreement with current studies but probably under-predict in most cases. Based on this reasonable agreement, the COVs indicated by participant responses are shown in Table 4.5. These COVs were determined by fitting a trend line to the response and the COV representation (i.e., a response of 4 meant the test had a COV of about 0.08).

Table 4.5 Concrete and Steel NDE Method Bias and COVs Indicated by Participant Responses

Concrete Methods		
	Bias Factor	COVs Indicated by Participant Responses
Cover Meters/Pachometer	1.000	0.078
Electrical Potential	-	0.093
Impact Echo	1.060	0.097
Mechanical Sounding	-	0.078
Radar	1.043	0.086
Rebound Hammer	-	0.147
Thermal	-	0.153
Ultrasonic Testing	1.029	0.078
Visual Inspection	1.011	0.080
Steel Methods		
	Bias Factor	COVs Indicated by Participant Responses
Acoustic Emission	1.050	0.080
Liquid Penetrant Testing	-	0.088
Magnetic Particle Testing - Crack Detection	-	0.080
Magnetic Particle Testing - Weld Inspection	-	0.085
Radiography	1.029	0.076
Ultrasonic Testing - Crack Detection	1.013	0.084
Ultrasonic Testing - Pin Inspection	1.011	0.083
Ultrasonic Testing - Weld Inspection	1.014	0.080
Visual Inspection	1.033	0.088

4.6 Results for the Accuracy of NDE Methods

As previously mentioned, participants were given three options for each method in the accuracy subsection: false positive, false negative, and true response. Participants were asked to estimate the percentage of times each test would have each result. They were told their percentages for each test should add to 100%. Since the data provided by respondents were open-ended and the sum of the averages could result in a total percentage of more than 100%, each participant's responses were normalized based on the group average to 100%. The normalized responses were then used to calculate the normalized average, normalized standard deviation, and normalized COV. Based on the convergence and stability data from round three, the questions were asked again for round four.

After round four, all responses except two (false negative response for electrical potential and thermal imaging) had a COV of less than 0.5, indicating convergence. For the two responses that were above this threshold, both COVs had dropped significantly and were now in the lower portion of the less than satisfactory range (between 0.5 and 0.8 COV). A breakdown of the COVs from round three and four can be seen in Table 4.6. Note that the highlighted cells indicate a COV of greater than 0.5. Furthermore, the no response for acoustic emission indicates only one person answered this question resulting in no COV.

Table 4.6 Coefficient of Variation Results from the Accuracy Subsection from Round 3 and Round 4

	Survey Three COV			Survey Four COV		
	False Positive	False Negative	True Response	False Positive	False Negative	True Response
Concrete Methods						
Electrical Potential	0.34	0.71	0.15	0.27	0.67	0.12
Mechanical Sounding	0.19	0.31	0.15	0.18	0.30	0.12
Thermal	0.35	0.81	0.38	0.43	0.57	0.29
Visual Inspection	0.41	0.21	0.14	0.38	0.20	0.11
Steel Methods						
Acoustic Emission	-	-	0.25	-	-	-
Liquid Penetrant	0.57	0.32	0.10	0.45	0.35	0.08
Magnetic Particle – Crack	0.44	0.44	0.12	0.33	0.31	0.11
Magnetic Particle – Weld	0.57	0.61	0.13	0.45	0.50	0.11
Radiography	0.37	0.56	0.11	0.39	0.28	0.07
Ultrasonic – Crack	0.31	0.31	0.10	0.31	0.31	0.10
Ultrasonic – Pin	0.22	0.22	0.11	0.20	0.20	0.10
Ultrasonic - Weld	0.36	0.25	0.10	0.36	0.25	0.10
Visual Inspection	0.41	0.46	0.24	0.37	0.38	0.23

After determining the responses had a relatively good convergence, the stability was then analyzed. Based on the stability results from round three to round four, it can be shown that the responses are becoming much more stable when compared with the results from round two to round three. The stability results from round three to round four can be seen in Table 4.7 (results from round two to round three can be seen in Appendix B). Note that the highlighted cells indicate a relatively unstable response.

Table 4.7 Stability Results from the Accuracy Subsection from Round 3 and Round 4

	False Positive	False Negative	True Response
Concrete NDE Method			
Electrical Potential	0.96	0.88	0.04
Mechanical Sounding	1.00	0.97	-1.69
Thermal	0.86	0.37	-0.15
Visual Inspection	1.00	0.97	-1.78
Steel NDE Method			
Acoustic Emission	1.00	1.00	-0.41
Liquid Penetrant	-0.92	0.82	-2.76
Magnetic Particle – Crack	0.78	0.53	0.11
Magnetic Particle – Weld	0.78	0.32	-0.03
Radiography	1.00	0.36	-0.42
Ultrasonic – Crack	1.00	1.00	1.00
Ultrasonic – Pin	0.80	0.80	-0.18
Ultrasonic - Weld	1.00	1.00	1.00
Visual Inspection	0.81	0.67	-5.39

While the responses were becoming more stable, there were still some issues. These issues arose in part because of the large scale used to identify accuracy. If a respondent changed his or her answer by a seemingly small 5%, this change is drastically increased due to the exponential nature of the stability equation. Furthermore, as the rounds progressed it was observed that the participants were becoming more reluctant to change their answers during the iteration process. Also, it stands to reason that if the false responses were becoming stable, the true response should trend towards stability as well. Since the responses were considered to be converging, it was determined that there would be little change if another round were implemented, and the responses would be considered stable if they were asked in a subsequent round. Thus the results from the fourth questionnaire were considered to be the final results. The normalized response average and standard deviation can be seen in Table 4.8.

Table 4.8 Concrete and Steel NDE Method Normalized Accuracy and Standard Deviation Indicated by Participant Responses

	Normalized Average			Normalized Standard Deviation		
	False Positive	False Negative	True Response	False Positive	False Negative	True Response
Concrete Methods						
Electrical Potential	8.40%	11.60%	80.00%	2.30	7.77	9.35
Mechanical Sounding	10.41%	12.22%	77.37%	1.84	3.72	9.06
Thermal	6.67%	33.33%	60.00%	2.89	18.93	17.32
Visual Inspection	9.76%	11.95%	78.29%	3.69	2.43	8.68
Steel Methods						
Acoustic Emission	-	-	100.00%	-	-	-
Liquid Penetrant	8.40%	8.16%	83.44%	3.81	2.86	6.94
Magnetic Particle – Crack	10.92%	12.26%	76.81%	3.61	3.84	8.30
Magnetic Particle – Weld	10.35%	10.51%	79.14%	4.70	5.23	9.09
Radiography	6.43%	9.07%	84.50%	2.49	2.52	6.32
Ultrasonic – Crack	8.00%	8.00%	84.00%	2.48	2.48	8.51
Ultrasonic – Pin	8.92%	8.92%	82.16%	1.81	1.81	8.47
Ultrasonic - Weld	6.89%	8.23%	84.88%	2.48	2.10	8.09
Visual Inspection	14.54%	14.83%	70.63%	5.37	5.62	16.26

There were very few comparative studies that provided information about the accuracy of bridge NDE methods. However, studies that did provide information tended to agree with the results. Gucunski et al. (2013) gave relative accuracy ratings for various concrete methods. It was shown that impact echo, ultrasonic, and electrical potential tended to have more accurate measurements (near the favorable rating) with ground penetrating radar, infrared, and chain drag being slightly less accurate (between the not favorable and favorable rating). The relative scales of these ratings tend to agree with the responses from the participants. Furthermore, Clark et al. (2003) compared thermal imaging and mechanical sounding and found that these methods agree about 62% of the time. This indicates some issues in accuracy for both of these methods. Based on the 77% true response for mechanical sounding and the 60% true response for thermal imaging as determined by the participants, these numbers seem reasonable. In addition, the participants could have been taking into account environmental factors. These factors were shown by various studies including Clark et al. (2003) and Yehia et al. (2007) to affect the accuracy of thermal imaging. This could have been the reason why thermal imaging was determined to be the least accurate of all the methods.

4.7 Results for the Costs Subsection of the Survey

Since the cost questions were changed from round two to round three, the convergence and stability values could not be computed until after round four. After round four, it was determined that 93% of the questions had a COV less than 0.5. The remaining questions were on the lower portion of the less than satisfactory range (between 0.5 and 0.8 COV). All the COVs that were in this range had also dropped significantly from round three. Furthermore, it was determined that all responses had become stable. Based on these factors, the responses from questionnaire four were considered to be the final values. The median of these final values, standard deviation, COV, and the range of costs corresponding to the median value for each cost can be seen in Table 4.9 through Table 4.13. Note that the highlighted cells indicate a COV of greater than 0.5. The stability calculations for this subsection can be seen in Appendix B.

The response range represents the range of costs based on the participant's median response and the cost scales provided during the third and fourth rounds of the survey. It was impossible to know what value within the provided range each respondent wanted to choose so the mean value of these responses could not be used to determine the exact cost using a trend line and interpolation (similar to the bias and repeatability subsections). Based on this, the median response was used in order to have mostly whole numbers to correlate to these cost scales. The median value corresponded to the range most participants indicated. This range was then considered the final value. In the case of a median value being between two whole numbers (1.5, 2.5, 3.5 etc.), the range for both whole numbers the median values were between was used. An example of this would be for a median value of 3.5 for the cost Time Spent to Analyze Data. A response of 3 would have a correlated range of 4 – 8 hours while a response of 4 would have a correlated range of 8 – 12 hours. Thus a response of 3.5 has a correlated range of 4 – 12 hours.

Table 4.9 Concrete and Steel NDE Method Time to Run a Test Median, Standard Deviation, COV, and Correlated Range Indicated by Participant Responses

Concrete Methods				
	Median	Standard Deviation	COV	Response Range
Cover Meters/Pachometer	3	1.04	0.38	8 - 10 Hours
Electrical Potential	4	1.03	0.31	10 - 15 Hours
Impact Echo	4	1.22	0.31	10 - 15 Hours
Mechanical Sounding	2	0.88	0.36	4 - 8 Hours
Radar	3	1.07	0.31	8 - 10 Hours
Rebound Hammer	2	0.76	0.33	4 - 8 Hours
Thermal	4	1.41	0.47	10 - 15 Hours
Ultrasonic Testing	3	0.98	0.28	8 - 10 Hours
Visual Inspection	2	0.60	0.32	4 - 8 Hours
Steel Methods				
	Median	Standard Deviation	COV	Response Range
Acoustic Emission	3	1.41	0.47	8 - 10 Hours
Liquid Penetrant Testing	3	1.27	0.41	8 - 10 Hours
Magnetic Particle Testing - Crack Detection	3	1.04	0.38	8 - 10 Hours
Magnetic Particle Testing - Weld Inspection	3	0.58	0.19	8 - 10 Hours
Radiography	4	0.71	0.17	10 - 15 Hours
Ultrasonic Testing - Crack Detection	3	0.76	0.25	8 - 10 Hours
Ultrasonic Testing - Pin Inspection	3	0.67	0.24	8 - 10 Hours
Ultrasonic Testing - Weld Inspection	3	0.82	0.27	8 - 10 Hours
Visual Inspection	2	0.71	0.35	4 - 8 Hours

Table 4.10 Concrete and Steel NDE Method Time to Analyze Data Median, Standard Deviation, COV, and Correlated Range Indicated by Participant Responses

Concrete Methods				
	Median	Standard Deviation	COV	Response Range
Cover Meters/Pachometer	3	0.92	0.39	4 - 8 Hours
Electrical Potential	3	0.95	0.29	4 - 8 Hours
Impact Echo	4	0.71	0.18	8 - 12 Hours
Mechanical Sounding	2	0.83	0.47	2 - 4 Hours
Radar	4	0.49	0.11	8 - 12 Hours
Rebound Hammer	2	0.90	0.42	2 - 4 Hours
Thermal	4	0.71	0.18	8 - 12 Hours
Ultrasonic Testing	3	0.89	0.26	4 - 8 Hours
Visual Inspection	2	0.78	0.41	2 - 4 Hours
Steel Methods				
	Median	Standard Deviation	COV	Response Range
Acoustic Emission	3.5	0.71	0.20	4 - 12 Hours
Liquid Penetrant Testing	2	0.46	0.26	2 - 4 Hours
Magnetic Particle Testing - Crack Detection	2	0.69	0.32	2 - 4 Hours
Magnetic Particle Testing - Weld Inspection	2	0.52	0.22	2 - 4 Hours
Radiography	3	0.75	0.24	4 - 8 Hours
Ultrasonic Testing - Crack Detection	2	0.53	0.22	2 - 4 Hours
Ultrasonic Testing - Pin Inspection	2.5	0.53	0.21	2 - 8 Hours
Ultrasonic Testing - Weld Inspection	2	0.52	0.22	2 - 4 Hours
Visual Inspection	2	0.52	0.32	2 - 4 Hours

Gucunski et al. (2013) provided a comparison of speeds for each NDE method. For the speed category, the Time to Run a Test and the Time to Analyze Data parameters from the survey were used to compare with the speed results from the Gucunski et al. study. It was shown by Gucunski et al. that radar, electrical potential, infrared imaging, and mechanical sounding all tended to be relatively quick (above the favorable rating) with impact echo and ultrasonic being slower with a rating between the not favorable and favorable rating. Based on the comparison of the relative scales of the study and the survey it can be seen the data tend to agree with only two inconsistencies. One inconsistency is in the case of infrared imaging. Gucunski et al. (2013) determined this method was relatively quick to use while the respondents indicated it took a relatively long time to perform. Furthermore, ultrasonic testing was shown by the study to be very time intensive, but the respondents indicated the test was near the midpoint in terms of time used relative to the other methods.

Table 4.11 Concrete and Steel NDE Method Time to Train an Inspector Median, Standard Deviation, COV, and Correlated Range Indicated by Participant Responses

Concrete Methods				
	Median	Standard Deviation	COV	Response Range
Cover Meters/Pachometer	2.5	1.41	0.54	2 - 14 Days
Electrical Potential	3	1.17	0.37	7 - 14 Days
Impact Echo	5	0.45	0.09	21 + Days
Mechanical Sounding	2	1.24	0.48	2 - 7 Days
Radar	5	0.53	0.12	21 + Days
Rebound Hammer	3	1.35	0.47	7 - 14 Days
Thermal	4	0.55	0.12	14 - 21 Days
Ultrasonic Testing	5	0.38	0.08	21 + Days
Visual Inspection	3	1.30	0.40	7 - 14 Days
Steel Methods				
	Median	Standard Deviation	COV	Response Range
Acoustic Emission	4.5	0.71	0.16	14 - 21+ Days
Liquid Penetrant Testing	2	1.30	0.55	2 - 7 Days
Magnetic Particle Testing - Crack Detection	3	1.25	0.46	7 - 14 Days
Magnetic Particle Testing - Weld Inspection	3	1.03	0.31	7 - 14 Days
Radiography	5	0.00	0.00	21 + Days
Ultrasonic Testing - Crack Detection	5	0.95	0.22	21 + Days
Ultrasonic Testing - Pin Inspection	5	0.92	0.21	21 + Days
Ultrasonic Testing - Weld Inspection	5	0.84	0.19	21 + Days
Visual Inspection	2	1.06	0.40	2 - 7 Days

Table 4.12 Concrete and Steel NDE Method Number of Inspectors Needed Median, Standard Deviation, COV, and Correlated Range Indicated by Participant Responses

Concrete Methods				
	Median	Standard Deviation	COV	Response Range
Cover Meters/Pachometer	2	0.64	0.34	2 Inspectors
Electrical Potential	2.5	1.21	0.45	2 - 3 Inspectors
Impact Echo	3	0.84	0.30	3 Inspectors
Mechanical Sounding	2	0.67	0.38	2 Inspectors
Radar	3	0.82	0.27	3 Inspectors
Rebound Hammer	2	0.58	0.29	2 Inspectors
Thermal	3	0.89	0.37	3 Inspectors
Ultrasonic Testing	2	0.53	0.22	2 Inspectors
Visual Inspection	2	0.50	0.25	2 Inspectors
Steel Methods				
	Median	Standard Deviation	COV	Response Range
Acoustic Emission	2	0.00	0.00	2 Inspectors
Liquid Penetrant Testing	2	0.52	0.32	2 Inspectors
Magnetic Particle Testing - Crack Detection	2	0.69	0.37	2 Inspectors
Magnetic Particle Testing - Weld Inspection	2	0.75	0.41	2 Inspectors
Radiography	3	0.75	0.27	3 Inspectors
Ultrasonic Testing - Crack Detection	2	0.58	0.29	2 Inspectors
Ultrasonic Testing - Pin Inspection	2	0.64	0.30	2 Inspectors
Ultrasonic Testing - Weld Inspection	2	0.63	0.32	2 Inspectors
Visual Inspection	1.5	0.74	0.46	1 - 2 Inspectors

Gucunski et al. (2013) also gave data for the ease of use of NDE methods. While this category did not exactly correlate to the costs used for the surveys, some comparisons can be made. For this category, the Number of Inspectors Needed and the Time to Train Inspectors parameters were used to compare. Along with this, the Time to Run a Test and Time to Analyze Data parameters were also considered but to a lesser degree because they did not directly correlate to the ease of use measures as determined by Gucunski et al. It was shown that electrical potential, infrared imaging, and mechanical sounding all tended to be relatively easy to use (above the favorable rating) with impact echo, radar, and ultrasonic being less easy with a rating between the not favorable and favorable rating. Comparing this to the survey results it can be seen that the relative scales of the study and the survey tends to agree. One inconsistency is in the case of infrared imaging. Gucunski et al. determined this method was relatively easy to use while the respondents indicated it was a relatively difficult test to perform.

Table 4.13 Concrete and Steel NDE Method Monetary Cost for Equipment Median, Standard Deviation, COV, and Correlated Range Indicated by Participant Responses

Concrete Methods				
	Median	Standard Deviation	COV	Response Range
Cover Meters/Pachometer	3	0.71	0.26	$1500 - $3000
Electrical Potential	3	1.41	0.47	$1500 - $3000
Impact Echo	5	0.55	0.12	$6000+
Mechanical Sounding	1	0.88	0.61	0 - $500
Radar	5	0.00	0.00	$6000+
Rebound Hammer	2.5	1.37	0.51	$500 - $3000
Thermal	5	0.45	0.09	$6000+
Ultrasonic Testing	5	0.52	0.11	$6000+
Visual Inspection	1	0.71	0.53	0 - $500
Steel Methods				
	Median	Standard Deviation	COV	Response Range
Acoustic Emission	5	0.00	0.00	$6000+
Liquid Penetrant Testing	1.5	0.53	0.36	0 - $1500
Magnetic Particle Testing - Crack Detection	2	0.98	0.38	$500 - $1500
Magnetic Particle Testing - Weld Inspection	2	1.03	0.39	$500 - $1500
Radiography	5	0.41	0.08	$6000+
Ultrasonic Testing - Crack Detection	5	0.79	0.18	$6000+
Ultrasonic Testing - Pin Inspection	4.5	0.74	0.17	$3000 - $6000+
Ultrasonic Testing - Weld Inspection	5	0.84	0.19	$6000+
Visual Inspection	1	1.16	0.67	0 - $500

Rens et al. (2005) showed that acoustic emission, radar, radiography, and thermography had relatively high costs, while electrical methods, impact echo, magnetic methods, mechanical sounding, surface hardness methods, and ultrasonic had relatively low costs. While it was not discussed what these relative cost were measuring, it was assumed these costs were monetary in nature. Based on this, the response results from the participants tend to agree with this study in terms of monetary cost. It should be noted that impact echo and ultrasonic testing were determined by Rens et al. to be relatively low in cost, but the survey participants indicated these tests have high monetary costs. This discrepancy could be due to the costs determined by Rens et al. being more than just monetary costs.

The final parameter discussed by Gucunski et al. (2013) was monetary cost. This measure could not be compared with the data collected by the survey. For the cost measure, participants were asked to provide a cost estimate for bridge decks with an area of 5,000 ft^2 and 10,000 ft^2. The only monetary cost that was collected by the survey was the cost of equipment, which was not included in the cost estimates performed by the study.

4.8 Conclusion for the Results of the Surveys

After all the data were collected, an initial comparison of the NDE methods was made. This comparison can be seen in Table 4.14. Each method was given a rank in each of the four categories measured: bias, repeatability, accuracy, and cost. Note there are a total of nine tests for both concrete and steel methods. These ranks were based on the most desirable outcome for each category. For bias, the lower the rank, the less biased the test was. Similarly, the lower the rank for the cost category, the cheaper the average cost of the method. Since the cost subsection measured five different costs, an average for all the costs for each method was used for this comparison. For repeatability and accuracy, the lower the rank, the more repeatable or more accurate the test was, respectively. Based on the results from these categories, an overall ranking was established. This ranking was determined by the sum of the rankings from each category. In the case of a test getting a ranking for both bias and accuracy, the average of these two was taken. It should be noted that this overall ranking was made with the assumption that each of the parameters can be weighted equally. For some circumstances this might not be the case. In some instances, one parameter may be more important than the others.

Table 4.14 Comprehensive Comparison by Ranking of Each NDE Method Indicated by Participant Responses

Concrete Methods					
	Bias	Repeatability	Accuracy	Average Cost	Overall Rank
Cover Meters/Pachometer	1	1	-	4	2
Electrical Potential	-	6	1	5	6
Impact Echo	2	7	-	8	7
Mechanical Sounding	-	1	3	1	1
Radar	5	5	-	7	8
Rebound Hammer	-	8	-	3	4
Thermal	-	9	4	7	9
Ultrasonic Testing	4	1	-	6	4
Visual Inspection	3	4	2	2	3
Steel Methods					
	Bias	Repeatability	Accuracy	Average Cost	Overall Rank
Acoustic Emission	1	2	1	8	2
Liquid Penetrant Testing	-	8	5	2	7
Magnetic Particle Testing - Crack	-	2	8	3	3
Magnetic Particle Testing - Weld	-	7	7	3	9
Radiography	5	1	3	9	4
Ultrasonic Testing - Crack Detection	3	6	4	5	6
Ultrasonic Testing - Pin Inspection	2	5	6	5	4
Ultrasonic Testing - Weld Inspection	4	2	2	5	1
Visual Inspection	6	8	9	1	8

By comparing the rankings of each of the four categories that were examined for each NDE method, it is possible to understand the relative differences between each test. Furthermore, a correlation of the costs of a method to the bias, accuracy and reliability can be made. In general, it was shown by the participants that the more expensive the method was, the better bias, accuracy, and reliability the method had and vice versa. However, there were a few exceptions to this rule. Both infrared imaging and radar tended to be relatively expensive. Infrared imaging tended to be relatively inaccurate and not very repeatable, and radar was relatively biased and also fairly unrepeatable. By evaluating these comparisons it can be seen that inspection planning choices should consider the quality of information a test provides as well as the costs (in terms of time and money).

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary

An extensive literature search was performed, and many NDE articles, papers and reports for bridge engineering applications were found and reviewed. Based on the articles found, it was determined that there is a significant amount of research being conducted to determine the best situations for specific NDE methods for bridges. Furthermore, it was determined there have been some studies conducted to determine various costs and statistical measures, including accuracy and reliability of many of these methods relative to one another. It was shown, however, that these comparative studies did not provide enough information to adequately quantify the accuracy and reliability of common bridge NDE methods.

Since it was determined that there is limited work being done to quantify the accuracy and reliability of common NDE methods or to compare various tests to one another, a comprehensive survey to gather expert opinion was identified as a means to gather the desired information. It was established that the Delphi method was an efficient and effective survey technique to gather this information. This survey aimed to provide quantitative descriptions of accuracy, reliability, bias, and costs to provide information to researchers and practitioners working in the fields of bridge management and inspection.

A total of four Delphi method rounds were conducted in order to determine quantitatively the accuracy, reliability, bias, and various costs of common NDE methods. The first survey was employed to determine background information of the participants and common NDE methods for bridges. The second and subsequent surveys were used along with various statistical scales to attempt to quantify the uncertainty present in results from NDE methods. The power of the survey was limited by the sample size, and further Delphi studies focusing on one particular NDE method in depth would enhance the quality of the findings presented here.

5.2 Conclusions

The results of these surveys and literature review were used to the extent possible to develop quantitative information for each method. Based on these results, the following conclusions can be drawn:

1. Most commonly used bridge NDE methods tend to be under-biased, meaning the majority of the measured results are slightly less than the true value. However, all these biases in the survey were shown to be less than 10%. These values tended to agree with the data from previous experimental studies.
2. Most commonly used bridge NDE methods tend to be relatively repeatable. Furthermore, it was shown that inspectors seem to have a relative understanding of the variability in different tests, but they tend to not have an understanding of the absolute scale of the variability. Based on the data provided by Barnes and Trottier (2000), it was shown participants were able to indicate which methods were more repeatable, but it also showed they felt all methods were more repeatable than what was determined by the previous studies. Barnes and Trottier showed that radar, chain drag (mechanical sounding), and electrical potential have a COV of 0.258, 0.183, and 0.536, respectively, while the participants responses indicated they thought the COV for radar, mechanical sounding, and electrical potential were 0.086, 0.078, and 0.093, respectively.
3. The accuracy of commonly used bridge NDE methods tends to be relatively variable. For concrete testing, most tests had a true response percentage of about 80%. The exception to this was infrared imaging with a true response percentage of 60%. Furthermore, most steel tests had a true response percentage of about 85%. However, there were a couple exceptions to this. Acoustic emission had a true response percentage of 100%, while visual inspection

and magnetic particle had a true response percentage of 70% and an average of 78%, respectively.

4. The various costs associated with the NDE methods examined tended to be quite variable, making this measure difficult to evaluate. However, there was a small trend that indicated tests that were cheaper in terms of equipment also tended to be easier and faster to perform.
5. By comparing the rankings of each of the four categories that were examined for each NDE method, it is possible to correlate the cost of a method to the bias, accuracy, and reliability. In general, it was shown by the participants that the more expensive the method was, the better bias, accuracy, and reliability the method had and vice versa. A risk-based approach to inspection planning would therefore need to carefully consider the level of information needed and the costs of obtaining that information.

5.3 Suggestions for Future Work

There are several topics that could be further investigated involving the accuracy and reliability of commonly used bridge NDE methods. The following is a description of these topics:

1. A follow-up survey could be conducted with a larger sample size. This follow-up survey could be used to see if the new, larger group's values are similar to the values found in this survey. This survey might also be enhanced by focusing on just one NDE method at a time and finding experts in that particular method as respondents.
2. Rather than conducting a survey, a more comprehensive examination of many commonly used NDE methods could be performed by conducting experiments with each NDE method in the field and comparing the results to one another. This comprehensive examination would need many more resources, including time and equipment in order to perform the tests.
3. More comparative studies need to be performed that actually measure the variation in each test. This information then needs to be distributed to inspectors and managers so they know what kind of variability each test tends to produce.
4. This study looked at NDE methods independently. Further research needs to be conducted to compare methods that can be used with one another. An example of this could be using a cheaper and easier method to detect broad areas of deterioration, and then use a more expensive method to examine that area. Another example could be examining the use of methods in conjunction with one another, which could increase the overall accuracy and reliability.
5. The information that was gathered from this survey could be used to better understand various parameters of commonly used bridge NDE methods. Future work could apply these findings to develop a risk-based approach to bridge inspection.

BIBLIOGRAPHY

ASCE/SEI-AASHTO Ad-hoc Group on Bridge Inspection, Rating, Rehabilitation, and Replacement. (2009). White Paper on Bridge Inspection and Rating. *Journal of Bridge Engineering, 14(1),* 1-5.

Algernon, D., Hiltunen, D. R., Ferraro, C. C., Ishee, C. (2011). "Rebar Detection with Cover Meter and Ultrasonic Pulse Echo Combined with Automated Scanning System." *Transportation Research Record, 2251,* 123-131.

ASTM International. (2009). Standard Test Method for Corrosion Potential of Uncoated Reinforcing Steel in Concrete, C876 - 09. West Conshohocken, PA.

Barnes, C. L. & Trottier, J. (2000). "Ground-Penetrating Radar for Network-Level Concrete Deck Repair Management." *Journal of Transportation Engineering, 126,* 257-262.

Barnes, C. L. & Trottier, J. (2004). "Effectiveness of Ground Penetrating Radar in Predicting Deck Repair Quantities." *Journal of Infrastructure Systems, 10,* 69-76.

Clark, M. R., McCann, D. M., Forde, M. C. (2003). "Application of Infrared Thermography to the Non-Destructive Testing Concrete and Masonry Bridges." *NDT&E International, 36,* 265-275.

English, M.J., & Kernan, J.L. (1976). "The Prediction of Air Travel and Aircraft Technology to the Year 2000 Using the Delphi Method." *Transportation Research, 10(1).*

Gad, G. M., & Shane, J. S. (2012). "A Delphi Study on the Effects of Culture on the Choice Dispute Resolution Methods in International Construction Contracts." *Construction Research Congress,* 1-10.

Golaski, L., Gebski, P., & Ono, K. (2002) "Diagnostics of Reinforce Concrete Bridges by Acoustic Emission." *Journal of Acoustic Emission, 20,* 83-98

Gong, Z., Nyborg, E. O., Oommen, G. (1992). "Acoustic Emission Monitoring of Steel Railroad Bridges." *Materials Evaluation, 50(7),* 883-887.

Green, P. J. (1981). "The content of a college-level leadership course for land-based outdoor pursuits in the pacific northwest: A Delphi consensus." Doctoral dissertation, University of Oregon

Green, P. J. (1982). *The content of a college-level outdoor leadership course*. Paper presented at the Conference of the Northwest District Association for the American Alliance for Health, Physical Education, Recreation, and Dance. Spokane, WA.

Gucunski, G., Romero, F., Kruschwitz, S., Feldmann, R., Abu-Hawash, A., & Dunn, M. (2010). "Multiple Complementary Nondestructive Evaluation Technologies for Condition Assessment of Concrete Bridge Decks." *Transportation Research Record, 2201,* 34-44. Retrieved February 16, 2013 from: http://www.trb.org/Main/Blurbs/167278.aspx

Gucunski, N., Imani, A., Romero, F., Nazarian, S., Yuan, D., Wiggenhauser, H., Shokouhi, P., Taffe, A., & Kutrubes, D. (2013). "Nondestructive Testing to Identify Concrete Bridge Deck Deterioration." Prepared by the Transportation Research Board. SHRP 2 Report S2-R06A-RR-1.

Hsu, C., & Sanford, B. A. (2007). “The Delphi Technique: Making Sense of Consensus.” *Practical Assessment, Research & Evaluation, 12(10).*

Kalaian, S.A., & Kasim, R.M. (2012). “Termination Sequential Delphi Survey Data Collection.” *Practical Assessment, Research & Evaluation, 17(5).*

Linstone, H. A., & Turoff, M. (2002). *The Delphi Method: Techniques and Applications.* Reading, MA: Addison-Wesley Publishing Co., Advanced Book Program.

Ludwig, B. (1997). Predicting the future: Have you considered using the Delphi methodology? *Extension Journal, 35* (5). Retrieved January 14, 2013 from http://www.joe.org/joe/1997october/tt2.php

Maierhofer, C., Reinhardt, H., & Dobmann, G. (2010). *Non-Destructive Evaluation of Reinforced Concrete Structures, Volume 2 - Non-Destructive Testing Methods*. Woodhead Publishing. Retrieved January 16, 2013 from http://www.knovel.com/web/portal/browse/display?_EXT_KNOVEL_DISPLAY_bookid=4368&VerticalID=0

Martin, J., Hardy, M. S. A., Usmani, A. S., & Forde, M. C. (1998). “Accuracy of NDE in Bridge Assessment.” *Engineering Strucutres, 20(11),* 979-984.

McCrea, A., Chamberlain, D., Navon, R. (2002). “Automated Inspection and Restoration of Steel Bridges – A Critical Review of Methods and Enabling Technologies.” *Automation in Construction, 11,* 351-373.

Minchin, R. E., Zayed, T., Boyd, A. J., & Mendoza, M. (2006) “Best Practices of Bridge System Management – A Synthesis.” *Journal of Management in Engineering, 22,* 186-195.

Mix, P. L. (2005). *Introduction to Nondestructive Testing: A Training Guide, Second Edition.* Hoboken, NJ: John Wily & Sons, Inc.

Moore, M., Rolander, D., Graybeal, B., Phares, B., Washer, G. (2001). Highway Bridge Inspection: State-of-the-Practice Study. Prepared by the United States Department of Transportation Federal Highway Administration. Final Report.

Nair, A. (2010). “Acoustic Emission Monitoring of Bridges: Review and Case Studies.” *Engineering Structures, 32,* 1704-1714.

Okoli, C. & Pawlowski, S. D. (2004). “The Delphi method as a research tool: an example, design considerations and applications.” *Information & Management, 42*, 15-29.

Phares, B. M., Rolander, D. D., Graybeal, B. A., & Washer, G. A. (2001). Reliability of Visual Bridge Inspection. *Public Roads (Publication of the Federal Highway Administration), 64(5).* Retrieved January 24, 2013 from http://www.fhwa.dot.gov/publications/publicroads/01marapr/bridge.cfm

Phares, B. M., Washer, G. A., Rolander, D. D., Graybeal, B. A., & Moore, M. (2004). “Routine Highway Bridge Inspection Condition Documentation Accuracy and Reliability.” *Journal of Bridge Engineering, 9*, 403-413.

Qasrawi, H. Y. (2000). “Concrete Strength by Combined Nondestructive Methods Simply and Reliably Predicted.” *Cement and Concrete Research, 30,* 739-746

Ramey, G. E., Wolff, A. R., & Wright, R. L. (1997). "DOT Management Actions to Enhance Bridge Durability/Longevity." *Practice Periodical on Structural Design and Construction, 2*, 125-130.

Rens, K. L., Kim, T. (2006). Quebec Bridge Inspection Using Common Nondestructive and Destructive Testing Techniques. Conference Proceeding Paper part of Structures Congress 2006: Structural Engineering and Public Safety, 1-16

Rens, K. L., Nogueira, C. L., & Transue, D. J. (2005). "Bridge Management and Nondestructive Evaluation." *Journal of Performance of Constructed Facilities, 19(1),* 3-16

Rens, K. L., & Transue, D. J. (1998). "Recent Trends in Nondestructive Inspections in State Highway Agencies." *Journal of Performance Constructed Facilities, 12*. 94-96.

Rens, K. L., Wipf, T. J., Klaiber, F. W. (1997). "Review of Nondestructive Techniques of Civil Infrastructure." *Journal of Performance Constructed Facilities, 11*. 152-160.

Sack, D. A. & Olson, L. D. (1995). "Advanced NDT Methods for Evaluating Concrete Bridges and Other Structures." *NDT&E International, 28(6),* 349-357.

Saito, M., & Sinha, K. C. (1991). "Delphi Study on Bridge Condition Rating and Effects of Improvements." *Journal of Transportation Engineering, 117,* 320-334.

Scott, M., Rezaizadeh, A., Delahaza, A., Santos, C.G., Moore, M., Graybeal, B., Washer, G. (2003). "A Comparison of Nondestructive Evaluation Methods for Bridge Deck Assessment." *NDT&E International, 36,* 245-255.

Sheshkin, D. J. (2004). Handbook of Parametric and Nonparametric Statistical Procedures, Third Edition. Boca Raton, FL: Chapman & Hall/CRC.

Shull, P. J. (2002). *Nondestructive Evaluation: Theory, Techniques, and Applications.* New York, NY: Marcel Dekker, Inc.

Song, H., Saraswathy, V. (2007). "Corrosion Monitoring of Reinforced Concrete Structures – A Review." *International Journal of Electrochemical Science, 2,* 1-28

Watanabe, T., Morita, T., Hashimoto, C., & Ohtsu, M. (2004). "Detecting Voids in Reinforced Concrete Slab by SIBIE." *Construction and Building Materials, 18,* 225-231

Washer, G. A. (1998). "Developments for the Non-Destructive Evaluation of Highway Bridges in the USA." *NDT&E International, 31(4),* 245-249.

Witkin, B. R., & Altschuld, J. W. (1995). *Planning and conducting needs assessment: A practical guide.* Thousand Oaks, CA: Sage Publications, Inc.

Wood, J. C., & Rens, K. L. (2006). "Nondestructive Testing of the Lawrence Street Bridge." *Structures.*

Yasamis-Speroni, F., Lee, D., & Arditi, D. (2012). "Evaluating the Quality Performance of Pavement Contractors." *Journal of Construction Engineering Management, 138,* 1114-1124.

Yehia, S., Abudayyeh, O., Nabulsi, S., & Abdelqader, I. (2005). “A Comparison between Ground Penetrating Radar (GPR) and Impact Echo (IE) for Detection of Common Concrete Bridge Decks.” *Computing in Civil Engineering*

Yehia, S., Abudayyeh, O., Nabulsi, S., & Abdelqader, I. (2007). “Detection of Common Defect in Concrete Bridge Decks Using Nondestructive Evaluation Techniques.” *Journal of Bridge Engineering, 12*, 215-225.

Yousuf, M. I. (2007). “Using Experts’ Opinions through Delphi Technique.” *Practical Assessment, Research & Evaluation, 12(4).*

APPENDIX A: EXAMPLES OF THE SURVEYS IMPLEMENTED

Colorado State University

Using Expert Opinion to Quantify Accuracy and Reliability of Nondestructive Evaluation on Bridges - Round One of a Delphi Survey

This questionnaire is the first round of an iterative voluntary survey with approximately four to five rounds. All results will be kept confidential and will be published in a thesis paper for Colorado State University. The purpose of this survey is to help determine the accuracy, precision, sensitivity, reliability and various costs associated with commonly used nondestructive evaluation (NDE) techniques on bridges. The goal of the researchers is to quantify and better understand these properties of the various NDE techniques for use in inspection planning. Please have the person at your firm with the most experience with NDE on bridges fill out this questionnaire. Each questionnaire will take approximately 20 minutes to complete. Please be thoughtful, thorough, and candid in your responses.

If you have any questions regarding this survey please contact Alex Hesse at Colorado State University by phone at (608) 354-5865 or email at aahesse@engr.colostate.edu. Please return the completed questionnaire two weeks after receiving this questionnaire either by email at aahesse@engr.colostate.edu or mailing to:

Department of Civil and Environmental Engineering
Colorado State University
1372 Campus Delivery
Fort Collins, CO 80523-1372

ATTN: Alex Hesse

Completed by: ____________________
Organization: ____________________
Position/Title: ____________________
Address: ____________________
City/State/Zip: ____________________
Phone Number: ____________________
Email Address: ____________________

Please answer all questions thoroughly. If more space is needed please write in the space provided at the end of this questionnaire or attach further responses on a separate sheet, and indicate which question you are responding to. PLEASE DO NOT USE ABBREVIATIONS OR ACRONYMS IN YOUR RESPONSES.

The following section contains a series of questions concerning your experience with NDE techniques on bridges.

Section 1 - NDE Experience with Bridges

1) What is your current level of education? *(mark only one)*

- ☐ High School/GED
- ☐ College Education - No Degree
- ☐ 2-year College Degree (Associates)
- ☐ 4-year college Degree (BA, BS)
- ☐ Master's Degree
- ☐ Doctoral Degree
- ☐ Other, Please Specify: ______________________

2) Mark any certifications for bridge inspection which you currently hold. For ASNT certification, please indicate specifically which NDE test certification you hold for each level. (*Mark all that apply. Note that ASNT refers to American Society for Nondestructive Testing and NICET refers to National Institute for Certification in Engineering Technologies (NICET) Bridge Safety Inspection)*

- ☐ P.E. License
- ☐ ASNT Level I ______________________
- ☐ ASNT Level II ______________________
- ☐ ASNT Level III ______________________
- ☐ NICET Level I
- ☐ NICET Level II
- ☐ NICET Level III
- ☐ NICET Level IV
- ☐ Other, Please Specify: ______________________

3) How many years of experience do you have using NDE on bridges?

____

4) In what states does your bridge inspection group practice NDE? Abbreviations are acceptable

____

5) What tasks do you do when it comes to NDE techniques on bridges? (e.g. Management, Data Analysis etc…)

____

6) Approximately how many bridges are inspected by your organization **each year** and how many of these do you help with?

____

The following two sections contain a series of questions concerning your usage of various NDE techniques. Section 2 will be on steel bridges and Section 3 will be on concrete bridges. For each technique, please indicate what condition the technique is used to identify or assess. If you have no experience with any of the following techniques, please skip that technique and go on to the next technique.

Section 2 - Use of NDE Techniques for Steel Bridges

Example:
Ultrasonic
Flaw detection in welds, Crack detection, and Corrosion detection

The following section only pertains to **Steel Bridges:**

1) Acoustic Emission

2) Liquid Penetrant

3) Thermal/Infrared

4) Visual Inspection

5) Eddy Current

6) Magnetic Particle

7) Ultrasonic

8) Radiography

Section 2 is continued on the next page.

9) Vibration Analysis

____

10) Other

____

11) Of these NDE techniques on steel bridges, which methods do you use at least once every month? If you do not use any techniques at least once every month, please list the two techniques you use the most.

____

Section 3 - Use of NDE Techniques for Concrete Bridges

The following section only pertains to **Concrete Bridges**:

1) Acoustic Emission

____

2) Mechanical Sounding (Chain Drag)

____

3) Rebound Hammer

____

4) Impact Echo

____

5) Cover Meters/Pachometers

____

6) Radar

____

7) Thermal/Infrared

____

8) Vibration Analysis

____

9) Electrical Potential Measurements

____

10) Radiography

____

Section 3 is continued on the next page.

11) Ultrasonics (Pulse Velocity)

____

12) Visual Inspection

____

13) Other

____

14) Of these NDE techniques on concrete bridges, which methods do you use at least once every month? If you do not use any techniques at least once every month, please list the two techniques you use the most.

____

15) Have you stopped using any NDE techniques on bridges in the past? If so, which technique and please explain why.

____

Please indicate if you would like to complete subsequent surveys by email or mail. *(mark only one)*

- ☐ Mail
- ☐ Email

End of Questionnaire. Please make sure you have completed all three sections of the questionnaire. After that, please return the completed survey either by mail or email as described above. If you have any comments please include them in the extra space provided.

Thank you for completing this questionnaire and successive questionnaires in this study. Your responses will help improve the use of NDE techniques.

Extra space if needed:

Colorado State University

Using Expert Opinion to Quantify Accuracy and Reliability of Nondestructive Evaluation on Bridges - Round Two of a Delphi Survey

---

This questionnaire is the second round of an iterative voluntary survey with approximately four to five rounds. All results will be kept confidential and will be published in a thesis paper for Colorado State University. Please be thoughtful, thorough, and candid in your responses.

If you have any questions regarding this survey please contact Alex Hesse at Colorado State University by phone at (608) 354-5865 or email at aahesse@engr.colostate.edu. Please return the completed questionnaire by **November 28th, 2012** (Note: Thanksgiving is November 22nd) either by email at aahesse@engr.colostate.edu or mailing to:

Department of Civil and Environmental Engineering
Colorado State University
1372 Campus Delivery
Fort Collins, CO 80523-1372

ATTN: Alex Hesse

Participant Code: A12

Please answer all questions thoroughly. All questions are **OPINION BASED**. You do not need to have extensive experience of a method to answer questions about that method. If you feel you have no opinion on a topic, please mark "Don't Know". Please answer all questions to the best of your knowledge. If more space is needed please write in the space provided at the end of this questionnaire or attach further responses on a separate sheet, and indicate which question you are responding to. PLEASE DO NOT USE ABBREVIATIONS OR ACRONYMS IN YOUR RESPONSES.

Section 1 - NDE Methods for Concrete Bridges

The following section contains a series of ranking questions concerning various aspects of NDE methods on **CONCRETE** bridges. ALL QUESTIONS ARE OPINION BASED. You do not need to have extensive experience of a method to answer questions about that method. If you feel you have no opinion on a topic, please mark "Don't Know". Please answer all questions to the best of your knowledge.

The following questions pertain to **BIAS** of each NDE test for concrete bridges. **BIAS** is defined as the tendency of a test to consistently measure either higher or lower than the actual or perceived value. Rate how **BIASED** you feel each method is based on the scale provided. *(mark one for each method)*

BIAS SCALE

-50% or More	-40%	-30%	-20%	-10%		+10%	+20%	+30%	+40%	+50% or More
1	2	3	4	5	6	7	8	9	10	11

LOWER THAN TRUE VALUE — ↑ TRUE VALUE — HIGHER THAN TRUE VALUE

Example:

If Cover Meters/Pachometers do not consistently measure a higher or lower value than the actual value they have no BIAS and the response is 6 (TRUE VALUE)

If Impact Echo consistently measures values that average about 20% higher than the actual value it has a BIAS of about 20% higher than the true value and the response is 8 (+20%)

	1	2	3	4	5	6	7	8	9	10	11	Don't Know
Cover Meters/Pachometer						X						
Impact Echo								X				

	1	2	3	4	5	6	7	8	9	10	11	Don't Know
1) Cover Meters/Pachometer	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
2) Impact Echo	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
3) Radar	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
4) Ultrasonic Testing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
5) Visual Inspection	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

The following questions pertain to **ACCURACY** of each NDE test for concrete bridges. **ACCURACY** is defined as the tendency of a test to measure false results. Please indicate, on average, the **ACCURACY** of each method for each category. *Note: Your percentage for each method should add to 100%.*

Definitions:

False Positive: Test measures damage, but there is no damage

False Negative: Test measures no damage, but there is damage

True Response: Test measures damage and there is damage OR test measures no damage and there is no damage

Example:

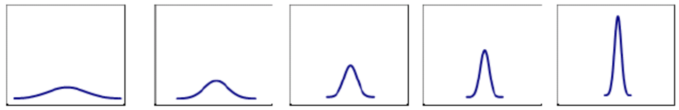
	False Positive	False Negative	True Response	Don't Know
Electrical Potential	0	0	100%	
Mechanical Sounding	5%	10%	85%	

	False Positive	False Negative	True Response	Don't Know
6) Electrical Potential				
7) Mechanical Sounding				
8) Thermal				
9) Visual Inspection				

The following questions pertain to **PRECISION** of each NDE test for concrete bridges. **PRECISION** is defined as the reproducibility of a test in a controlled environment (i.e. a lab setting). Please mark the **PRECISION** of each method based on the scale provided. *(mark one for each method)*

PRECISION SCALE

Response:	1	2	3	4	5



Mean:	100	100	100	100	100
Standard Deviation:	35	22	12	8	6
Coefficient of Variation:	0.35	0.22	0.12	0.08	0.06

Note: The coefficient of variation is a normalized measure of dispersion of a probability distribution.

Example:

If Cover Meters/Pachometers are used to repeat a test multiple times in a controlled environment, and the results have a coefficient of variation of about 0.12 they have an average PRECISION and the response would be 3.

If Electrical Potential is used to repeat a test multiple times in a controlled environment, and the results have a coefficient of variation of about 0.06 it has a high PRECISION and the response would be 5.

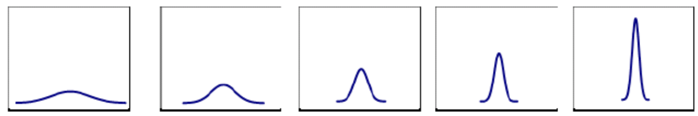
	1	2	3	4	5	Don't Know
Cover Meters/Pachometer			X			
Electrical Potential					X	

	1	2	3	4	5	Don't Know
10) Cover Meters/Pachometer	☐	☐	☐	☐	☐	☐
11) Electrical Potential	☐	☐	☐	☐	☐	☐
12) Impact Echo	☐	☐	☐	☐	☐	☐
13) Mechanical Sounding	☐	☐	☐	☐	☐	☐
14) Radar	☐	☐	☐	☐	☐	☐
15) Rebound Hammer	☐	☐	☐	☐	☐	☐
16) Thermal	☐	☐	☐	☐	☐	☐
17) Ultrasonic Testing	☐	☐	☐	☐	☐	☐
18) Visual Inspection	☐	☐	☐	☐	☐	☐

The following questions pertain to **RELIABILITY** of each NDE test for concrete bridges. For the purposes of this survey **RELIABILITY** is defined as the reproducibility of a test in an uncontrolled environment (i.e. in the field). Please mark the **RELIABILITY** of each method based on the scale provided. *(mark one for each method)*

RELIABILITY SCALE

Response:	1	2	3	4	5



	1	2	3	4	5
Mean:	100	100	100	100	100
Standard Deviation:	35	22	12	8	6
Coefficient of Variation:	0.35	0.22	0.12	0.08	0.06

Note: The coefficient of variation is a normalized measure of dispersion of a probability distribution.

Example:

If Cover Meters/Pachometers are used to repeat a test multiple times in a controlled environment, and the results have a coefficient of variation of about 0.12 they have an average RELIABILITY and the response would be 3.

If Electrical Potential is used to repeat a test multiple times in a controlled environment, and the results have a coefficient of variation of about 0.06 it has a high RELIABILITY and the response would be 5.

	1	2	3	4	5	Don't Know
Cover Meters/Pachometer			X			
Electrical Potential					X	

	1	2	3	4	5	Don't Know
19) Cover Meters/Pachometer	☐	☐	☐	☐	☐	☐
20) Electrical Potential	☐	☐	☐	☐	☐	☐
21) Impact Echo	☐	☐	☐	☐	☐	☐
22) Mechanical Sounding	☐	☐	☐	☐	☐	☐
23) Radar	☐	☐	☐	☐	☐	☐
24) Rebound Hammer	☐	☐	☐	☐	☐	☐
25) Thermal	☐	☐	☐	☐	☐	☐
26) Ultrasonic Testing	☐	☐	☐	☐	☐	☐
27) Visual Inspection	☐	☐	☐	☐	☐	☐

The following questions pertain to various **COSTS** of each NDE test for concrete bridges. For each segment you will be developing your own scale and marking your opinion of **COST** for each method based on this scale. This process will give us a better idea of your opinion on each cost relative to other participants. After making your scale, please mark the **COSTS** of each method. *(mark one for each)*

Example:

Personal Scale Made For **TIME SPENT RUNNING A TEST**

	Very Low	Low	Moderate	High	Very High
Scale	0 - 4 Hours	3 - 8 Hours	8 - 10 Hours	10 - 15 Hours	15+ Hours

If Cover Meters/Pachometer takes about 4 hours to run a test on the example bridge the TIME SPENT RUNNING A TEST (3 - 8 Hours) response would be Low.

If Electrical Potential takes about 13 hours to run a test on the example bridge the TIME SPENT RUNNING A TEST (10 - 15 Hours) response would be High.

	Very Low	Low	Moderate	High	Very High	Don't Know
Cover Meters/Pachometer		X				
Electrical Potential				X		

Please develop your scale and mark the average **TIME SPENT RUNNING A TEST** for each method for a 200 ft, 4 lane Prestressed Concrete Bridge with a reinforced concrete deck. *(mark one for each method)*

Personal Scale Made For **TIME SPENT RUNNING A TEST**

	Very Low	Low	Moderate	High	Very High
Scale					

	Very Low	Low	Moderate	High	Very High	Don't Know
28) Cover Meters/Pachometer	☐	☐	☐	☐	☐	☐
29) Electrical Potential	☐	☐	☐	☐	☐	☐
30) Impact Echo	☐	☐	☐	☐	☐	☐
31) Mechanical Sounding	☐	☐	☐	☐	☐	☐
32) Radar	☐	☐	☐	☐	☐	☐
33) Rebound Hammer	☐	☐	☐	☐	☐	☐
34) Thermal	☐	☐	☐	☐	☐	☐
35) Ultrasonic Testing	☐	☐	☐	☐	☐	☐
36) Visual Inspection	☐	☐	☐	☐	☐	☐

Please develop your scale and mark the average **TIME SPENT ANALYZING DATA** for each method for a 200 ft, 4 lane Prestressed Concrete Bridge with a reinforced concrete deck. *(mark one for each method)*

Personal Scale Made For **TIME SPENT ANALYZING DATA**

	Very Low	Low	Moderate	High	Very High
Scale					

		Very Low	Low	Moderate	High	Very High	Don't Know
37)	Cover Meters/Pachometer	☐	☐	☐	☐	☐	☐
38)	Electrical Potential	☐	☐	☐	☐	☐	☐
39)	Impact Echo	☐	☐	☐	☐	☐	☐
40)	Mechanical Sounding	☐	☐	☐	☐	☐	☐
41)	Radar	☐	☐	☐	☐	☐	☐
42)	Rebound Hammer	☐	☐	☐	☐	☐	☐
43)	Thermal	☐	☐	☐	☐	☐	☐
44)	Ultrasonic Testing	☐	☐	☐	☐	☐	☐
45)	Visual Inspection	☐	☐	☐	☐	☐	☐

Please develop your scale and mark the average **TIME TO TRAIN AN INSPECTOR** to be considered able to competently perform the specified test method for concrete. *(mark one for each method)*

Personal Scale Made For **TIME TO TRAIN AN INSPECTOR**

	Very Low	Low	Moderate	High	Very High
Scale					

		Very Low	Low	Moderate	High	Very High	Don't Know
46)	Cover Meters/Pachometer	☐	☐	☐	☐	☐	☐
47)	Electrical Potential	☐	☐	☐	☐	☐	☐
48)	Impact Echo	☐	☐	☐	☐	☐	☐
49)	Mechanical Sounding	☐	☐	☐	☐	☐	☐
50)	Radar	☐	☐	☐	☐	☐	☐
51)	Rebound Hammer	☐	☐	☐	☐	☐	☐
52)	Thermal	☐	☐	☐	☐	☐	☐
53)	Ultrasonic Testing	☐	☐	☐	☐	☐	☐
54)	Visual Inspection	☐	☐	☐	☐	☐	☐

Please develop your scale and mark the average **MONETARY COST FOR THE EQUIPMENT** for each method for a 200 ft, 4 lane Prestressed Concrete Bridge with a reinforced concrete deck. *(mark one for each method)*

Personal Scale Made For **MONETARY COST FOR THE EQUIPMENT**

	Very Low	Low	Moderate	High	Very High
Scale					

		Very Low	Low	Moderate	High	Very High	Don't Know
55)	Cover Meters/Pachometer	☐	☐	☐	☐	☐	☐
56)	Electrical Potential	☐	☐	☐	☐	☐	☐
57)	Impact Echo	☐	☐	☐	☐	☐	☐
58)	Mechanical Sounding	☐	☐	☐	☐	☐	☐
59)	Radar	☐	☐	☐	☐	☐	☐
60)	Rebound Hammer	☐	☐	☐	☐	☐	☐
61)	Thermal	☐	☐	☐	☐	☐	☐
62)	Ultrasonic Testing	☐	☐	☐	☐	☐	☐
63)	Visual Inspection	☐	☐	☐	☐	☐	☐

Please develop your scale and mark the average **NUMBER OF INSPECTORS NEEDED** for each method for a 200 ft, 4 lane Prestressed Concrete Bridge with a reinforced concrete deck. *(mark one for each method)*

Personal Scale Made For **NUMBER OF INSPECTORS NEEDED**

	Very Low	Low	Moderate	High	Very High
Scale					

		Very Low	Low	Moderate	High	Very High	Don't Know
64)	Cover Meters/Pachometer	☐	☐	☐	☐	☐	☐
65)	Electrical Potential	☐	☐	☐	☐	☐	☐
66)	Impact Echo	☐	☐	☐	☐	☐	☐
67)	Mechanical Sounding	☐	☐	☐	☐	☐	☐
68)	Radar	☐	☐	☐	☐	☐	☐
69)	Rebound Hammer	☐	☐	☐	☐	☐	☐
70)	Thermal	☐	☐	☐	☐	☐	☐
71)	Ultrasonic Testing	☐	☐	☐	☐	☐	☐
72)	Visual Inspection	☐	☐	☐	☐	☐	☐

Section 2 - NDE Methods for Steel Bridges

The following section contains a series of ranking questions concerning various aspects of NDE methods on **STEEL** bridges. ALL QUESTIONS ARE OPINION BASED. You do not need to have extensive experience of a method to answer questions about that method. If you feel you have no opinion on a topic, please mark "Don't Know". Please answer all questions to the best of your knowledge.

The following questions pertain to **BIAS** of each NDE test for steel bridges. **BIAS** is defined as the tendency of a test to measure either higher or lower than the actual or perceived value. Rate how **BIASED** you feel each method is based on the scale provided. *(mark one for each method)*

BIAS SCALE

-50% or More	-40%	-30%	-20%	-10%		+10%	+20%	+30%	+40%	+50% or More
1	2	3	4	5	6	7	8	9	10	11

LOWER THAN TRUE VALUE ↑ TRUE VALUE HIGHER THAN TRUE VALUE

Example:

If Acoustic Emission does not consistently measure a higher or lower value than the actual value it has no BIAS and the response is 6 (TRUE VALUE)

If Radiography consistently measures values that average about 20% higher than the actual value it has a BIAS of about 20% higher than the true value and the response is 8 (+20%)

	1	2	3	4	5	6	7	8	9	10	11	Don't Know
Acoustic Emission						X						
Radiography								X				

	1	2	3	4	5	6	7	8	9	10	11	Don't Know
73) Acoustic Emission	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
74) Radiography	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
75) Ultrasonic Testing - Crack Detection	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
76) Ultrasonic Testing - Pin Inspection	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
77) Ultrasonic Testing - Weld Inspection	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
78) Visual Inspection	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

The following questions pertain to **ACCURACY** of each NDE test for steel bridges. **ACCURACY** is defined as the tendency of a test to measure false results. Please indicate, on average, the **ACCURACY** of each method for each category. *Note: Your percentage for each method should add to 100%.*

Definitions:

False Positive: Test measures damage, but there is no damage
False Negative: Test measures no damage, but there is damage
True Response: Test measures damage and there is damage OR test measures no damage and there is no damage

---

Example:

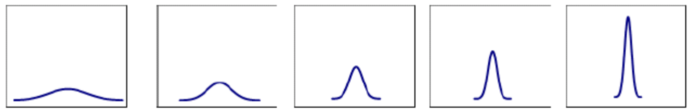
	False Positive	False Negative	True Positive	Don't Know
Acoustic Emission	0	0	100%	
Liquid Penetrant Testing	5%	10%	85%	

	False Positive	False Negative	True Positive	Don't Know
79) Acoustic Emission				
80) Liquid Penetrant Testing				
81) Magnetic Particle Testing - Crack Detection				
82) Magnetic Particle Testing - Weld Inspection				
83) Radiography				
84) Ultrasonic Testing - Crack Detection				
85) Ultrasonic Testing - Pin Inspection				
86) Ultrasonic Testing - Weld Inspection				
87) Visual Inspection				

The following questions pertain to **PRECISION** of each NDE test for steel bridges. **PRECISION** is defined as the reproducibility of a test in a controlled environment (i.e. a lab setting). Please mark the **PRECISION** of each method based on the scale provided. *(mark one for each method)*

PRECISION SCALE

Response:	1	2	3	4	5



Mean:	100	100	100	100	100
Standard Deviation:	35	22	12	8	6
Coefficient of Variation:	0.35	0.22	0.12	0.08	0.06

Note: The coefficient of variation is a normalized measure of dispersion of a probability distribution.

Example:

If Acoustic Emission is used to repeat a test multiple times in a controlled environment, and the results have a coefficient of variation of about 0.12 they have an average PRECISION and the response would be 3.

If Liquid Penetrant Testing is used to repeat a test multiple times in a controlled environment, and the results have a coefficient of variation of about 0.06 it has a high PRECISION and the response would be 5.

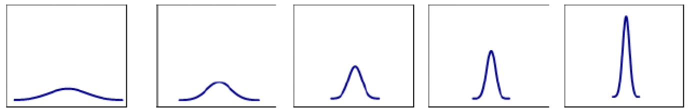
	1	2	3	4	5	Don't Know
Acoustic Emission			X			
Liquid Penetrant Testing					X	

	1	2	3	4	5	Don't Know
88) Acoustic Emission	☐	☐	☐	☐	☐	☐
89) Liquid Penetrant Testing	☐	☐	☐	☐	☐	☐
90) Magnetic Particle Testing - Crack Detection	☐	☐	☐	☐	☐	☐
91) Magnetic Particle Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
92) Radiography	☐	☐	☐	☐	☐	☐
93) Ultrasonic Testing - Crack Detection	☐	☐	☐	☐	☐	☐
94) Ultrasonic Testing - Pin Inspection	☐	☐	☐	☐	☐	☐
95) Ultrasonic Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
96) Visual Inspection	☐	☐	☐	☐	☐	☐

The following questions pertain to **RELIABILITY** of each NDE test for steel bridges. For the purposes of this survey **RELIABILITY** is defined as the reproducibility of a test in an uncontrolled environment (i.e. in the field). Please mark the **RELIABILITY** of each method based on the scale provided. *(mark one for each method)*

RELIABILITY SCALE

Response:	1	2	3	4	5



Mean:	100	100	100	100	100
Standard Deviation:	35	22	12	8	6
Coefficient of Variation:	0.35	0.22	0.12	0.08	0.06

Note: The coefficient of variation is a normalized measure of dispersion of a probability distribution.

Example:

If Acoustic Emission is used to repeat a test multiple times in a controlled environment, and the results have a coefficient of variation of about 0.12 they have an average RELIABILITY and the response would be 3.

If Liquid Penetrant Testing is used to repeat a test multiple times in a controlled environment, and the results have a coefficient of variation of about 0.06 it has a high RELIABILITY and the response would be 5.

	1	2	3	4	5	Don't Know
Acoustic Emission			X			
Liquid Penetrant Testing					X	

		1	2	3	4	5	Don't Know
97)	Acoustic Emission	☐	☐	☐	☐	☐	☐
98)	Liquid Penetrant Testing	☐	☐	☐	☐	☐	☐
99)	Magnetic Particle Testing - Crack Detection	☐	☐	☐	☐	☐	☐
100)	Magnetic Particle Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
101)	Radiography	☐	☐	☐	☐	☐	☐
102)	Ultrasonic Testing - Crack Detection	☐	☐	☐	☐	☐	☐
103)	Ultrasonic Testing - Pin Inspection	☐	☐	☐	☐	☐	☐
104)	Ultrasonic Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
105)	Visual Inspection	☐	☐	☐	☐	☐	☐

The following questions pertain to various **COSTS** of each NDE test for concrete bridges. For each segment you will be developing your own scale and marking your opinion of **COST** for each method based on this scale. This process will give us a better idea of your opinion on each cost relative to other participants. After making your scale, please mark the **COSTS** of each method. *(mark one for each)*

Example:

Personal Scale Made For **TIME SPENT RUNNING A TEST**

	Very Low	Low	Moderate	High	Very High
Scale	0 - 4 Hours	3 - 8 Hours	8 - 10 Hours	10 - 15 Hours	15+ Hours

If Acoustic Emission takes about 4 hours to run a test on the example bridge the TIME SPENT RUNNING A TEST (3 - 8 Hours) response would be Low.

If Liquid Penetrant Testing takes about 13 hours to run a test on the example bridge the TIME SPENT RUNNING A TEST (10 - 15 Hours) response would be High.

	Very Low	Low	Moderate	High	Very High	Don't Know
Acoustic Emission	X					
Liquid Penetrant Testing				X		

Please develop your scale and mark the average **TIME SPENT RUNNING A TEST** for each method for a 200 ft, 4 lane Steel Girder Bridge with a reinforced concrete deck. *(mark one for each method)*

Personal Scale Made For **TIME SPENT RUNNING A TEST**

	Very Low	Low	Moderate	High	Very High
Scale					

	Very Low	Low	Moderate	High	Very High	Don't Know
106) Acoustic Emission	☐	☐	☐	☐	☐	☐
107) Liquid Penetrant Testing	☐	☐	☐	☐	☐	☐
108) Magnetic Particle Testing - Crack Detection	☐	☐	☐	☐	☐	☐
109) Magnetic Particle Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
110) Radiography	☐	☐	☐	☐	☐	☐
111) Ultrasonic Testing - Crack Detection	☐	☐	☐	☐	☐	☐
112) Ultrasonic Testing - Pin Inspection	☐	☐	☐	☐	☐	☐
113) Ultrasonic Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
114) Visual Inspection	☐	☐	☐	☐	☐	☐

Please develop your own scale and mark the average **TIME SPENT ANALYZING DATA** for each method for a 200 ft, 4 lane Steel Girder Bridge with a reinforced concrete deck. *(mark one for each method)*

Personal Scale Made For **TIME SPENT ANALYZING DATA**

	Very Low	Low	Moderate	High	Very High
Scale					

		Very Low	Low	Moderate	High	Very High	Don't Know
115)	Acoustic Emission	☐	☐	☐	☐	☐	☐
116)	Liquid Penetrant Testing	☐	☐	☐	☐	☐	☐
117)	Magnetic Particle Testing - Crack Detection	☐	☐	☐	☐	☐	☐
118)	Magnetic Particle Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
119)	Radiography	☐	☐	☐	☐	☐	☐
120)	Ultrasonic Testing - Crack Detection	☐	☐	☐	☐	☐	☐
121)	Ultrasonic Testing - Pin Inspection	☐	☐	☐	☐	☐	☐
122)	Ultrasonic Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
123)	Visual Inspection	☐	☐	☐	☐	☐	☐

Please develop your own scale and mark the average **TIME TO TRAIN AN INSPECTOR** to be considered able to competently perform the test method for a 200 ft, 4 lane Steel Girder Bridge with a reinforced concrete deck. *(mark one for each method)*

Personal Scale Made For **TIME TO TRAIN AN INSPECTOR**

	Very Low	Low	Moderate	High	Very High
Scale					

		Very Low	Low	Moderate	High	Very High	Don't Know
124)	Acoustic Emission	☐	☐	☐	☐	☐	☐
125)	Liquid Penetrant Testing	☐	☐	☐	☐	☐	☐
126)	Magnetic Particle Testing - Crack Detection	☐	☐	☐	☐	☐	☐
127)	Magnetic Particle Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
128)	Radiography	☐	☐	☐	☐	☐	☐
129)	Ultrasonic Testing - Crack Detection	☐	☐	☐	☐	☐	☐
130)	Ultrasonic Testing - Pin Inspection	☐	☐	☐	☐	☐	☐
131)	Ultrasonic Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
132)	Visual Inspection	☐	☐	☐	☐	☐	☐

Please develop your own scale and mark the average **MONETARY COST FOR THE EQUIPMENT** for each method for a 200 ft, 4 lane Steel Girder Bridge with a reinforced concrete deck. *(mark one for each method)*

Personal Scale Made For **MONETARY COST FOR THE EQUIPMENT**

	Very Low	Low	Moderate	High	Very High
Scale					

	Very Low	Low	Moderate	High	Very High	Don't Know
133) Acoustic Emission	☐	☐	☐	☐	☐	☐
134) Liquid Penetrant Testing	☐	☐	☐	☐	☐	☐
135) Magnetic Particle Testing - Crack Detection	☐	☐	☐	☐	☐	☐
136) Magnetic Particle Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
137) Radiography	☐	☐	☐	☐	☐	☐
138) Ultrasonic Testing - Crack Detection	☐	☐	☐	☐	☐	☐
139) Ultrasonic Testing - Pin Inspection	☐	☐	☐	☐	☐	☐
140) Ultrasonic Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
141) Visual Inspection	☐	☐	☐	☐	☐	☐

Please develop your own scale and mark the average **NUMBER OF INSPECTORS NEEDED** for each method for a 200 ft, 4 lane Steel Girder Bridge with a reinforced concrete deck. *(mark one for each method)*

Personal Scale Made For **NUMBER OF INSPECTORS NEEDED**

	Very Low	Low	Moderate	High	Very High
Scale					

	Very Low	Low	Moderate	High	Very High	Don't Know
142) Acoustic Emission	☐	☐	☐	☐	☐	☐
143) Liquid Penetrant Testing	☐	☐	☐	☐	☐	☐
144) Magnetic Particle Testing - Crack Detection	☐	☐	☐	☐	☐	☐
145) Magnetic Particle Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
146) Radiography	☐	☐	☐	☐	☐	☐
147) Ultrasonic Testing - Crack Detection	☐	☐	☐	☐	☐	☐
148) Ultrasonic Testing - Pin Inspection	☐	☐	☐	☐	☐	☐
149) Ultrasonic Testing - Weld Inspection	☐	☐	☐	☐	☐	☐
150) Visual Inspection	☐	☐	☐	☐	☐	☐

End of Questionnaire. Please make sure you have completed both sections of the questionnaire. After that, please return the completed survey either by mail or email as described previously. If you have any comments please include them in the extra space provided.

Thank you for completing this questionnaire and successive questionnaires in this study. Your responses will help improve the use of NDE techniques.

Extra space if needed:

Example of Cover Letter

3/4/2013

Dear Potential Participant,

My name is Alex Hesse and I am a researcher from Colorado State University in the Civil and Environmental Engineering Department. We are conducting a research study to attempt to better understand the use of nondestructive evaluation methods (NDE) on bridges. The title of our project is *Using Expert Opinion to Quantify Accuracy and Reliability of Nondestructive Evaluation on Bridges*. The Principal Investigator is Rebecca Atadero in the Civil and Environmental Engineering Department and the Co-Principal Investigator is Alex Hesse from the Civil and Environmental Engineering Department. This study is funded by the Mountain Plains Consortium, the Region 8 University Transportation Center sponsored by the USDOT.

The purpose of this survey is to help determine the accuracy, precision, sensitivity, reliability, and various costs associated with commonly used NDE techniques on bridges. The goal of the researchers is to quantify and better understand these properties of the various NDE techniques for use in inspection planning. Ultimately we hope to encourage the use of NDE on bridges by developing risk based plans for inspection. We would like the most experienced personnel to complete this survey. If you feel you are not experienced in NDE techniques on bridges please pass this along to a more experienced person.

It is our belief that the routine visual inspections conducted every two years on most bridges are not the most effective allocation of resources, but in order to design a more efficient process we need better information about the accuracy and reliability of NDE methods. Your information was found from a public source that indicated you might be knowledgeable in this area. Only a small portion of NDE experts are being contacted and your experiences and opinions on the subject are very important. Results from this survey will hopefully contribute to making our nation's bridges safer for the public.

We would like you to complete a Delphi method survey about nondestructive evaluation techniques on bridges. The process will involve the completion of approximately four to five surveys over the course of about five months. Participation will take approximately 20 minutes for each survey totaling about 1.5 hours of participation throughout the process. Your participation in this research is voluntary. If you decide to participate in the study, you may withdraw your consent and stop participation at any time without penalty.

All surveys will be kept confidential and only the research team will have access to the responses. All subsequent surveys will only use number and/or letter identifiers to distinguish between participants. All responses will be stored separately from the identification information and all information will be either locked or password protected. While there are no direct benefits to you, we hope to gain knowledge allowing us to improve the bridge inspection process.

There are no known risks associated with this process. As stated previously, all responses will be kept confidential and will only be presented in aggregate form. It is not possible to identify all potential risks in research procedures, but the researchers have taken reasonable safeguards to minimize any known and potential, but unknown, risks.

Please complete the enclosed questionnaire and return it by October 19, 2012 either by email to aahesse@engr.colostate.edu or by mail in the provided postage-paid envelope:

Department of Civil and Environmental Engineering
Colorado State University
1372 Campus Delivery
Fort Collins, CO 80523-1372
ATTN: Alex Hesse

If you have any questions, please contact Alex Hesse at (608) 354-5865 or aahesse@engr.colostate.edu or Rebecca Atadero at (970) 491-3584 or ratadero@engr.colostate.edu. If you have any questions about your rights as a volunteer in this research, contact Janell Barker, Human Research Administrator, at 970-491-1655.

Sincerely,

Rebecca Atadero	Alex Hesse
Assistant Professor	Master's Student

APPENDIX B: STABILITY RESULTS

Table 1A. Stability Results from the Bias Subsection from Round 2 to Round 3

Concrete Method	r_S
Cover Meters/Pachometer	1.00
Impact Echo	1.00
Radar	1.00
Ultrasonic Testing	0.99
Visual Inspection	0.85
Steel Method	r_S
Acoustic Emission	1.00
Radiography	1.00
Ultrasonic Testing - Crack Detection	0.99
Ultrasonic Testing - Pin Inspection	1.00
Ultrasonic Testing - Weld Inspection	1.00
Visual Inspection	0.98

Table 2A. Stability Results from the Reliability Subsection from Round 2 to Round 3

Concrete Method	r_S
Cover Meters/Pachometer	1.00
Electrical Potential	1.00
Impact Echo	0.99
Mechanical Sounding	0.98
Radar	0.99
Rebound Hammer	1.00
Thermal	1.00
Ultrasonic Testing	0.97
Visual Inspection	0.99
Steel Method	r_S
Acoustic Emission	0.99
Liquid Penetrant Testing	0.97
Magnetic Particle Testing - Crack Detection	0.93

Magnetic Particle Testing - Weld Inspection	0.97
Radiography	0.99
Ultrasonic Testing - Crack Detection	0.98
Ultrasonic Testing - Pin Inspection	0.98
Ultrasonic Testing - Weld Inspection	0.99
Visual Inspection	0.99

Table 3A. Stability Results from the Reliability Subsection from Round 2 to Round 3 (Unstable Results are Highlighted)

r_s			
	False Positive	False Negative	True Response
Concrete NDE Method			
Electrical Potential	0.81	-49.79	-4.94
Mechanical Sounding	0.63	0.82	-2.84
Thermal	0.95	-23.94	-3.23

Visual Inspection	-0.27	-46.47	-11.99
Steel NDE Method			
Acoustic Emission	1.00	1.00	1.00
Liquid Penetrant	0.25	0.19	-4.94
Magnetic Particle – Crack	0.09	-0.72	-1.19
Magnetic Particle – Weld	0.10	-0.69	-0.96
Radiography	0.76	0.53	-3.46
Ultrasonic – Crack	-1.07	-1.07	-8.47
Ultrasonic – Pin	-1.03	-0.85	-6.59
Ultrasonic - Weld	0.81	0.93	0.23
Visual Inspection	-0.01	-1.05	-4.15

Table 4A. Stability Results from the Costs Subsection from Round 3 to Round 4 for the Concrete NDE Methods

r_s					
	Time Spent Running a Test	Time Spent Analyzing Data	Time to Train an Inspector	Monetary Cost for Equipment	Number of Inspectors Needed
Cover Meters/Pachometer	0.98	0.99	0.98	0.98	0.98
Electrical Potential	0.99	0.99	0.99	0.99	0.99
Impact Echo	1.00	0.98	1.00	0.97	0.99
Mechanical Sounding	0.98	0.99	0.99	1.00	0.97
Radar	0.96	0.99	1.00	0.99	0.93
Rebound Hammer	1.00	0.98	0.98	1.00	0.98
Thermal	0.98	1.00	0.99	1.00	0.98
Ultrasonic Testing	0.96	0.99	1.00	1.00	0.99
Visual Inspection	0.98	0.98	0.98	1.00	0.97

Table 5A. Stability Results from the Costs Subsection from Round 3 to Round 4 for the Steel NDE Methods

r_s					
	Time Spent Running a Test	Time Spent Analyzing Data	Time to Train an Inspector	Monetary Cost for Equipment	Number of Inspectors Needed
Acoustic Emission	1.00	1.00	1.00	1.00	1.00
Liquid Penetrant	0.95	0.98	0.98	0.99	1.00
Magnetic Particle – Crack	0.99	0.98	0.98	0.99	0.99
Magnetic Particle – Weld	0.99	0.98	0.98	0.99	0.99
Radiography	0.99	0.98	1.00	0.99	0.95
Ultrasonic – Crack	0.99	0.98	1.00	0.97	0.99
Ultrasonic – Pin	0.95	0.98	1.00	0.97	0.99
Ultrasonic - Weld	0.98	0.98	1.00	0.97	0.99
Visual Inspection	0.92	0.99	0.98	0.99	1.00